A New Method for Estimating the Incidence of Infectious Diseases

Hamish McManus, Denton Callander, Jason Asselin, James McMahon, Jennifer F. Hoy, David J. Templeton, Christopher K. Fairley, Basil Donovan, Alisa E. Pedrana, Phillip Keen, David P. Wilson, Julian Elliott, John Kaldor, Siaw-Teng Liaw, Kathy Petoumenos, Martin Holt, Margaret E. Hellard, Andrew E. Grulich, Andrew Carr, Mark A. Stoove, and Rebecca J. Guy

Correspondence to Dr. Hamish McManus, Kirby Institute, Level 6, Wallace Wurth Building, UNSW Sydney, NSW Australia 2052 (e-mail: hmcmanus@kirby.unsw.edu.au)

Author affiliations: The Kirby Institute, UNSW Sydney, Sydney, NSW, Australia (Hamish McManus, Denton Callander, David J. Templeton, Basil Donovan, Phillip Keen, David P. Wilson, John Kaldor, Kathy Petoumenos, Andrew E. Grulich, Rebecca J. Guy); Mailman School of Public Health, Columbia University, New York NY (Denton Callander); Disease Elimination Program, The Burnet Institute, Melbourne, VIC, Australia (Jason Asselin, Alisa E. Pedrana, David P. Wilson, Margaret E. Hellard, Mark A. Stoove); Department of Infectious Diseases, The Alfred Hospital and Monash University, Melbourne, VIC, Australia (James McMahon, Jennifer F. Hoy, Julian Elliott, Margaret E. Hellard); Department of Sexual Health Medicine, Sydney Local Health District, Camperdown, NSW, Australia (David J. Templeton); Discipline of Medicine, Central Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia (David J. Templeton); Melbourne Sexual Health Centre, Alfred Health Melbourne, VIC, Australia (Christopher K. Fairley); Central Clinical School, Monash University, Melbourne, VIC, Australia (Christopher K. Fairley); Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia (Alisa E. Pedrana); School of Population Health, UNSW Sydney, Sydney, NSW, Australia (Siaw-Teng Liaw); Centre for Social Research in Health, UNSW Sydney, Sydney, NSW, Australia (Martin Holt); Centre for Applied Medical Research, St Vincent's Hospital, Sydney, NSW, Australia (Andrew Carr)

TAIPAN is funded through a peer-reviewed project grant from Australia's National Health and Medical Research Centre (DPAPP1082336).

Conflict of interest: none declared

Running head: Estimating the Incidence of Infectious Diseases

ABSTRACT

Ambitious World Health Organization targets for disease elimination require monitoring of epidemics using routine health data in settings of decreasing and low incidence. We evaluated two methods commonly applied to routine testing results to estimate incidence rates that assume uniform probability of infection between consecutive negative and positive tests based on: 1. the midpoint of this interval; and 2. a randomly selected point on this interval. We compared these with an approximation to the Poisson-binomial distribution which assigns partial incidence to time-periods based on the uniform probability of occurrence in these intervals. We assessed bias, variance and convergence of estimates using simulations of Weibull distributed failure times with systematically varied baseline incidence, and varying trend. We considered results for quarterly, half-yearly and yearly incidence estimation frequencies. We applied methods to assess human immunodeficiency virus (HIV) incidence in HIV-negative patients from the Treatment with Antiretrovirals and their Impact on Positive And Negative men study between 2012 and 2018. The Poisson-binomial method had reduced bias and variance at low levels of incidence and for increased estimation frequency, with increased consistency of estimation. Application of methods to real-world assessment of HIV incidence found decreased variance in Poisson-binomial model estimates, with observed incidence declining to levels where simulation results had indicated bias in midpoint and random-point methods.

KEY WORDS

disease incidence rates, routine diagnostic testing, Poisson-binomial distribution, longitudinal data

List of Abbreviations

HIV	human immunodeficiency virus
MSM	men who have sex with men
PrEP	pre-exposure prophylaxis
PY	person years
STI	sexually transmissible infection
WHO	World Health Organization

Recent advances in the treatment and prevention of many infectious diseases have led to ambitious World Health Organization (WHO) targets which include ending the global human immunodeficiency virus (HIV)/ AIDS (1, 2) and tuberculosis (3, 4) epidemics, interrupting

malaria (5) and measles transmission (6) , eliminating mother-to-child-transmission of HIV (7) and syphilis (7) , and eliminating hepatitis-B and C as a public health threat. (8, 9) In low incidence settings, ambitious target incidence thresholds have been set as low as <0.1 per 100 000 population. (10, 11) This has created increased programmatic interest in more frequently measured incidence estimates. In Australia, there is an ambitious goal to virtually eliminate HIV transmission by 2022, with HIV pre-exposure prophylaxis (PrEP) one of the key strategies. (12) This requires frequent evaluation of incidence to assess the response and make necessary adjustments to PrEP and/or other strategies implemented as required. (13) However more frequent incidence estimation and reporting, such as at quarterly intervals for PrEP, leads to smaller increments in incident case numbers compared to annual reporting time frames, which presents analytical challenges and necessitates a review of current methodologies.

