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.

1   **Estimating the incidence and diagnosed proportion of HIV infections in Japan: a statistical**

2   **modeling study**

4   Hiroshi Nishiura[1,2]

6   [1] Graduate School of Medicine, Hokkaido University, Sapporo, Japan

7   [2] CREST, Japan Science and Technology Agency, Saitama, Japan

9   Corresponding Author:

10   Hiroshi Nishiura

11   Graduate School of Medicine, Hokkaido University, Sapporo, Japan

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

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

15   **Abstract**

16   **Background.** Epidemiological surveillance of HIV infection in Japan involves two technical

17   problems for directly applying a classical backcalculation method, i.e., (i) all AIDS cases are not

18   counted over time and (ii) people diagnosed with HIV have received antiretroviral therapy,

19   extending the incubation period. The present study aimed to address these issues and estimate the

20   HIV incidence and the proportion of diagnosed HIV infections, using a simple statistical model.

21   **Methods.** From among Japanese nationals, yearly incidence data of HIV diagnoses and patients

22   with AIDS who had not previously been diagnosed as HIV positive, from 1985 to 2017, were

23   analyzed. Using the McKendrick partial differential equation, general convolution-like equations

24   were derived, allowing estimation of the HIV incidence and the time-dependent rate of

25   diagnosis. A likelihood-based approach was used to obtain parameter estimates.

26   **Results.** Assuming that the median incubation period was 10.0 years, the cumulative number of

27   HIV infections was estimated to be 29,613 (95% confidence interval (CI): 29,059, 30,167) by the

28   end of 2017, and the proportion of diagnosed HIV infections was estimated at 80.3% (95% CI:

29   78.7%, 82.0%). Allowing the median incubation period to range from 7.5 to 12.3 years, the

30   estimate of the proportion diagnosed can vary from 77% to 84%.

31   **Discussion.** The proportion of diagnosed HIV infections appears to have not yet reached 90%

32   among Japanese nationals. Compared with the peak incidence from 2005–2008, new HIV

33   infections have clearly been in a declining trend; however, there are still more than 1,000 new

34   HIV infections per year in Japan. To increase the diagnosed proportion of HIV infections, it is

35   critical to identify people who have difficulty accessing consultation, testing, and care, and to

36   explore heterogeneous patterns of infection.

37

38   **Introduction**

39   Following an infection with human immunodeficiency virus (HIV), development of acquired

40   immunodeficiency syndrome (AIDS) takes about 10 years (Munoz et al., 1997). The long

41   incubation period makes it difficult to directly observe the incidence and prevalence of HIV

42   infections over time. To offer insights into the epidemiology of HIV-infected and -incubating

43   individuals over time, and to evaluate public health control programs, various statistical

44   modeling approaches have been proposed to date (Brookmeyer & Gail, 1994; Donnelly & Cox,

45   2001; Jewell et al., 1992). Of these, a backcalculation method using a simple integral equation to

46   model AIDS incidence as arising from the HIV incidence convoluted with the independently and

47   identically distributed incubation period allows estimation of the HIV incidence based on

48   epidemiological surveillance data (Brookmeyer & Gail, 1986; Gail & Brookmeyer, 1988).

49   Assuming that the reported number of AIDS cases certainly and accurately captures the actual

50   number of AIDS incidence in industrialized countries, the backcalculation method greatly

51   improves our understanding of the epidemiology of HIV infection, attributing the observed

52   AIDS data to HIV infection events as a function of time.

53   Understanding the transmission dynamics of HIV using such statistical models is in line

54   with the concept of treatment cascade, introduced by the Joint United Nations Programme on

55   HIV/AIDS (UNAIDS). The so-called care cascade aims to identify and fill gaps in the

56   continuum of services for testing, care, and effective treatment of HIV (UNAIDS, 2014). In

57   relation to this, the UNAIDS report has led to the global initiative "90–90-90" by 2020 that sets

58   out goals in care cascades to achieve the following: 90% of people living with HIV know their

59   HIV status, 90% of people diagnosed with HIV have access to antiretroviral therapy (ART), and

60   90% of people receiving ART have suppressed viral loads (UNAIDS, 2014). UNAIDS even

61   aims to achieve 95-95-95 at a global level by the year 2030, contributing to successfully

62   controlling HIV and AIDS, as supported by the so-called test-and-treat strategy (Granich et al.,

63   2009; Granich et al., 2017). To quantify the situation of each country, monitoring diagnosed

64   individuals is essential; moreover, estimation of the diagnosed proportion of HIV infections must

65   be supported by firm scientific methods, to estimate the first part of the three 90-90-90 targets,

66   i.e., 90% of HIV-infected people know their HIV status. In Japan, an analysis of blood donors

67   took place in 2017, in which it was estimated that 85.6% of HIV-infected individuals, regardless

68   of nationality, were diagnosed (Iwamoto et al., 2017). Nevertheless, it is known that the analysis

69   of voluntary blood donation data is prone to sampling bias of donors owing to the tendency of

70   people with high risk to repeatedly undertake anonymous laboratory testing through the practice

71   of blood donation (Kihara et al., 2000), and moreover, an exclusion of repeaters can result in

72   underestimation of the prevalence, resulting in overestimation of the fraction diagnosed, even

73   though such screening of a large number of people is very costly. Considering the need to

74   achieve continued monitoring of the diagnosed proportion of HIV-infected individuals,

75   development of a reasonable yet scientifically rigorous method based on other datasets would be

76   beneficial, especially using epidemiological surveillance data (Hsieh et al., 2012; Cuadros &

77   Abu-Raddad, 2016; Hsieh & Lin, 2016; Mumtaz et al., 2018).

78   Despite the clear need for epidemiological estimation of the number of undiagnosed HIV

79   infections, the surveillance data in Japan possesses two technical problems. First, while the

80   definition of AIDS has remained nearly unchanged over time, reporting AIDS cases that were

81   previously diagnosed as HIV-infected cases has never been mandated (Nishiura, 2007). This

82   makes it impossible to directly apply the simplest convolution equation to the data because the

83   backcalculation method requires the count of all AIDS cases over time. Surveillance in Japan has

84   only consistently counted (i) HIV infections without AIDS at the time of diagnosis and (ii) AIDS

85   cases without previous diagnosis of HIV infection. Second, ART has been widespread since

86   1997 and has continuously improved the prognosis of HIV infection. Explicit incorporation of

87   treatment requires us to account for not only the treatment coverage but also the treatment details

88   (e.g., details of combination therapy), adherence, and many other factors. While there are a

89   number of possible methods to address these issues, including those using CD4 data (e.g. van

90   Sighem et al., 2017) or molecular biomarkers, a simple yet tractable estimation method that rests

91   on epidemiological surveillance data and that can reasonably overcome these problems is called

92   for.

