Preplanned Studies

Late HIV Diagnosis and Associated Factors Among Newly Reported HIV/AIDS Cases Aged ≥50 Years — China, 2022–2024

Duo Shan[1]; Yalan Wang[1]; Yuan Zhao[1]; Chunpeng Zang[1]; Juan Yang[1]; Hui Li[1]; Jiarong Ren[1]; Mengchi Liu[1]; Yufen Liu[1,#]

Summary

What is already known about this topic?

Late human immunodeficiency virus (HIV) diagnosis remains prevalent among older adults in China; however, recent national-level evidence characterizing its correlates remains limited.

What is added by this report?

Using nationwide surveillance data spanning 2022–2024 and employing two complementary analytical approaches, this study identified transmission route and geographic region as primary correlates of late HIV diagnosis. Decision-tree analysis further identified two distinct subgroups in eastern China exhibiting particularly elevated proportions of late diagnosis.

What are the implications for public health practice?

These findings provide epidemiological evidence to inform more targeted HIV screening strategies tailored to specific population subgroups, geographic regions, and diagnostic settings.

ABSTRACT

Introduction: Late HIV diagnosis represents a critical public health challenge among older adults in China. Identifying its correlates is essential for enhancing timely detection and improving health outcomes.

Methods: We analyzed newly reported human immunodeficiency virus (HIV)-infected cases aged ≥50 years using national surveillance data from 2022–2024. Logistic regression was employed to examine factors associated with late diagnosis, while decision tree modeling captured complex variable interactions.

Results: Among 162,026 cases, 77.78% were diagnosed late. Late diagnosis was more frequently observed among males, older individuals, non-migrants, ethnic minorities, those without sexually transmitted disease (STD) history, individuals with higher education, cases diagnosed in medical institutions, residents of eastern China, and those infected through non-marital commercial heterosexual contact (adjusted odds ratio=1.05–1.60). The decision-tree model identified transmission route and region as primary stratifiers. Within the eastern China branch, two terminal subgroups exhibited particularly high proportions of late diagnosis: individuals infected through non-marital commercial heterosexual contact (74.6%) and those infected through other heterosexual or men who have sex with men (MSM) routes who were diagnosed in medical institutions (73.0%).

Conclusions: Late HIV diagnosis among older adults in China remains persistently high. These findings identify specific population subgroups and diagnostic settings with elevated proportions of late diagnosis and underscore the need for targeted screening interventions, strengthened testing protocols in medical settings, and enhanced risk awareness among older heterosexual adults.

Late human immunodeficiency virus (HIV) diagnosis remains a major global challenge, despite substantial progress in HIV prevention and treatment (1). The European Late Diagnosis Consensus Working Group (2) defines late HIV diagnosis as an initial HIV diagnosis with a CD4 count <350 cells/μL or the presence of an acquired immunodeficiency syndrome (AIDS)-defining event, regardless of CD4 level. This phenomenon is particularly prevalent among adults aged ≥50 years, who often underestimate their infection risk and may engage in high-risk sexual behaviors without seeking timely testing (3). Among older adults, late diagnosis leads to compromised immune recovery, increased susceptibility to

opportunistic infections, and elevated mortality. At the population level, delayed diagnosis undermines the public health benefits of Treatment as Prevention, increases healthcare expenditures, and heightens the risk of ongoing transmission (*4*). Although numerous studies have examined late HIV diagnosis, most have been regional or population-specific, leaving a gap in recent national-level evidence. Furthermore, prior national investigations have predominantly focused on individual risk factors without exploring interaction-based risk stratification among older adults. To address these limitations, we analyzed recent nationwide surveillance data from adults aged ≥50 years using a classification and regression tree (CRT) model. Our objectives were to identify high-risk subgroups through interaction-based analysis and to characterize factors associated with late diagnosis, thereby informing more targeted screening and intervention strategies.