The gold standard method of measuring disease incidence requires repeat testing at prespecified frequent and regular intervals, typically in the context of clinical trials or prospective cohort studies. (13) However, such studies are very costly and thus have not been widely adopted, especially for population level surveillance and assessment.(14) Many high-income countries rely on passive notification (otherwise known as case reporting) of diagnoses to assess trends in disease epidemiology, but these data do not reflect trends in incidence because they include both recent and established infections (2) and do not take account of changes in diagnostic testing frequency. (15)

Instead, the WHO/UNAIDS has recommended greater use of routinely collected health facility data to evaluate epidemics. WHO is actively developing health statistics and information systems toolkits designed to optimize the analysis and use of routine data available through health management information systems; including for HIV, hepatitis, malaria, tuberculosis, and

neglected tropical diseases.(13, 16) This is increasingly feasible as more low- and middle-income countries adopt comparable electronic data management systems to those used in high-income countries, making input data more accessible for public health surveillance. (17-19) (20) Applications of such systems to disease incidence measurement and surveillance have been reported for at-risk populations including female sex workers, people who inject drugs and men who have sex with men (MSM) in both high- and low-middle income countries. (20-25)

Therefore, we assessed the robustness of widely used incidence estimation methods (the midpoint method and the random-point method) based on routinely collected testing data with unknown date of infection, and compared these with a new method which uses Poisson-binomial distribution assumptions about the timing of infection. We hypothesis that the number of incident cases gets smaller the more incidence is measured, leading to increases in the variance around estimates for all the methods, but less so for the Poisson-binomial than the others. We firstly assessed the different methodologies using generalised simulated populations with systematically varied levels and trends in disease incidence to compare variance, bias and consistency. We then applied these methodologies to a real-world HIV-negative cohort in Australia to assess HIV incidence among MSM over a period of expanded HIV prevention program coverage and newly introduced prevention methods, such as treatment as prevention and PrEP. This setting provided an example of the emerging challenges associated with accurate and timely incidence estimations in settings of reducing numbers of detectable cases. (26)

METHODS

We applied 3 different methods to estimate the date of incidence between consecutive negative and positive HIV antibody tests (Figure 1):

1. Midpoint method: We assumed infections occurred at the midpoints (expected value for the uniform distribution) of negative-positive testing intervals.
2. Random point method: We assumed infections occurred at random points uniformly distributed on negative-positive testing intervals
3. Poisson-binomial method: We assumed an equal probability of infection at each time point within negative-positive testing intervals (equal to the probability density function of the uniform distribution). This method assigns partial incidence to different follow-up periods in instances where negative-positive test intervals fall across multiple follow-up periods. The distribution of the total number of new infections at any given time point was then approximated using a Poisson distribution, with expected incident case numbers equal to the sum of individual probabilities of seroconversion. For large sample size, but small and not necessarily equal incidence probability, the total number of incident cases is well approximated by a Poisson distribution by the Law of Small numbers. (27) Time at risk in subintervals between negative and positive tests was reduced by 50% to reflect expected time of failure for each interval. Where presented, we developed 95% confidence bounds for estimates based on the incomplete gamma function. (28)

Study data

Simulated populations.

Using the Stata 15.1 (Stata Corporation, College Station, Texas, USA) "survsim.ado" program we simulated survival times in populations with incident infection using assumed Weibull distributions of estimated positive test times for a population size of 10,000 over 6 years of follow-up. A quasi-normal distribution of test interval durations was assumed for intervals

between last negative test and first positive test, with average interval duration of 0.5 years (σ^2=0.25), but with minimum interval length restricted to positive durations. Analysis time was restricted to the last 5 years of follow-up, with left censoring applied to the first year of follow-up to reduce effects of incomplete follow-up associated with negative test dates occurring prior to start of the analysis.