93   In the present study, the aim was to address the abovementioned issues, estimating the

94   HIV incidence among Japanese nationals, and also to offer statistical estimates of undiagnosed

95   HIV infections and the proportion of diagnosed HIV infections over time.

96   **Materials & Methods**

97   **Surveillance data of HIV and AIDS in Japan**

98   The present study investigated the epidemiological surveillance data of HIV and AIDS in Japan,

99   which is publicly reported by the Committee of AIDS Trends (2018), belonging to the Ministry

100   of Health, Labor and Welfare, Japan. Of the reported datasets, our analyses are focused on

101   Japanese nationals because estimation of infection among foreigners requires accounting for

102   human migration, and the decision of migratory behavior (e.g. leaving Japan) is highly

103   dependent on the diagnosis of HIV infection and AIDS. As of the end of 2017, there were 16,663

104   HIV infections and 7,587 AIDS cases among Japanese nationals (Committee of AIDS Trends,

105   2018). As mentioned, HIV diagnoses reflect HIV-infected individuals who undertook voluntary

106   diagnostic testing before the onset of AIDS. An AIDS case indicates a patient who has never

107   been diagnosed with HIV infection prior to an AIDS diagnosis and who meets the clinical

108   diagnostic criteria: (i) confirmed HIV infection and (ii) the presence of one of 23 indicator

109   diseases representing opportunistic infections or tumors. According to the Infectious Disease

110   Law, HIV and AIDS are classified as a category V notifiable disease, and once diagnosed,

111   physicians must notify the case within 7 days of diagnosis. In the present study, the yearly

112   incidence of HIV infections and AIDS diagnoses from 1985 to 2017 was used. The data are

113   structured by sex and also by the most likely route of transmission (e.g., heterosexual,

114   homosexual or intravenous drug use, based on a physician's interview of patients). The latter

115   information, i.e., the mode of transmission, is discarded because it is believed that a substantial

116   proportion of men having sex with men do not disclose the actual contact and inform physicians

117   that they acquired infection through heterosexual contact (Inoue et al., 2015). Thus, a stratified

118   estimation by sex was conducted. Although the magnitude of the epidemic in Japan is relatively

119   small compared with that in Western industrialized countries, the incidence of HIV infection in

120   Japan is believed to have steadily increased over time, especially among men who have sex with

121   men and young adults (Kihara et al., 2003; Nemoto, 2004).

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

123   The proposed statistical model is derived from the following partial differential equation

124   (PDE) model, which is referred to as the McKendrick equation (Nishiura & Inaba, 2011; Ejima

125   et al., 2014). Figure 1 shows a compartmental diagram of the data-generating process. Once

126   infected with HIV, individuals who are undiagnosed and in the incubation period experience two

127   different hazards, i.e., the force of HIV diagnosis $\alpha(t)$ that depends on calendar time t and the

128   hazard of illness onset $\rho(\tau)$ that depends on the time elapsed since infection τ. Let $h(t,\tau)$ be

129   undiagnosed incubating HIV infections at calendar time t and the time since infection τ (i.e.,

130   undiagnosed HIV infection without AIDS), the dynamics of HIV diagnosis and illness onset are

131   described by

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

133   with a boundary condition

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

135   where $\lambda(t)$ represents the HIV incidence (i.e., the number of new HIV infections) at calendar time

136   t. It should be noted that $\rho(\tau)$ yields $f(\tau)$, the probability density function of the incubation period

137   as follows:

$$f(s) = \rho(s)exp\left(-\int_{0}^{s}\rho(y)dy\right), \quad (3)$$

139   for $s>0$. It is well known that the McKendrick equation can be solved along the characteristic

140   line, i.e.,

$$h(t,s) = \lambda(t-s)exp\left(-\int_{t-s}^{t}\alpha(x)dx - \int_{0}^{s}\rho(y)dy\right), \quad (4)$$

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

143   $u(t)$, is written as

$$u(t) = \int_{0}^{t}\alpha(t)h(t,s)ds$$

$$= \int_{0}^{t}\alpha(t)\lambda(t-s)exp\left(-\int_{t-s}^{t}\alpha(x)dx - \int_{0}^{s}\rho(y)dy\right)ds, \quad (5)$$

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

147   $a(t) = \int_{0}^{t}\rho(s)h(t,s)ds$

148
$$= \int_0^t \rho(s)\lambda(t-s)exp\left(-\int_{t-s}^t \alpha(x)dx - \int_0^s \rho(y)dy\right)ds. \quad (6)$$

149 Equations (5) and (6) read similarly to the so-called extended backcalculation (Hall et al., 2008),

150 which is derived from a competing risk model (Marschner, 1994; Cui & Becker, 2000). The

151 abovementioned process can be used as the generalization.

Statistical model and estimation

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

154 and (6) as

155
$$u_t = \sum_{s=1}^t \lambda_{t-s}\alpha_t \prod_{x=t-s-1}^{t-1}(1-\alpha_x)\prod_{y=1}^{s-1}(1-\rho_y), \quad (7)$$

156 and

157
$$a_t = \sum_{s=1}^t \lambda_{t-s}\rho_s \prod_{x=t-s-1}^{t-1}(1-\alpha_x)\prod_{y=1}^{s-1}(1-\rho_y). \quad (8)$$

158 There is no prior notion as to the shape of the epidemic curve (i.e. the frequency of transmission)

159 over time. Thus, the incidence of HIV infection in year t, λ_t, is modeled as a step function:

160
$$\lambda_t = \begin{cases} \lambda_1 & \text{for } t < 1989, \\ \lambda_2 & \text{for } 1989 \leq t < 1993, \\ \vdots & \\ \lambda_9 & \text{for } 2013 \leq t, \end{cases} \quad (9)$$

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

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

163
$$\alpha_t = \begin{cases} \alpha_1 & \text{for } t < 1989, \\ \alpha_2 & \text{for } 1989 \leq t < 1993, \\ \vdots & \\ \alpha_9 & \text{for } 2013 \leq t. \end{cases} \quad (10)$$

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

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

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

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

168   continuous version as

169

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

170   and

171

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

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

173   computed as

174

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

175   The diagnosed proportion of HIV infections is calculated either as $\Sigma(a+u)/\Sigma(x+a+u)$ or $\Sigma u/$

176    $\Sigma(x+u)$, taking the summations over time. The former calculates the proportion of diagnosed

177   HIV-positive individuals out of the cumulative number of HIV-positive individuals. This

178   calculation has the drawback of including patients with AIDS who have already died by the year

179   of calculation. As of 2017, it has been reported that a total of 2,321 cases resulted in death

180   (Iwamoto et al., 2017). Alternatively, the latter calculates the fraction of individuals who are HIV

181   positive but have not yet developed AIDS out of the cumulative number of HIV-positive

182   individuals but including undiagnosed individuals, considering that the incubation period in most

183   cases of HIV infection is now considerably extended by ART. The drawback of the latter

184   calculation is that patients with AIDS who have survived and have received ART are excluded;

185   thus, the calculated proportion may not be strictly in line with the target figure in the first goal of

186   the 90-90-90 initiative. Therefore, when estimating the undiagnosed number of HIV infections

187   and the diagnosed proportion at the end of 2017, both calculations are made, and the former is

188   adjusted by subtracting 2,321 AIDS deaths from the cumulative count of AIDS cases.

189   To quantify the proposed system of equations, we estimate parameters λ_t and α_t by means

190   of the maximum likelihood method. Considering that HIV infections are generated as the

191   nonhomogenous Poisson process, the resulting HIV diagnoses and AIDS cases would also

192   follow Poisson distributions. The likelihood function of HIV diagnoses is

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

194   where r_t denotes the reported (observed) number of HIV diagnoses in year t in the surveillance

195   record. Similarly, the likelihood of new AIDS diagnoses is

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

197   where w_t denotes the reported number of new AIDS diagnoses in year t. Consequently, the total

198   likelihood L is given by

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

200   Maximum likelihood estimates of parameters are obtained by minimizing the negative logarithm

201   of Eq. (16). As mentioned above, the incubation period distribution is assumed as known, and to

202   address the uncertainty, three different estimates are derived from published studies (Boldson et

203   al., 1988; Brookmeyer & Goedert, 1989; Munoz & Xu, 1996). A widely cited estimate by

204   Brookmeyer & Goedert (1989) was derived from the study of patients with hemophilia over 20

205   years of age with η=11.6 and k=2.5, resulting in a median incubation period of 10.0 years.