Data were obtained from the Chinese HIV/AIDS Comprehensive Response Information Management System. Among newly reported cases from January 1, 2022 to December 31, 2024 with a final review status of "approved" and aged ≥50 years at diagnosis, late diagnosis was defined using five established criteria (*5*): 1) HIV/AIDS deaths from non-accidental causes; 2) surviving or accidental-death HIV/AIDS cases with CD4 count <350 cells/μL; 3) surviving or accidental-death AIDS cases with CD4 count between 350–499 cells/μL; 4) surviving or accidental-death AIDS cases without CD4 testing; 5) surviving or accidental-death HIV cases without CD4 testing. For the fifth category, late diagnosis status was estimated based on the proportion of CD4 <350 cells/μL among tested cases within the same demographic stratum. Because this estimation approach could not be verified at the individual level and introduced potential measurement error, we excluded the fifth category from the multivariable analysis to ensure diagnostic accuracy.

Regions were grouped into eastern, central, and western zones according to economic development level. The variable 'source of detection' was included as a surveillance-based classification to capture differences in case-finding contexts. "Transmission route" was classified using standardized categories from the national HIV surveillance system at the time of case reporting. Variables were selected for analysis based on two criteria: their established relevance in prior surveillance-based studies of late HIV diagnosis and their availability within the national HIV surveillance system with standardized reporting protocols. Data

were cleaned using Excel 2021 (Microsoft Corporation, Redmond, Washington, USA) and analyzed in R version 4.2 (version 4.2, R Foundation for Statistical Computing, Vienna, Austria). Chi-square tests were used for group comparisons, and logistic regression was applied for both univariate and multivariate analyses. Adjusted odds ratios (*aOR*s) and 95% confidence intervals (*CI*s) were calculated. Statistical significance was set at *P*<0.05. A classification and regression tree (CRT) model was developed to predict late diagnosis using recursive binary splitting, which partitions data based on variables that maximize homogeneity within resulting subgroups (measured by reduction in Gini impurity, a metric quantifying classification accuracy). Data were randomly divided into training (70%) and validation (30%) sets, with cross-validation pruning applied to prevent overfitting. Variable importance was assessed by evaluating each variable's split frequency and contribution to model performance, with region and transmission route emerging as key stratifiers.

A total of 162,026 newly reported HIV/AIDS cases aged ≥50 years were included in the analysis, of whom 71.13% were male (115,279/162,026). Overall, 77.78% (126,018/162,026) were diagnosed late. Demographic characteristics revealed that nearly half were aged 50–59 years (47.09%), 65.0% were farmers, and the majority were of Han ethnicity (87.93%), had education below primary school level (61.45%), and were married (57.10%). Geographically, eastern China accounted for 48.45% of all cases, and 96.19% were non-migrants. Heterosexual transmission was the predominant route (90.23%), with non-marital non-commercial contact representing the most common subtype (42.03%, 68,106/162,026). Most cases were identified through medical institutions (69.51%, 112,619/162,026).

After excluding individuals without CD4 results, 145,741 cases remained for multivariable analysis. Among these cases with complete CD4 data, 75.29% (109,733/145,741) were diagnosed late, a slightly lower proportion than the overall sample due to the exclusion of estimated late diagnoses without laboratory confirmation. Late diagnosis demonstrated positive associations with male sex (*aOR*=1.13), non-migration status (*aOR*=1.18), absence of STD history (*aOR*=1.11), higher educational attainment (*aOR*=1.08–1.10), and residence in eastern China (*aOR*=1.14). Weaker positive associations were observed for ethnic minority status (*aOR*=1.06) and

older age (*aOR*=1.08–1.11). Regarding transmission routes, non-marital commercial heterosexual contact exhibited the highest risk relative to injection drug use (IDU) (*aOR*=1.60, 95% *CI*: 1.25, 2.05). Cases detected in medical institutions also demonstrated elevated odds of late diagnosis compared with those identified through key population screening (*aOR*=1.05, 95% *CI*: 1.02, 1.08) (Table 1).