Decreasing and constant incidence rate scenarios were simulated by variation of the Weibull shape parameter (η=0.5, 1.0), while different baseline incidence levels were simulated by variation of the Weibull scale parameter (ρ=0.001, 0.010, and 0.100). Scenarios were simulated using different incidence estimation frequency: 4-times per year (quarterly), twice per year (half-yearly), and once per year (yearly).

The effect of testing frequency (quarterly, half-yearly, and yearly) on incidence estimates was estimated for constant, low incidence rate scenarios (η=1, ρ=0.001).

Real world HIV negative population: TAIPAN study data.

We used data from the Treatment with Antiretrovirals and their Impact on Positive And Negative men (TAIPAN) study of HIV incidence in MSM to examine the utility of models applied to real world data. TAIPAN is a retrospective, longitudinal, open cohort study created using routine HIV testing data. (29) Inclusion in the cohort was based on having more than 1 HIV test recorded within the participating network of clinical services during the study period 1 January 2011 to 30 June 2019, with the first test being HIV-negative. MSM status was based on recorded sexual behaviour or male patients and records indicating testing for a rectal sexually transmissible infection (STI). (30) Study entry date was the date of first negative HIV test in the study period based on screening using 4th generation HIV immunoassays. Study exit was set as

the last negative test (if no positive tests) or first positive test during the study period. Incident cases were assigned to the time period between last HIV negative test and first HIV positive test.

Data were extracted from 52 clinics (comprised of primary health care, sexual health and hospital clinics) in the 2 most populous states in Australia (New South Wales and Victoria), where an estimated two-thirds (64%) of MSM reside and where over 60% of national annual HIV diagnoses are reported. (31, 32) The clinics were participating in a sentinel surveillance program called the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance of Sexually Transmissible Infections and Blood Borne Viruses (ACCESS). (33) These data are recorded in health management information systems as part of routine care and extracted and transferred regularly to research institutes via GRHANITE™ data extraction software (GRHANITE™ Health Informatics Unit, University of Melbourne, Victoria, Australia) which creates an anonymised unique identifier for everyone. (34) The anonymised unique identifier allows individuals' medical records from multiple clinics and laboratories to be reliably linked. (34, 35) The analysis was conducted using data from the period 1 January 2011 to 30 June 2019 to estimate HIV incidence over the period 1 January 2012 to 31 December 2018, encompassing a period before and during the scale up of coverage of treatment as prevention (from 2012) and, more recently, PrEP in Australia (from 2016). (36-39)

Ethics approval

Ethics approval for the ACCESS was granted by the human research ethics committees of St Vincent's Hospital, Sydney (reference, 08/51), the University of New South Wales (reference, 8076) and the Alfred Hospital (ref 248/17); site-specific assessment approvals were obtained from local research governance offices as required.

Statistical Analysis

Simulations were conducted using systematically varied estimation frequency, trend and incidence levels described above, and incidence rates then calculated as cases per 100 person years (PY) for the 3 methods. Monte Carlo simulation of incidence rates for respective methods (median, 2.5% and 97.5% percentile) was conducted using the Stata "simulate.ado" program based on 10,000 repetitions per model. Relevant Stata code is provided in appendices (Web Appendices 1 and 2).

For real world data we compared estimates for the 3 methods to sampling (10,000 repetitions) from simulated uniformly distributed failure times across testing intervals for 2012-18 based on follow-up testing censored at 30-June-2019. This allowed estimates to more closely approximate a true uniformly distributed set of failure times based on extended follow-up compared to model estimates.

RESULTS

Simulation disease incidence

Constant incidence rate.

For simulated incidence rates that were constant over time, estimates showed higher variance when estimation frequency was higher or incidence rates were lower (Figure 2, Web Figure 1-2). Variance in the Poisson binomial model was lower than that observed in midpoint and random-point models for half-yearly and quarterly estimation frequency and for lower incidence rates. Models showed consistency with simulated distributions except at lowest incidence rates and quarterly estimation frequencies where the midpoint method and random-point method showed

negative bias leading to underestimation of incidence, while the Poisson-binomial method more closely approximated simulated incidence.

Decreasing incidence rate

For simulated incidence rates that were decreasing over time, estimates showed higher variance when estimation frequency was higher or incidence rates were lower, with lower variance in the Poisson binomial model than other models (Figure 3, Web Figure 3-4). At low rates the midpoint and random-point models showed some departure from simulated rates while the Poisson binomial method was relatively consistent for all methods, although at lowest levels of incidence and higher estimation frequencies there was indication of negative-bias.

Testing frequency

For simulated variation in testing frequency, estimates were constant across different testing rates. The Poisson binomial method estimates showed increased variance at increased testing frequencies and bounds on estimates tended towards those of the midpoint and random-point methods at higer testing frequencies (Web Figure 5).