206   Boldson et al. (1988) investigated a cohort of AIDS cases in San Francisco with η=14.3 and

207   k=2.5, yielding a median incubation period of 12.3 years. The estimate by Munoz and Xu (1996)

208   was obtained from the Multicenter AIDS Cohort Study with η=10.0 and k=1.3, and the median

209   incubation period is 7.5 years. All three estimates have been used in the present study to address

210   uncertainty with respect to the incubation period. In addition to (14) and (15), we have also

211   explored the over-dispersed likelihood function, employing the negative binomial distribution

212   (Althaus, 2015; Riou et al., 2018) and compared the Akaike Information Criterion (AIC) against

213   Poisson distributed likelihood, as part of sensitivity analysis.

214   The 95% confidence interval (CI) of parameters was derived from the profile likelihood.

215   The 95% CI of model estimates (e.g., the number of undiagnosed HIV infections and the

216   proportion diagnosed) was derived using a parametric bootstrap method. In the bootstrapping

217   exercise, model parameters were resampled from a multivariate normal distribution with vectors

218   of mean $\boldsymbol{\theta}$ and standard deviation $\boldsymbol{\sigma}$. The latter vector was derived from the covariance matrix,

219   taking diagonal elements of the inverse Hessian matrix ($\sigma^2 = \mathrm{diag}(\mathrm{H}^{-1}(\theta))$). For each set of

220   parameters, the model solution is obtained, and 1,000 times of parameter resampling results in a

221   simulated distribution of model solutions. By taking the 2.5th and 97.5th percentile points of the

222   simulated distribution, the 95% CI is obtained. All statistical data were analyzed using R version

223   3.1 (Comprehensive R Archive Network) (R Core Team, 2016) and JMP version 12.0.1

224   statistical software (SAS Institute Inc., Cary, NC, USA).

225   **Ethical considerations**

226   In the present study, the analyzed data are publicly available (Committee of AIDS Trends, 2018).

227   As such, the datasets used in our study are deidentified and fully anonymized in advance, and the

228   analysis of publicly available data with no identifying information does not require ethical

229   approval.

230   **Results**

231   Estimated parameters, i.e., yearly incidence and yearly probability of diagnosis, are

232   shown in Fig. 2. With the assumed median incubation period of 10.0 years, the yearly incidence

233   was the highest from 2005–2008, with an estimated 1,972 (95% CI: 1,829, 2,115) infections per

234   year (Fig. 2A). Subsequently, the incidence began to decline; the yearly estimate in the most

235   recent interval (from 2013–2017) was 1,179 (95% CI: 1,047, 1,293) infections. The yearly

236   probability of diagnosis has monotonously improved over time (Fig. 2B). The estimated

237   diagnosis probability by 1999 was less than 10%, but the latest estimate from 2013–2017 was

238   15.6% (95% CI: 14.8%, 16.4%). The qualitative patterns of HIV incidence and diagnosis did not

239   vary greatly, even when shorter and longer median incubation periods were used (Figs. 2C and

240   2D). Figures 2E and 2F show maximum likelihood estimates of the incidence and probability of

241   diagnosis by sex. The incidence in males was the highest from 2005–2008; the latest estimate

242   from 2013–2017 ranged from 1,015 to 1,363 infections per year, assuming a median incubation

243   period from 7.5 to 12.3 years. Similarly, the incidence in females was highest from 1993–1996,

244   ranging from 86 to 97 infections per year; the latest yearly incidence ranged from 31 to 54

245   infections with a median incubation period of 7.5 to 12.3 years. The yearly probability of

246   diagnosis among males behaved similarly to that of the entire population, but there was no

247   apparent improvement in the frequency of diagnosis among females. In general, female enjoyed

248   higher rate of diagnosis than male. For the entire population with median incubation period at

249   10.0 years, AIC with the Poisson distributed likelihood was 650.7, while that with negative

250   binomially distributed likelihood was 847.8, indicating that the dataset was not over-dispersed.

251   Figure 3 shows a comparison between the observed and predicted number of HIV

252   diagnoses and AIDS cases. All three models with different median incubation periods yielded

253   almost identically good fit to the data (Fig. 3A), reflecting mutual compensations between λ and

254   α to fit to the data. Even though the number of diagnosed HIV infections and AIDS cases was

255   relatively small for females, the proposed model successfully captured the observed patterns of

256   HIV diagnoses and AIDS cases by sex (Fig. 3B).

257   Figure 4 shows the estimated undiagnosed number of HIV infections and the estimated

258   proportion of diagnosed HIV-positive individuals over time, among Japanese nationals. Using

259   the median incubation period of 10.0 years (Fig. 4A), undiagnosed HIV infection was estimated

260   to have peaked in 2009 with 7,532 (95% CI: 6,911, 8,152) infections. In the latest time interval,

261   from 2013–2017, it was estimated that 5,363 (95% CI: 4,809, 5,917) infections remained

262   unrecognized. Varying the median incubation period from 7.5 to 12.3 years, the maximum

263   likelihood estimate of undiagnosed HIV infections in the latest time interval ranged from 4,041

264   to 6,552 infections. These findings indicate that the cumulative number of HIV infections by the

265   end of 2017 was 29,613 (95% CI: 29,059, 30,167) Japanese nationals, using the median

266   incubation period of 10.0 years, and can range from 28,291 to 30,802 individuals.

267   Including and excluding AIDS cases, the estimated proportions of diagnosed HIV

268   infections are shown in Figs. 4C and 4D. Including AIDS cases, the diagnosed proportion was

269   estimated at 81.9% (range 78.7% to 85.7%) using the median incubation period of 10.0 (7.5 to

270   12.3) years. Excluding AIDS cases, the estimate was 75.7% (range 71.8% to 80.5%). Figures 4E

271   and 4F show the estimated number of undiagnosed HIV infections and the diagnosed proportion

272   by sex, excluding AIDS cases. Estimates of undiagnosed HIV infections among males behaved

273   similarly to the entire population of Japanese nationals, whereas those of females peaked in the

274   year 2001. In the latest time interval (2013–2017), it was estimated that 5,150 infections (range

275   3,881 to 6,287) in males and 210 infections (range 162 to 255) in females remained

276   unrecognized, using the median incubation period of 10.0 (with the range of 7.5 to 12.3) years.

277   The diagnosed proportion of both males and females increased with time, and females tended to

278   yield higher estimates than males. In the latest time interval from 2013–2017, the diagnosed

279   proportion (excluding AIDS cases) was estimated at 75.3% (range 71.4% to 80.2%) among

280   males and 82.1% (range 79.1% to 85.6%) among females.