The decision-tree model identified transmission route, region, migration status, sex, and source of detection as the primary stratifiers of late diagnosis, yielding 14 terminal subgroups (referred to as "nodes" in tree-based classification). The overall proportion of late diagnosis at the root node was 67.7%. Following stratification by region, the model identified two particularly high-risk terminal nodes within the eastern China branch (Node 3). The first high-risk pathway comprised individuals in eastern China infected through non-marital commercial heterosexual contact (Node 7), with a late diagnosis proportion of 74.6%. The second high-risk pathway included individuals in eastern China infected through transmission from an HIV-positive spouse/regular partner, men who have sex with men (MSM), or non-marital non-commercial heterosexual contact, who were subsequently diagnosed in medical institutions (Node 11), with a late diagnosis proportion of 73.0% (Figure 1).

The logistic regression and decision-tree models demonstrated comparable overall discrimination, with AUCs of 0.710 (95% *CI*: 0.708–0.713) and 0.718 (95% *CI*: 0.715–0.721), respectively (*P*<0.001). However, the decision-tree model exhibited superior sensitivity (0.976 *vs.* 0.859) and a larger Youden index (0.451 *vs.* 0.326), indicating moderately better performance in identifying late diagnosis among adults aged ≥50 years (Figure 2).

DISCUSSION

Based on nationwide surveillance data from 2022 to 2024, a substantial proportion of adults aged ≥50 years were diagnosed at a late stage (77.78%), a rate that has remained persistently elevated compared with earlier reports in which late diagnosis consistently exceeded 65% during 2015–2019 (*3*). Although internal data suggested a temporary decline during 2020–2021, potentially attributable to coronavirus disease 2019 (COVID-19) pandemic-related disruptions in testing services, the persistently high proportion observed in recent years indicates that timely HIV diagnosis among

older adults continues to present significant challenges. Previous studies have generally demonstrated that late HIV diagnosis among older adults is associated with a range of demographic and epidemiological factors, including male sex, heterosexual transmission, and facility-based diagnosis, most often identified through logistic regression analyses (*3,6*). Consistent with this literature, our regression results confirmed similar associations. Through decision-tree modeling, we extended these findings by illustrating how such factors co-occur within the observed data. Beyond identifying independent associations, the decision-tree model delineated two high-risk terminal nodes in eastern China — non-marital commercial heterosexual transmission and other heterosexual or MSM transmission routes combined with diagnosis in medical institutions — thereby uncovering interaction-based risk patterns not captured by regression analysis alone.

Notably, the eastern region exhibited the highest proportion of late diagnosis, contrasting with earlier studies that reported greater risks in southwestern China (*7–8*). This regional difference likely reflects variation in HIV service delivery pathways. Long-standing high-burden areas in the southwest have developed more established surveillance systems and proactive testing practices, whereas routine HIV screening remains less systematically integrated into outpatient care across many eastern settings (*9*). Additionally, a higher proportion of late diagnosis was observed among cases detected in medical institutions, suggesting persistent reliance on symptom-driven testing rather than routine screening protocols. Non-migrants demonstrated a higher proportion of late diagnosis, potentially reflecting reduced access to testing opportunities or support services. The variable "no history of sexually transmitted disease" showed only weak associations and did not emerge as a splitting variable in the decision-tree model, indicating limited contribution to risk stratification. While earlier studies predominantly identified individual-level factors such as sex, education, and transmission route (*3,8*), the decision-tree analysis in this study further delineated specific factor combinations — particularly heterosexual transmission in eastern regions combined with diagnosis in medical institutions — thereby providing a more nuanced characterization of heterogeneity in late diagnosis patterns.

This study has several limitations. First, reliance on case-reporting data without serological markers for

TABLE 1. Factors associated with late diagnosis among cases aged ≥50 years, 2022–2024 (*n*=145,741).