Real world HIV incidence: TAIPAN study data.

Population characteristics

HIV incidence estimates in the TAIPAN study were based on 46,604 MSM and included 885 HIV incident cases (1.90% of patients) over the course of the analysis period 2012-2018 inclusive. The median time between repeat tests was 105 days (interquartile range [IQR] 71-217) with longer time between repeat tests for participants who subsequently became HIV positive (227 days (IQR 109-566)). Participants tested a mean of 2.0 times per year (SD 1.34) increasing

from 1.5 times per year in 2012 to 2.5 times per year in 2018. This increase was less evident in participants who subsequently became positive with a change from 1.73 times per year to 1.50 times per year (Table 1).

HIV incidence estimates

Estimated incidence rates trended downwards across the 7-year duration of the study from over 1/100PY to close to 0.10/100PY (Figure 4, Web Table 1-3). There was increased variance in the estimates associated with the midpoint and random-point methods compared to the Poisson-binomial method and this increased as estimation frequency increased (Web Table 1-3). For quarterly estimation frequencies, the range in incidence rates over the interval 2012-2018 for the Poisson-binomial method was 1.18 to 0.23 cases/100 PY, for the midpoint method it was 1.26 to 0.22 cases/100 PY, and for the random-point method it was 1.48 to 0.19 cases/100 PY (Figure 4, Web Table 1-3) .

DISCUSSION

In settings where disease incidence is low and decreasing, we found that repeated testing methods (midpoint, random point and the new Poisson-binomial method) produced similar annual estimates of HIV incidence in MSM. However, the Poisson-binomial method more closely approximated a uniform distribution of disease transmission which was more robust to increases in incidence estimation frequency, with more continuous distributions of risk and lower variance and bias than the other repeat testing methods evaluated. This method is therefore, likely to be more useful in evaluating epidemics where transmission is low and decreasing, or when reporting frequency increases, such as quarterly. The programatic monitoring of progress

towards WHO disease elimination targets, including for global epidemics such as HIV and Hepatitis C virus, is therefore likely to benefit from application of this method.

Methods compared in this paper all assigned the date of incidence based on an assumed uniform distribution across the interval between the last negative and first positive test. However, the midpoint method and random point methods allocate incident cases to a single point of time (respectively the midpoint or a random point on negative-positive testing intervals), which does not reflect uncertainty about the precise incidence date. Instead the Poisson-binomial method allocates time weighted proportions of incident cases to more than 1 unit of time when testing intervals overlap multiple time periods. While all methods were found to be accurate when incidence rates or time units were increased, the Poisson-binomial method was more robust when numbers of incident cases were low and/or when there was more frequent estimation. This reflects an advantageous property of the Poisson-binomial distribution in a setting of decreasing disease incidence. In particular, for small and non-identical incidence rates it is closely approximated by the Poisson distribution.(40) Another useful property of the Poisson-binomial method is that this approximation has maximum and calculable bounds on error which allow the quality of estimates to be easily quantified. (41)

The application of estimation methods to a real world cohort monitoring testing for HIV showed similar broad trends by all methods but with less variance between consecutive time periods for the Poisson-binomial method, especially when estimation frequency increased. We presented data for quarterly time intervals which reflects commonly used HIV testing intervals for PrEP. There was more frequent testing in more recent periods for this cohort associated with increased availability of PrEP which requires quarterly HIV testing. Since 2016 PrEP was available through state government funded PrEP demonstration programs, and in 2018, nationally via the

pharmaceutical benefits scheme. From 2016 to 2018, self-reported PEP use increased from 4.6% to 20.9% (42) This trend was associated with reduced differences between estimates according to methods when incidence rates had fallen to around 0.3/100 PY. Nevertheless, results from the Poisson-binomial distribution suggested clearer trends than other methods. As rates fall further we expect differences between methods to become more apparent as reflected by simulations using baseline rates of 0.1/100 PY where the Poisson-binomial more closely approximated simulated distributions but other methods showed a negative bias associated with low numbers of incident cases.