281   Figure 5 shows the undiagnosed number of HIV infections and the proportion of

282   diagnosed infections at the end of 2017. The uncertainty bound was greatest with an assumed

283   median incubation period of 12.3 years, with an estimated 6,552 infections (95% CI: 5,632,

284   7,471). Figure 5B shows the diagnosed proportion, including and excluding AIDS cases, with

285   95% confidence intervals. Even when AIDS cases were included, the 2,321 deaths known up to

286   that point were subtracted from AIDS cases in advance of the calculation. Assuming that the

287   median incubation period was 10.0 years, the calculation, inclusive of surviving AIDS cases,

288   yielded 80.3% (95% CI: 78.7%, 82.0%); when excluding AIDS cases, the proportion was 75.7%

289   (95% CI: 73.8%, 77.6%).

290   **Discussion**

291   The present study estimated the incidence and diagnosed fraction of HIV infections among

292   Japanese nationals, devising an original model that captures the data generating process of HIV

293   and AIDS in the epidemiological surveillance. By the end of 2017, the cumulative number of

294   HIV infections was estimated to be about 30,000 cases, of which 4,000 to 6,000 were considered

295   to have remained undiagnosed. Assuming that the median incubation period was 10.0 years, 80%

296   of infections have ever been diagnosed; accounting for the uncertainty in a median incubation

297   period ranging from 7.5 to 12.3 years, the estimate of the diagnosed proportion can range from

298   77% to 84%. To the author's knowledge, the present study is the first to offer firm statistical

299   estimates of the incidence and diagnosed proportion of HIV infections based on epidemiological

300   surveillance data in Japan, using an explicit mathematical modeling approach.

301   There are two take-home messages from the results of this study. First, regardless of

302   whether AIDS cases are included, the proportion of diagnosed HIV infections appears not to

303   have reached 90% among Japanese nationals. Although some estimates exceed 80%, even after

304   subtraction of known deaths owing to AIDS, the findings echo those of a published study that

305   analyzed blood donor data (Iwamoto et al., 2017). The published blood donor-based estimate

306   indicated that 85.6% have been diagnosed at the end of 2015, which was prone to sampling bias

307   with substantial potential of both over- and underestimation, and the present study validated that

308   the surveillance-based estimate was slightly below and not too far from the published figure.

309   These findings pose a critical problem in Japan for controlling HIV and AIDS. In the present

310   study, the rate of diagnosis was shown to have improved with time, and the trend was

311   particularly apparent among men, mainly comprising men having sex with men (MSM). The

312   findings of the present study indicate that there would be a certain number of infected individuals

313   who may not have proper access to consultation, testing, and care with privacy protection. To

314   identify the attributes of such HIV-infected individuals in greater detail, the investigation must

315   be extended to explore heterogeneous patterns, including age-dependence, spatial heterogeneity,

316   and other background characteristics. These are my ongoing research interests.

317   Second, compared with the peak from 2005–2008, the incidence showed a declining

318   trend. Compared with the estimate in 2005–2008, the upper 95% CIs of the next two time

319   periods (2009–2012 and 2013–2017) were significantly lower than those in the peak period. In

320   fact, a declining trend has also been seen in other datasets, including the incidence of counseling

321   and blood testing at local health centers and the proportion of HIV-positive blood donors over

322   time (Committee of AIDS Trends, 2018). The present study results support that these observed

323   declines are partially attributable to actual decreases in the incidence of HIV infection in Japan.

324   The underlying mechanisms of such decreases have yet to be explored using a mathematical

325   model, perhaps requiring modeling of the saturated effect (Heesterbeek & Metz, 1993) together

326   with statistical estimates of the effective reproduction number (Kretzschmar et al., 2013). The

327   success of Japan's controlling HIV transmission among the core population, i.e. MSM, reflecting

328   the set-up of gay community centers and scale-up of gay non-governmental organizations'

329   activities (Sherriff et al., 2017), could potentially be objectively demonstrated in such an

330   analysis. In addition, it must be remembered that the yearly incidence still remains above 1,000

331   infections; moreover, such a declining trend is not evident among females, although the rate of

332   diagnosis in female has steadily been higher than that in male. Higher frequency of diagnosis

333   among female than male might reflect better awareness of the risk that results from heterosexual

334   transmission (e.g. through foreign partner's diagnosis).

335   Although the present study was motivated by the need for quantifying the care cascade in

336   Japan, in accordance with the goals of 90-90-90, a few technical issues must be noted to interpret

337   the estimates and apply the present results to the evaluation. First, Japanese estimates of the latter

338   two goals of the 90-90-90 initiative, i.e., access to ART and virus suppression, rest on

339   questionnaire surveys conducted in the prefectures, which do not distinguish between infected

340   individuals who are Japanese nationals and those who are not (Iwamoto et al., 2017). Thus, our

341   estimates of the diagnosed proportion of HIV infections among Japanese nationals alone cannot

342   immediately be compared with subsequent existing proportions as if they were sampled from the

343   same population. Whereas estimation of the HIV incidence among non-Japanese nationals is an

344   ongoing research subject, it is frequently the case that infection with HIV or illness onset of

345   AIDS acts as a trigger for foreigners to leave the country; therefore, incorporation of their

346   involvement in the transmission dynamics of Japan requires that very careful attention be paid to

347   migration events, and ideally, that information is supported by individual-based data. Second, the

348   clinical definition of AIDS in Japan depends on indicator diseases, imposing a certain extent of

349   uncertainty in diagnosis. For instance, Japan has a number of designated AIDS Core Hospitals,

350   and HIV diagnoses in those institutes involve screening of common opportunistic infections

351   upon diagnosis of HIV infection, which sensitively leads to the diagnosis of AIDS. Compared

352   with HIV-infected individuals diagnosed at local health centers, the frequency of AIDS diagnosis

353   may be higher in the designated hospitals, calling for the validation of estimates using other

354   methods. Using other datasets including biomarkers or CD4 count data can act as another

355   potential work to be built on this study. Third, in the present study, we struggled with subtraction

356   of AIDS deaths from the calculation of the diagnosed proportion of infections; this problem

357   essentially stems from the absence of a case registration system in Japan. Once diagnosed,

358   infected individuals are never longitudinally monitored by the government, considerably

359   complicating prevalence estimation. With a registration system of HIV-infected individuals,

360   statistical monitoring of the second and third goals of 90-90-90 can be achieved in real time and

361   in a very reasonable manner.

362   Five technical limitations must be noted. First, the present study did not account for

363   uncertainties other than variations in length of the incubation period. There has been a concern

364   that the incubation period has probably shortened over time (Nakamura et al., 2011), but I did

365   not have substantial data to support this issue. Second, the natural history of HIV infection has

366 yet to be explored in-depth; an explicit proportion of HIV individuals who never develop AIDS

367 over the course of infection is missing. Third, other than sex, the present study accepted

368 homogeneity in the natural course and diagnosis of infection. Our future studies will address

369 several heterogeneities. Fourth, estimates rested on yearly data, and the precision was limited

370 (e.g., with use of the step function for every 4 years). The use of smoothing with nonparametric

371 back-projection is another of our ongoing studies (Becker, 1997). Fifth, the present study focused

372 on the incidence estimation, and more explicit modeling of the transmission dynamics, including

373 those highlighting the role of MSM, is the subject for future studies.

374   Despite these limitations, the present study successfully estimated the incidence of HIV

375 infections, undiagnosed number of infections, and the proportion diagnosed in real time, using

376 limited but readily available epidemiological surveillance data. Improved estimates using age

377 and geographical data, as well as estimates based on other methods, are to follow, which will

378 boost studies of epidemiological estimation in this area in Japan.