Variables	Late diagnosis		Multivariate analysis				
	N	Proportion (%)	*β*	S.E.	*Z*	*P*	*aOR* (95% *CI*)
Sex							
Female	30,393	70.51					1.00 (Reference)
Male	79,340	77.30	0.13	0.01	8.90	<0.05	1.13 (1.10, 1.16)
Age (years)							
50–59	50,996	76.08					1.00 (Reference)
60–69	35,961	80.54	0.10	0.01	8.28	<0.05	1.11 (1.08, 1.13)
70–79	19,092	78.34	0.08	0.02	5.15	<0.05	1.08 (1.05, 1.11)
≥80	3,684	73.34	0.03	0.03	−4.87	<0.05	0.87 (0.82, 0.92)
Migration status							
Yes	3,932	75.27					1.00 (Reference)
No	105,801	76.02	0.17	0.03	6.05	<0.05	1.18 (1.12, 1.25)
Occupation							
Farmer	71,656	75.64					1.00 (Reference)
Others	38,077	74.64	−0.03	0.01	−2.22	<0.05	0.97 (0.95, 0.99)
Ethnicity							
Han	96,299	74.31					1.00 (Reference)
Ethnic minorities	13,434	75.43	0.05	0.02	3.16	<0.05	1.06 (1.02, 1.09)
History of sexually transmitted diseases (STDs)							
Yes	14,126	75.08					1.00 (Reference)
No	87,098	75.35	0.10	0.02	6.63	<0.05	1.11 (1.08, 1.14)
Unknown	8,509	75.05	0.07	0.02	3.06	<0.05	1.08 (1.03, 1.13)
Education level							
Illiterate	12,952	74.08					1.00 (Reference)
Primary school	54,191	75.06	0.09	0.02	5.13	<0.05	1.09 (1.06, 1.13)
Junior high school	31,131	75.85	0.10	0.02	4.98	<0.05	1.10 (1.06, 1.14)
Senior high school or above	11,459	74.63	0.07	0.02	3.04	<0.05	1.08 (1.03, 1.13)
Source of detection							
Key population screening	21,291	71.75					1.00 (Reference)
Medical institutions	76,576	76.99	0.05	0.01	3.43	<0.05	1.05 (1.02, 1.08)
Others	11,866	71.49	−0.03	0.02	−1.26	0.207	0.97 (0.94, 1.01)
Regional distribution							
Western	51,944	74.12					1.00 (Reference)
Central	17,544	74.78	0.04	0.02	2.49	<0.05	1.06 (1.03, 1.09)
Eastern	40,245	76.49	0.13	0.01	10.38	<0.05	1.14 (1.11, 1.17)
Route of transmission							
Injecting drug use	157	72.59					1.00 (Reference)
Others or unknown	998	70.66	−0.26	0.13	−1.92	0.055	0.77 (0.59, 1.01)
Male-to-male sexual contact	9,407	73.42	0.33	0.13	2.63	<0.05	1.40 (1.09, 1.79)
Unclassified heterosexual transmission	5,451	71.04	−0.18	0.13	−1.40	0.161	0.84 (0.65, 1.07)
HIV-positive spouse or regular partner	8,747	75.21	0.37	0.13	2.91	<0.05	1.45 (1.13, 1.86)
Non-marital commercial heterosexual contact	39,201	77.92	0.47	0.13	3.73	<0.05	1.60 (1.25, 2.05)
Non-marital non-commercial heterosexual contact	45,772	74.81	0.34	0.13	2.72	<0.05	1.41 (1.10, 1.80)