There are some limitations to the method. First, it is possible that the Poisson-binomial model may produce underdispersed estimates as reflected by comparison with the midpoint and random point methods. However, low incidence settings mean that the Poisson mean and variance are closely alligned, and we observed a low rate of error using simulated distributions of risk. Second, our study does not make any assumptions about the distribution of seroconversion other than it lying with equal probability at any point between consecutive negative-positive tests. In instances of long duration between tests this may lead to substantial inaccuracy in probabilities associated with individual cases. It is also possible that patients might be more likely to test following high risk events upon recognising symptoms of primary HIV infection which would skew the distribution of seroconversions on the interval between tests. In such cases adaptation of the methods discussed in this paper to incorporate distributions other than the uniform distribution may be more appropriate, although we do not investigate that here. However considering the relatively high testing frequency in our real world HIV-cohort of men (over 2 tests per year) it is likely any such influence would be limited.

In summary, we found estimates of incidence calculated using the repeat testing methods based on the Poisson-binomial distribution were accurate and superior to those estimated by the midpoint method and random point method. This was mainly attributable to increased robustness at high estimation frequency, and to low case numbers. Given widespread uptake of routine HMIS data in program monitoring, the method seems well suited to evaluate incidence of diseases associated with elimination goals, such as for HIV PrEP programs, or for Hepatitis C virus direct-acting antiviral implementation.

ACKNOWLEDGMENTS

Author affiliations: The Kirby Institute, UNSW Sydney, Sydney, New South Wales, Australia (Hamish McManus, Denton Callander, David J. Templeton, Basil Donovan, Phillip Keen, David P. Wilson, John Kaldor, Kathy Petoumenos, Andrew E. Grulich, Rebecca J. Guy); Burnet Institute, Disease Elimination Program, Melbourne, Victoria, Australia (Jason Asselin, Alisa E. Pedrana, David P. Wilson, Margaret E. Hellard, Mark A. Stoove); Monash University Department of Medicine Alfred Hospital, Department of Infectious Diseases Melbourne, Victoria, Australia (James McMahon, Jennifer F. Hoy, Julian Elliott, Margaret E. Hellard); Royal Prince Alfred Hospital, Sydney Local Health District, Camperdown, New South Wales, Australia (David J. Templeton); The University of Sydney, Central Clinical School, Sydney, New South Wales, Australia (David J. Templeton); Melbourne Sexual Health Centre, Alfred Health Melbourne, Victoria, Australia (Christopher K. Fairley); Monash University, Central Clinical School Melbourne, Victoria, Australia (Christopher K. Fairley); Monash University, Department of Epidemiology and Preventive Medicine, Melbourne, Victoria, Australia (Alisa E. Pedrana); UNSW Sydney, School of Public Health and Community Medicine, Sydney, New

South Wales, Australia (Siaw-Teng Liaw); UNSW Sydney, Centre for Social Research in Health, Sydney, New South Wales, Australia (Martin Holt); St Vincent's Hospital, Centre for Applied Medical Research, Sydney, New South Wales, Australia (Andrew Carr).

TAIPAN is funded through a peer-reviewed project grant from Australia's National Health and Medical Research Centre (DPAPP1082336)

The authors acknowledge the help of Dr Tobias Vickers, Dr Carol El Hayek, and Lucy Watchirs Smith in recruiting study sites, implementing the study software, data management and data extraction and providing feedback. We also acknowledge statistical advice given by Dr Matthew G Law in development of study methods. We also acknowledge the contributions of Douglas Boyle and his team at HaBIC R2 (University of Melbourne) for their work in the design and development of the study software.

The authors declare that they have no competing interests.