379 **Conclusions**

380 In the present study, a statistical modeling method was developed for the estimation of HIV

381 incidence in Japan and estimates made of the undiagnosed number of HIV infections and the

382 proportion of diagnosed HIV infections over time. Using the McKendrick equation, a general

383 convolution-like equation was derived, allowing for joint estimation of the HIV incidence and

384 time-dependent rate of diagnosis. By the end of 2017, the cumulative number of HIV infections

385 was estimated to be about 30,000, and about 80% of infections have ever been diagnosed.

386 Accounting for the uncertainty in the median incubation period ranging from 7.5 to 12.3 years,

387 estimates of the diagnosed proportion of HIV infections can range from 77% to 84%. The

388 proportion of diagnosed HIV infections appears not to have reached 90% among Japanese

389 nationals.

390

391 **Acknowledgment**

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

393

394 **References**

395 Althaus C. 2015. Ebola superspreading. *Lancet Infect Dis* 15:507-508.

396 Becker NG. 1997. Uses of the EM algorithm in the analysis of data on HIV/AIDS and other

397 infectious diseases. *Statistical Methods in Medical Research* 6:24-37.

398 Boldson JL, Jensen JL, Sogarrd J, Sorensen M. 1988. On the incubation time distribution and

399 the Danish AIDS data. *Journal of the Royal Statistical Society, Series A* 151:42-43.

400 Brookmeyer R, Gail MH. 1986. Minimum size of the acquired immunodeficiency syndrome

401 (AIDS) epidemic in the United States. *Lancet* 2:1320-1322.

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

403 Oxford University Press.

404 Brookmeyer R, Goedert JJ. 1989. Censoring in an epidemic with an application to

405 haemophilia-associated AIDS. *Biometrics* 45:325-335.

406 Committee of AIDS Trends, Ministry of Health, Labor and Welfare, Japan. 2018. Annual

407 Report of the Trend of HIV and AIDS in Japan. Tokyo: Ministry of Health, Labor and

408 Welfare. *Available at: http://api-net.jfap.or.jp/status/* (accessed 10 October 2018).

409   Cuadros DF, Abu-Raddad LJ. 2016. Geographical Patterns of HIV Sero-Discordancy in High

410   HIV Prevalence Countries in Sub-Saharan Africa. *International Journal of*

411   *Environmental Research and Public Health* 13:865. DOI: 10.3390/ijerph13090865.

412   Cui JS, Becker NG. 2000. Estimating HIV incidence using dates of both HIV and AIDS

413   diagnoses. *Statistics in Medicine* 19:1165-1177.

414   Donnelly CA, Cox DR. 2001. Mathematical biology and medical statistics: contributions to

415   the understanding of AIDS epidemiology. *Statistical Methods in Medical Research*

416   10:141-154.

417   Ejima K, Aihara K, Nishiura H. 2014. Probabilistic differential diagnosis of Middle East

418   respiratory syndrome (MERS) using the time from immigration to illness onset among

419   imported cases. *Journal of Theoretical Biology* 346:47-53.

420   Gail MH, Brookmeyer R. 1988. Methods for projecting course of acquired

421   immunodeficiency syndrome epidemic. *Journal of the National Cancer Institute* 80:900-

422   911.

423   Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. 2009. Universal voluntary HIV

424   testing with immediate antiretroviral therapy as a strategy for elimination of HIV

425   transmission: a mathematical model. *Lancet* 373:48-57.

426   Granich R, Williams B, Montaner J, Zuniga JM. 2017. 90-90-90 and ending AIDS: necessary

427   and feasible. *Lancet* 390:341-343.

428   Heesterbeek JAP, Metz J. 1993. The saturating contact rate in marriage-and epidemic

429   models. *Journal of Mathematical Biology* 31:529-539.

430   Hsieh YH, Lin PC. 2016. Current Trends and Future Projection of HIV/AIDS Epidemic in

431   Taiwan: A Modeling Analysis. *Current HIV Research* 14:138-147.

432   Hsieh YH, Ruan Y, Chen CW, Shi W, Li D, Luo F, Shao Y. 2012. HIV prevalence and

433   underreporting of men who have sex with men in Beijing. *International Journal of STD*

434   *& AIDS* 23:606-607.

435   Inoue Y, Takaku Y, Yajima T, Ikushima Y. 2015. A survey comparing the route of HIV

436   transmission reported by patients to their physicians and the self-reported "actual route".

437   *Nihon Koshu Eisei Zasshi* 62:106-116 (in Japanese).

438   Iwamoto A, Taira R, Yokomaku Y, Koibuchi T, Rahman M, Izumi Y, Tadokoro K. 2017.

439   The HIV care cascade: Japanese perspectives. *PLoS One* 12:e0174360.

440   Jewell NP, Dietz K, Farewell VT. 1992. *AIDS Epidemiology: Methodological Issues*. Berlin,

441   Birkhäuser.

442   Kihara M, Imai M, Shimizu M. 2000. Trends of HIV infections among blood donors.

443   *Infectious Agents Surveillance Report* 21:140-141 (in Japanese).

444   Kihara M, Ono-Kihara M, Feldman MD, Ichikawa S, Hashimoto S, Eboshida A, Yamamoto

445   T, Kamakura M. 2003. HIV/AIDS surveillance in Japan, 1984-2000. *Journal of Acquired*

446   *Immune Deficiency Syndromes* 32:S55-S62.

447   Kretzschmar ME, Schim van der Loeff MF, Birrell PJ, De Angelis D, Coutinho RA. 2013.

448   Prospects of elimination of HIV with test-and-treat strategy. *Proceedings of the National*

449   *Academy of Sciences of the United States of America* 110:15538-15543.

450   Marschner IC. 1994. Using time of first positive HIV test and other auxiliary data in

451   back‐projection of AIDS incidence. *Statistics in Medicine* 13:1959-1974.

452   Mumtaz GR, Awad SF, Feizzadeh A, Weiss HA, Abu-Raddad LJ. 2018. HIV incidence

453   among people who inject drugs in the Middle East and North Africa: mathematical

454   modelling analysis. *Journal of the International AIDS Society* 21:e25102.

455   Muñoz A, Sabin CA, Phillips AN. 1997. The incubation period of AIDS. *AIDS* 11:S69-S76.

456   Munoz A, Xu J. 1996. Models for the incubation of AIDS and variations according to age

457   and period. *Statistics in Medicine* 15:2459-2473.

458   Nakamura H, Teruya K, Takano M, Tsukada K, Tanuma J, Yazaki H, Honda H, Honda M,

459   Gatanaga H, Kikuchi Y, Oka S. 2011. Clinical Symptoms and Courses of Primary HIV-1

460   Infection in Recent Years in Japan. *Internal Medicine* 50:95-101.

461   Nemoto T. 2004. HIV/AIDS surveillance and prevention studies in Japan: summary and

462   recommendations. *AIDS Education and Prevention* 16:27-42.

463   Nishiura H. 2007. Lessons from previous predictions of HIV/AIDS in the United States and

464   Japan: epidemiologic models and policy formulation. *Epidemiologic Perspectives and*

465   *Innovations* 4:3.

466   Nishiura H, Inaba H. 2011. Estimation of the incubation period of influenza A (H1N1-2009)

467   among imported cases: addressing censoring using outbreak data at the origin of

468   importation. *Journal of Theoretical Biology* 272:123-130.

469   R Core Team. 2016. The R Project for Statistical Computing. Vienna, Austria. *Available at:*

470   *https://www.R-project.org/* (accessed 11 October 2018).

471   Riou J, Poletto C, Boelle PY. 2018. Improving early epidemiological assessment of emerging

472   Aedes-transmitted epidemics using historical data. *PLoS Negl Trop Dis* 12:e0006526.

473   Sherriff N, Koerner J, Kaneko N, Shiono S, Takaku M, Boseley R, Ichikawa S. 2017.

474   Everywhere in Japan: an international approach to working with commercial gay

475   businesses in HIV prevention. *Health Promot Int.* 32(3):522-534.

476   UNAIDS. 2014. 90-90-90: an ambitious treatment target to help end the AIDS epidemic.

477   UNAIDS: Geneva. *Available at:*

478   *http://www.unaids.org/en/resources/documents/2014/90-90-90* (accessed 11 October

479   2018).

480   van Sighem A, Pharris A, Quinten C, Noori T, Amato-Gauci AJ; The Ecdc HIv/Aids

481   Surveillance And Dublin Declaration Monitoring Networks. 2017. Reduction in

482   undiagnosed HIV infection in the European Union/European Economic Area, 2012 to

483   2016. *Eurosurveillance* 22: doi: 10.2807/1560-7917.ES.2017.22.48.17-00771.

484

485

486   **Figure legends**

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

488   New HIV infections occur at rate $\lambda(t)$. While going undiagnosed as $h(t,\tau)$, there would be an

489   increase in the time since infection τ. Diagnosis of HIV takes place at a time-dependent rate $\alpha(t)$,

490   and AIDS illness onset occurs at rate $\rho(\tau)$, which depends on the time since infection. Newly

491   diagnosed HIV infections, and AIDS cases that had not been previously diagnosed with HIV,

492   were notified to the surveillance system.

493

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

495   A. The yearly incidence of HIV infection, assuming that the median incubation period is 10.0

496   years. The step function for every 4 years was used to model the incidence. The 95% confidence

497   intervals were derived from profile likelihood. B. The yearly rate of diagnosis of HIV infection,

498   assuming that the median incubation period is 10.0 years. C. Maximum likelihood estimates of

499   the yearly incidence with different median incubation periods: 7.5, 10.0, and 12.3 years. D.

500   Maximum likelihood estimates of the yearly rate of diagnosis with different median incubation

501   periods: 7.5, 10.0, and 12.3 years. E. Yearly incidence estimates by sex and different median

502   incubation periods. Maximum likelihood estimates are shown. Note that a common logarithmic

503   scale is used on the vertical axis, to ease comparisons. F. Yearly rate of diagnosis estimates by

504   sex using different median incubation periods. Unfilled marks are for female, while filled marks

505   are for male. Maximum likelihood estimates are shown.

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

507   A. Comparisons between observed and predicted yearly number of HIV diagnoses and AIDS

508   cases. Different median incubation periods (i.e., 7.5, 10.0, and 12.3 years) were assumed, but

509   predicted values are mostly overlapped. B. Comparisons between observed and predicted values

510   by sex. Circles represent the observed number of HIV diagnoses whereas triangles represent that

511   of AIDS cases. Solid marks represent males; empty marks represent females. A common

512   logarithmic scale is used on the vertical axis. In both panels, bold grey lines represent lower and

513   upper 95% confidence intervals with the median incubation period of 10.0 years based on the

514   parametric bootstrap method.

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

516   A. Estimates of undiagnosed HIV infections, assuming that the median incubation period is 10.0

517   years. The 95% confidence intervals were derived from profile likelihood. B. Maximum

518   likelihood estimates of undiagnosed HIV infections with different median incubation periods:

519   7.5, 10.0, and 12.3 years. C. Proportion of diagnosed infections out of the cumulative number of

520   HIV infections, inclusive of AIDS cases. D. Proportion of diagnosed infections out of the

521   cumulative number of HIV infections, excluding AIDS cases. E. Maximum likelihood estimates

522   of undiagnosed HIV infections by sex, with different median incubation periods: 7.5, 10.0, and

523   12.3 years. Note that common logarithmic scale is used on the vertical axis. D. Proportion of

524   diagnosed infections out of the cumulative number of HIV infections, excluding AIDS cases, by

525   sex.

526   **Figure 5. Estimated undiagnosed HIV infections and proportion of diagnosed infections at**

527   **the end of 2017.**

528   A. Estimates of undiagnosed HIV infections with different incubation periods. Whiskers extend

529   to lower and upper 95% confidence intervals derived using a parametric bootstrapping method.

530   B. Proportion of diagnosed infections out of the cumulative number of HIV infections, excluding

531   AIDS cases (solid circles) or including AIDS cases but subtracting 2,321 deaths (empty circles).

532   Whiskers extend to lower and upper 95% confidence intervals derived using a parametric

533   bootstrapping method.

534

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.

ript to be reviewed
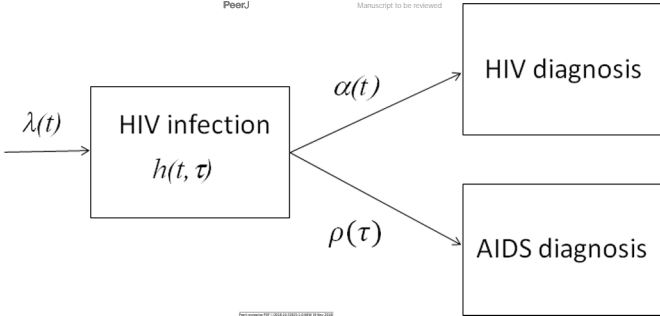
Figure 2. Estimated HIV incidence and rate of diagnosis in Japan.