REFERENCES

1. World Health Organization. Global health sector strategy on HIV 2016-2021. Towards ending AIDS. World Health Organization, 2016. (Available at:https://apps.who.int/iris/bitstream/handle/10665/246178/WHO-HIV-2016.05-eng.pdf). (Accessed August 15, 2020).
2. UNAIDS. Understanding fast-track: Accelerating action to end the AIDS epidemic by 2030. UNAIDS Geneva, 2015. (Available at:https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf). (Accessed August 15, 2020).
3. World Health Organization. Implementing the end TB strategy: the essentials. World Health Organization, 2015. (Available at:https://apps.who.int/iris/bitstream/handle/10665/206499/9789241509930_eng.pdf). (Accessed August 15, 2020).
4. World Health Assembly 67. Draft global strategy and targets for tuberculosis prevention, care and control after 2015: Report by the Secretariat. 2014. (Available at:https://apps.who.int/iris/bitstream/handle/10665/172989/B134_CONF4Rev1-en.pdf). (Accessed August 15, 2020).
5. World Health Organization. A framework for malaria elimination. World Health Organization, 2017. (Available at:https://apps.who.int/iris/bitstream/handle/10665/254761/9789241511988-eng.pdf). (Accessed October 20, 2020).
6. World Health Organization. Global measles and rubella strategic plan: 2012-2020. 2012. (Available at: https://apps.who.int/iris/bitstream/handle/10665/44855/9789241503396_eng.pdf?sequence=1&isAllowed=y). (Accessed April 26, 2020).
7. World Health Organization. Global guidance on criteria and processes for validation: elimination of mother-to-child transmission of HIV and syphilis. *Global guidance on criteria and processes for validation: elimination of mother-to-child transmission of HIV and syphilis*: World Health Organization, 2017. (Available at:https://www.who.int/reproductivehealth/publications/emtct-hiv-syphilis/en/). (Accessed August 15, 2020).
8. World Health Organization. Global health sector strategy on viral hepatitis 2016-2021. Towards ending viral hepatitis. World Health Organization, 2016. (Available at:https://apps.who.int/iris/bitstream/handle/10665/246177/WHO-HIV-2016.06-eng.pdf). (Accessed October 20, 2020).
9. World Health Organization. Technical considerations and case definitions to improve surveillance for viral hepatitis: technical report. World Health Organization, 2016. (Available at:https://apps.who.int/iris/bitstream/handle/10665/204501/9789241549547_eng.pdf). (Accessed October 20, 2020).
10. World Health Organization. Towards tuberculosis elimination: an action framework in low-incidence countries. World Health Organization, 2014. (Available at:https://apps.who.int/iris/bitstream/handle/10665/132231/9789241507707_eng.pdf). (Accessed October 20, 2020).

11. Department of Health. Eighth National HIV Strategy 2018-2022. Commonwealth of Australia Canberra, Australia, 2018. (Available at:https://www1.health.gov.au/internet/main/publishing.nsf/Content/ohp-bbvs-1/$File/HIV-Eight-Nat-Strategy-2018-22.pdf). (Accessed August 25, 2020).
12. Eighth National HIV Strategy 2018-2022. Commonwealt of Australia; 2018. (https://www1.health.gov.au/internet/main/publishing.nsf/Content/ohp-bbvs-1/$File/HIV-Eight-Nat-Strategy-2018-22.pdf). (Accessed October 20 2020).
13. HIV Incidence Measurement and Data Use. Boston, MA, USA: World Health Organization: WHO Working Group on HIV Incidence Measurement and Data Use, 2018. (Available at:https://www.who.int/diagnostics_laboratory/links/180622_boston_meeting_report.pdf?ua=1). (Accessed 30 June 2019).
14. Kassanjee R, De Angelis D, Farah M, et al. Cross-Sectional HIV Incidence Surveillance: A Benchmarking of Approaches for Estimating the 'Mean Duration of Recent Infection'. *Stat Commun Infect Dis* 2017;9(1).
15. Feigin A, El-Hayek C, Hellard M, et al. Increases in newly acquired HIV infections in Victoria, Australia: epidemiological evidence of successful prevention? *Sex Health* 2013;10(2):166-70.
16. WHO. Analysis and Use of Health Facility Data. World Health Organization; 2019. (https://www.who.int/healthinfo/tools_data_analysis_routine_facility/en/). (Accessed 20/11/2019 2019).
17. UNAIDS. Global AIDS Monitoring 2019 Indicators for monitoring the 2016 Political Declaration on Ending AIDS. UNAIDS; 2018. (http://www.unaids.org/en/resources/documents/2018/Global-AIDS-Monitoring). (Accessed 11 March 2019 2019).
18. World Health Organization. Consolidated strategic information guidelines for HIV in the health sector. 2015. (Available at:https://apps.who.int/iris/bitstream/handle/10665/164716/9789241508759_eng.pdf). (Accessed August 15, 2020).
19. Holmes C, Pillay Y, Mwango A, et al. Health systems implications of the 2013 WHO consolidated antiretroviral guidelines and strategies for successful implementation. *AIDS* 2014;28 Suppl 2(Suppl 2):S231-9.
20. Groseclose SL, Buckeridge DL. Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation. *Annu Rev Public Health* 2017;38(1):57-79.
21. van Griensven F, Holtz TH, Thienkrua W, et al. Temporal trends in HIV-1 incidence and risk behaviours in men who have sex with men in Bangkok, Thailand, 2006-13: an observational study. *Lancet HIV* 2015;2(2):e64-70.
22. Guy RJ, Spelman T, Stoove M, et al. Risk factors for HIV seroconversion in men who have sex with men in Victoria, Australia: results from a sentinel surveillance system. *Sex Health* 2011;8(3):319-29.
23. Bernstein KT, Marcus JL, Nieri G, et al. Rectal gonorrhea and chlamydia reinfection is associated with increased risk of HIV seroconversion. *J Acquir Immune Defic Syndr* 2010;53(4):537-43.
24. Annual Surveillance Report of HIV, viral hepatitis, STIs 2017. The Kirby Institute, UNSW Sydney, Sydney, NSW 2052, 2017. (Available

at:https://kirby.unsw.edu.au/report/annual-surveillance-report-hiv-viral-hepatitis-and-stis-australia-2017). (Accessed November 10, 2020).
25. Kwon JA, Iversen J, Maher L, et al. The impact of needle and syringe programs on HIV and HCV transmissions in injecting drug users in Australia: a model-based analysis. *J Acquir Immune Defic Syndr* 2009;51(4):462-9.
26. Grulich AE, Guy R, Amin J, et al. Population-level effectiveness of rapid, targeted, high-coverage roll-out of HIV pre-exposure prophylaxis in men who have sex with men: the EPIC-NSW prospective cohort study. *Lancet HIV* 2018;5(11):e629-e37.
27. Chen SX, Liu JS. Statistical Applications of the Poisson-Binomial and Conditional Bernoulli Distributions. *Statistica Sinica* 1997;7(4):875-92.
28. Scholz F. Confidence Bounds & Intervals for Parameters Relating to the Binomial, Negative Binomial, Poisson and Hypergeometric Distributions with Applications to Rare Events. In: Washington Uo, ed. University of Washington, 2019. (Available at:http://faculty.washington.edu/fscholz/DATAFILES498B2008/ConfidenceBounds.pdf).
29. Callander D, Stoove M, Carr A, et al. A longitudinal cohort study of HIV 'treatment as prevention' in gay, bisexual and other men who have sex with men: the Treatment with Antiretrovirals and their Impact on Positive And Negative men (TAIPAN) study protocol. *BMC Infect Dis* 2016;16(1):752.
30. Wilkinson A, El-Hayek C, Fairley CK, et al. Incidence and risk factors associated with chlamydia in men who have sex with men: a cohort analysis of Victorian Primary Care Network for Sentinel Surveillance data. *Sex Transm Infect* 2012;88(5):319-24.
31. Prestage G, Ferris J, Grierson J, et al. Homosexual men in Australia: population, distribution and HIV prevalence. *Sex Health* 2008;5(2):97-102.
32. HIV, viral hepatitis and sexually transmissible infections in Australia: Annual surveillance report 2018. The Kirby Institute, UNSW Sydney, Sydney, NSW 2052, 2018. (Available at:https://kirby.unsw.edu.au/report/hiv-viral-hepatitis-and-sexually-transmissible-infections-australia-annual-surveillance). (Accessed November 10, 2020).
33. Guy R, Lim MS, Wang YH, et al. A new surveillance system for monitoring HIV infection in Victoria, Australia. *Sex Health* 2007;4(3):195-9.
34. Callander D, Moreira C, El-Hayek C, et al. Monitoring the Control of Sexually Transmissible Infections and Blood-Borne Viruses: Protocol for the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance (ACCESS). *JMIR Res Protoc* 2018;7(11):e11028.
35. Boyle D, Kong F. A Systematic mechanism for the collection and interpretation of display format pathology test results from Australian primary care recordscords. *Electronic Journal of Health Informatics* 2011;6(2).
36. Zablotska IB, Selvey C, Guy R, et al. Expanded HIV pre-exposure prophylaxis (PrEP) implementation in communities in New South Wales, Australia (EPIC-NSW): design of an open label, single arm implementation trial. *BMC Public Health* 2018;18(1):210.
37. Holt M, Lea T, Mao L, et al. Community-level changes in condom use and uptake of HIV pre-exposure prophylaxis by gay and bisexual men in Melbourne and Sydney, Australia: results of repeated behavioural surveillance in 2013-17. *Lancet HIV* 2018;5(8):e448-e56.
38. NSW HIV Strategy 2012-2015 Quarter 3 2015 Data Report. NSW Ministry of Health, 2015. (Available at:https://www.health.nsw.gov.au/endinghiv/Publications/q3-2015-hiv-data-report.pdf). (Accessed November 10, 2020).