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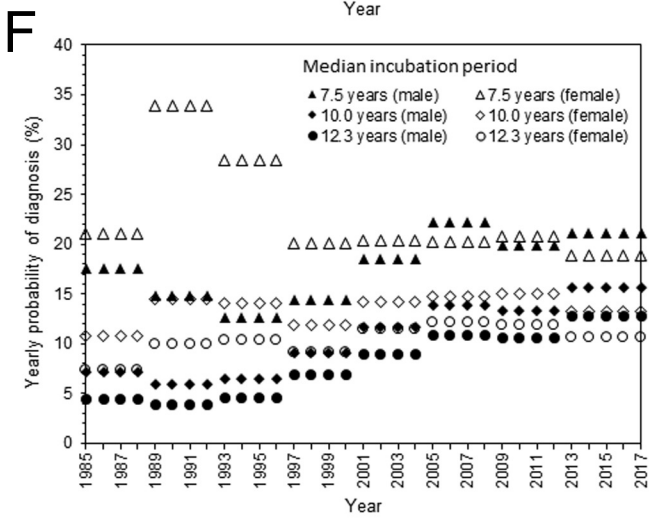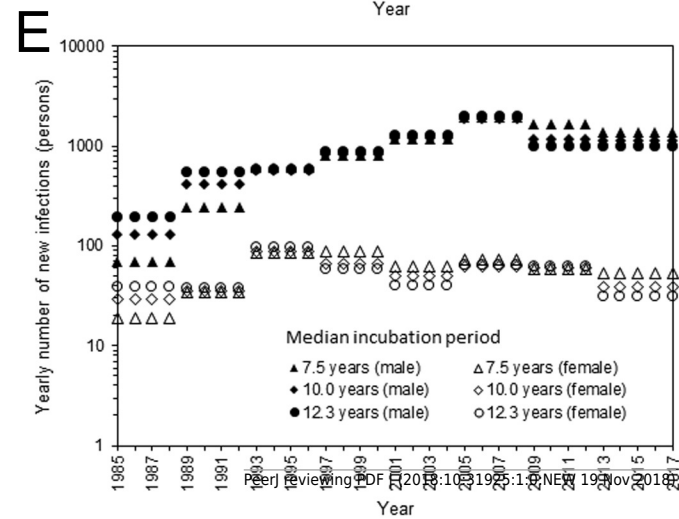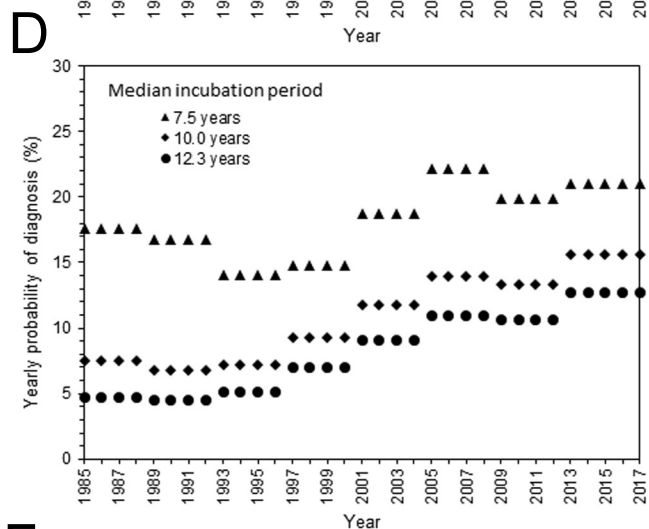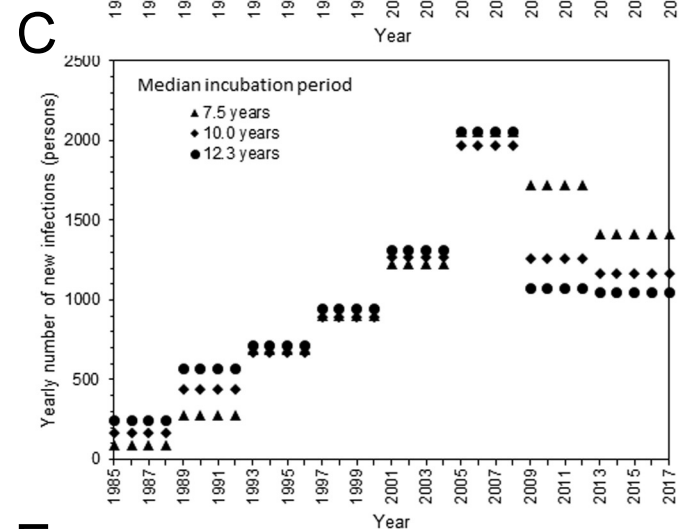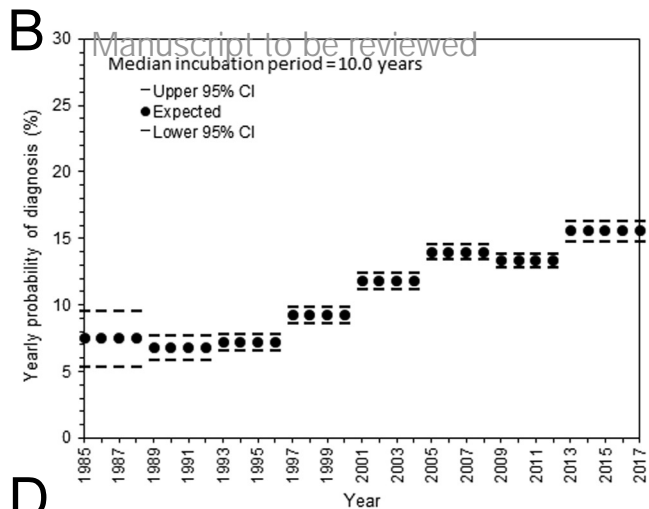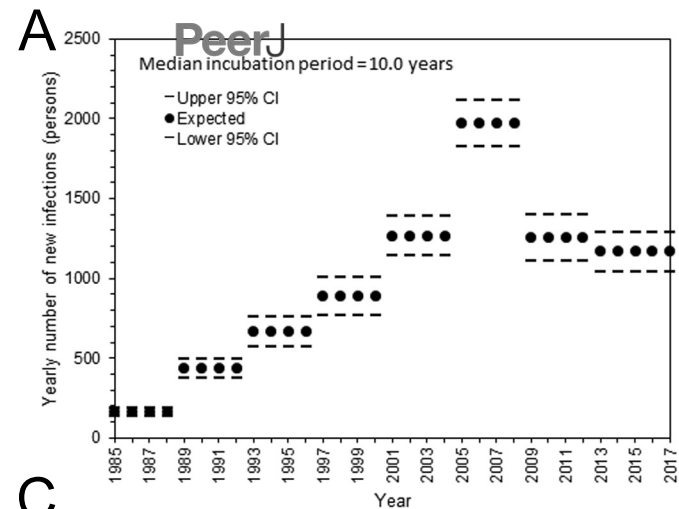
Figure 3. HIV diagnoses and AIDS cases in Japan, 1985–2017.

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. In both panels, bold grey lines represent lower and upper 95% confidence intervals with the median incubation period of 10.0 years based on the parametric bootstrap method.
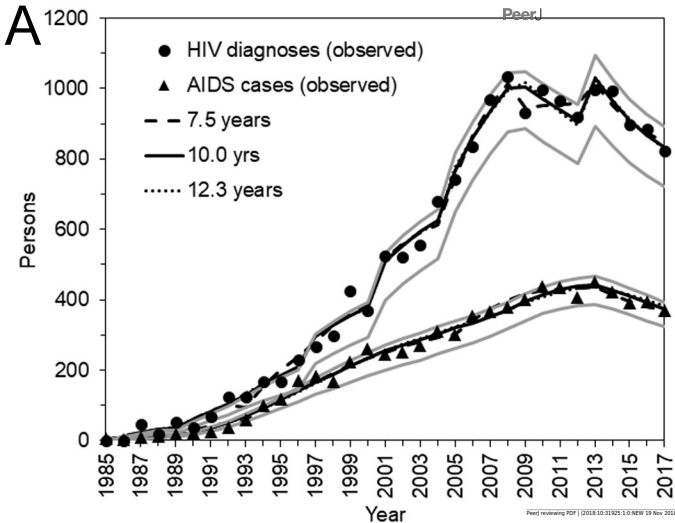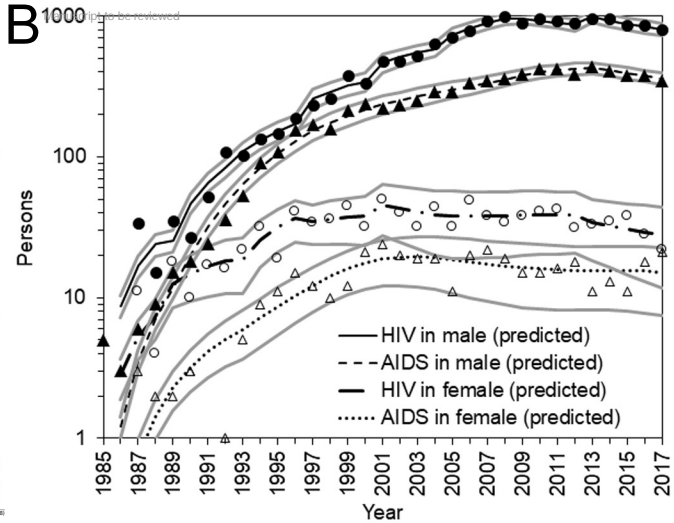
Figure 4. Undiagnosed number and proportion of HIV infections in Japan, 1986–2017.