39. NSW HIV Strategy 2016-2020 Quarter 1 2018 Data Report. NSW Ministry of Health, 2018. (Available at:https://www.health.nsw.gov.au/endinghiv/Publications/q1-hiv-2018-hiv-data-report.pdf). (Accessed November 10, 2020).
40. Le Cam L. An Approximation Theorem for the Poisson Binomial Distribution. *Pacific Journal of Mathematics* 1960;10(4):1181-97.
41. Barbour AD, Holst L, Janson S. *Poisson approximation*. Oxford, Oxfordshire, United Kingdom: The Clarendon Press Oxford University Press; 1992.
42. Bavinton BR, Grulich AE, Broady T, et al. Increases in HIV Testing Frequency in Australian Gay and Bisexual Men are Concentrated Among PrEP Users: An Analysis of Australian Behavioural Surveillance Data, 2013-2018. *AIDS Behav* 2020;24(9):2691-702.

Table 1: Characteristics of the TAIPAN study population, New South Wales and Victoria, Australia, 2012-2018

Characteristic	**Remained HIV negative (45,719)**			**Seroconverters (885)**			**All (46,604)**		
	No.	%	mean (SD)	No.	%	mean (SD)	No.	%	mean (SD)
Age at entry[a] (years)	32 (26-43)								
Age group at entry (years) <30	18,470	40		339	38		18,809	40	
30-34	7,691	17		155	18		7,846	17	
35-44	9,679	21		217	25		9,896	21	
45+	9,879	22		174	20		10,053	22	
Interval neg-neg tests[a] (days)	104 (71-216)			117 (71-216)			105 (71-217)		
Interval neg-pos tests[a] (days)	N/A			227 (109-566)			N/A		
Tests per year[b] 2011			1.49 (0.85)			1.73 (1.05)			1.50 (0.86)
2012			1.58 (0.92)			1.74 (1.00)			1,58 (0.93)
2013			1.60 (0.97)			1.81 (1.08)			1.61 (0.98)
2014			1.68 (1.04)			1.84 (1.14)			1.68 (1.05)
2015			1.79 (1.15)			1.75 (1.01)			1.79 (1.15)
2016			2.19 (1.48)			1.71 (1.22)			2.18 (1.48)
2017			2.43 (1.52)			1.53 (0.98)			2.43 (1.51)

2018			2.49 (1.47)			1.50 (0.91)			2.48 (1.47)
All			2.04 (1.35)			1.74 (1.07)			2.04 (1.34)

IQR=inter quartile range; SD=standard deviation; N/A=not applicable; TAIPAN=Treatment with Antiretrovirals and their Impact on Positive And Negative men

a. Values are expressed as median (interquartile range)

b. Sequences of repeat tests with maximum separation between consecutive tests of 7 days were considered as one testing event and allocated to first test date in given sequence

Figure 1 Illustration of incidence methods: A shows testing intervals between last negative and first positive test (dotted lines) for 2 patients recorded in a time period over years 1 and 2 of a study. Red crosses represent midpoint method assigned dates of infection. Blue crosses represent random point method assigned dates of infection. The green shaded area represents Poisson-binomial assigned period of infection. B shows the count of incident infections for yearly units of time based on the dates of infection.

Figure 2: Estimated incidence per 100 person years for constant simulated incidence and quarterly units of time by simulated baseline incidence rate and method of estimation for 1,000 simulations (A=0.10/100PY Poisson-Binomial; B=0.10/100PY Midpoint; C=0.10/100PY Random Point; D=1.0/100PY Poisson-Binomial; E=1.0/100PY Midpoint; F=1.0/100PY Random Point; G=10/100PY Poisson-Binomial; and H=10/100PY Midpoint; I=10/100PY Random Point. Bold dashed lines represent simulated estimate, and dash-dot lines represent 2.5th and 97.5th percentiles respectively. Upright dashed line represents simulated rate.

Figure 3: Estimated incidence per 100 person years for decreasing simulated incidence and quarterly units of time by simulated baseline incidence rate and method of estimation for 1,000 simulations (A=0.10/100PY Poisson-Binomial; B=0.10/100PY Midpoint; C=0.10/100PY Random Point; D=1.0/100PY Poisson-Binomial; E=1.0/100PY Midpoint; F=1.0/100PY Random Point; G=10/100PY Poisson-Binomial; and H=10/100PY Midpoint; I=10/100PY Random Point. Bold dashed lines represent simulated estimate, and dash-dot lines represent 2.5th and 97.5th percentiles respectively. Upright dashed line represents simulated rate.

Figure 4: Estimated incidence rates per 100 person years for 2012-2018 in the TAIPAN cohort by method based on quarterly unit of time with 6 months extra follow-up. Solid line with circular markers represents Poisson-Binomial method, dash-dot line with square markers represents Midpoint method, and long dashed line with triangular markers represents Random Point method. Markers or shown at quarter 1 of each year.