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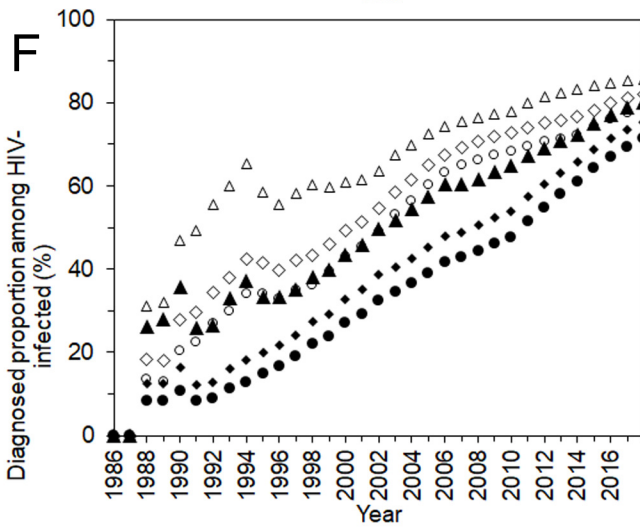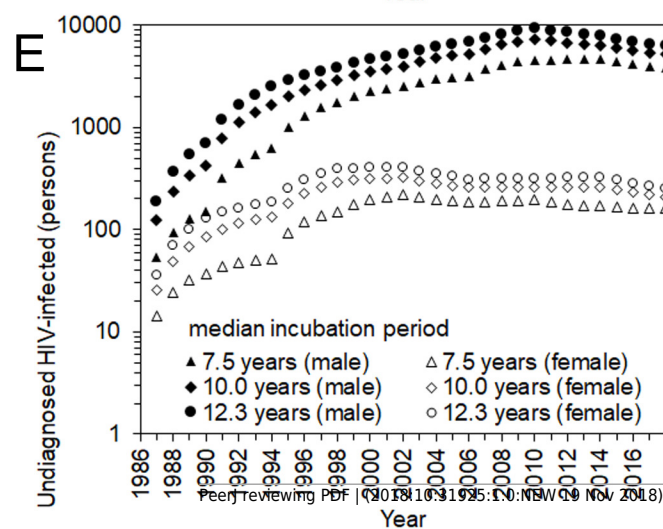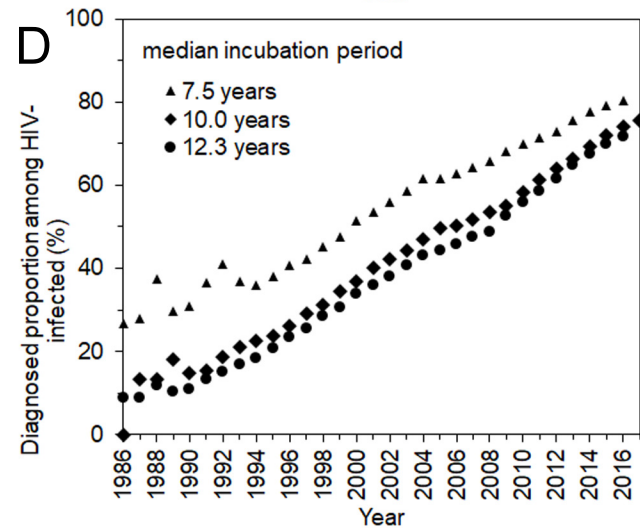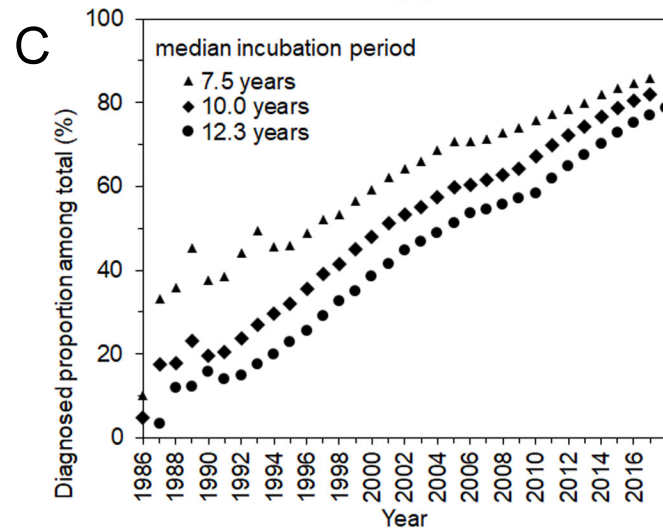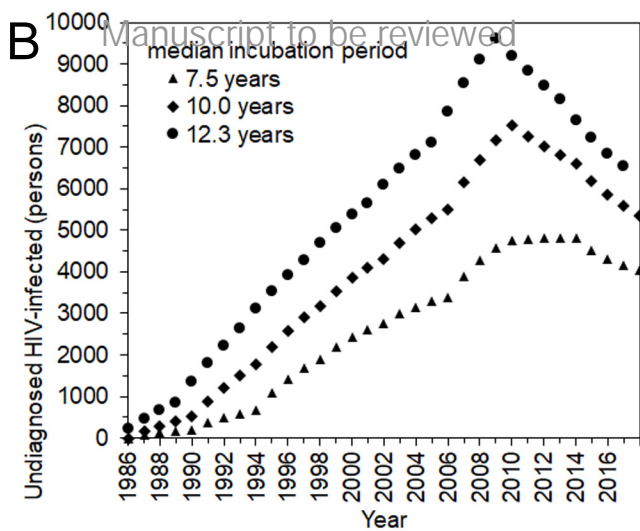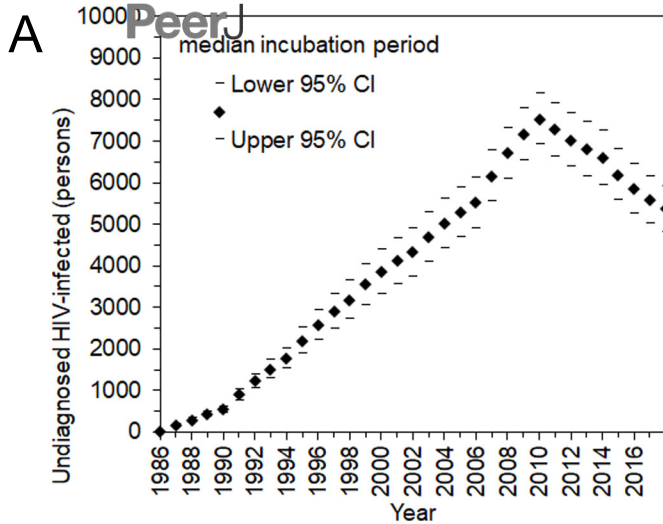
Figure 5. Estimated undiagnosed HIV infections and proportion of diagnosed infections at the end of 2017.

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.