Abbreviation: *aOR*=adjusted odds ratio; *CI*=confidence interval; S.E.=standard error; *Z*=Z statistic.
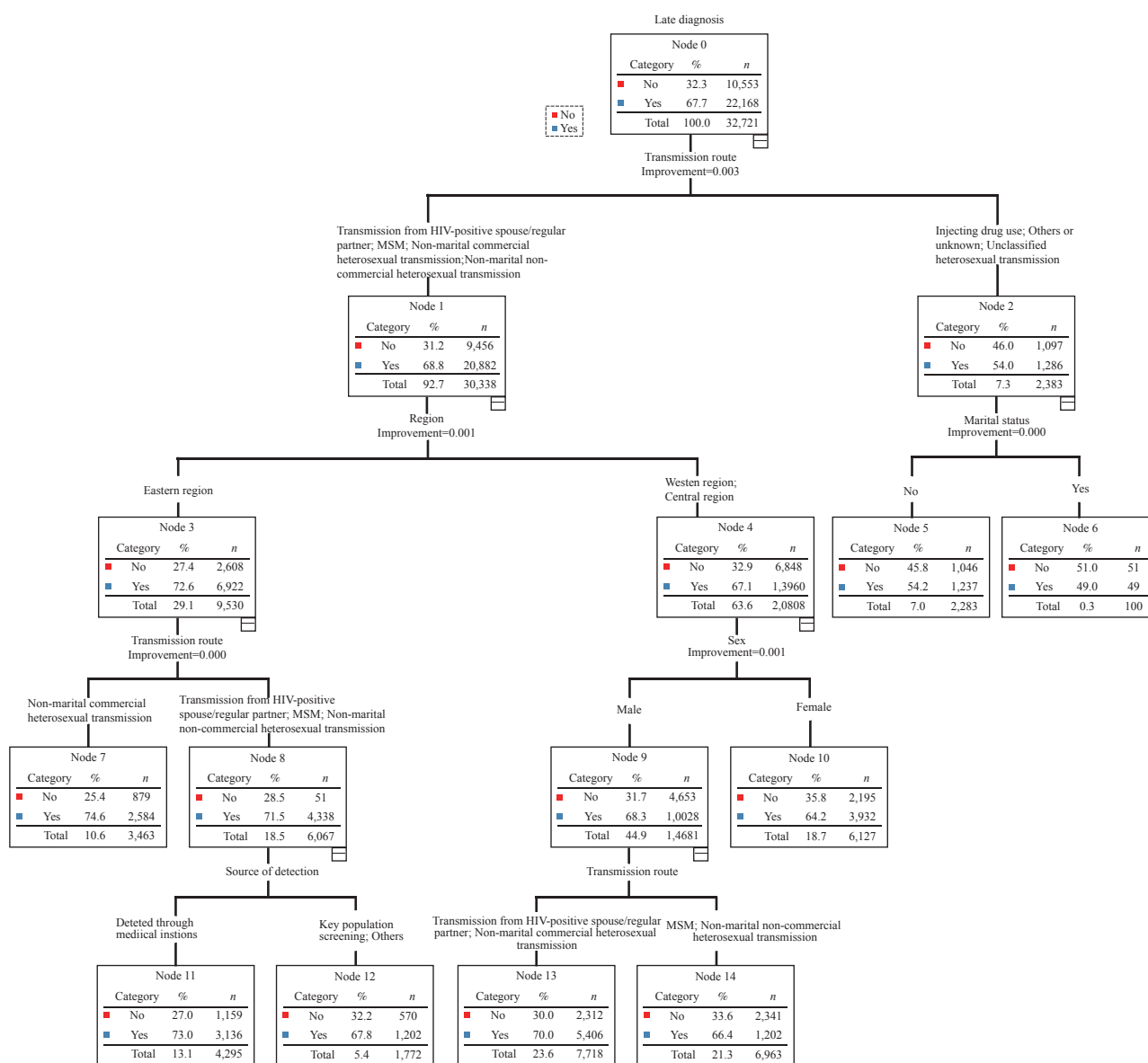
FIGURE 1. Decision tree model analysis results of late diagnosis aged ≥50 years, 2022–2024.

recent infection (e.g., limiting antigen avidity assays) may introduce misclassification bias in distinguishing recent from long-standing infections at the time of diagnosis. Although excluding estimated cases reduced measurement error, this approach may have also reduced statistical power and limited generalizability. Second, potential misclassification of sexual transmission categories may exist (*10*), as self-reported behavioral information is susceptible to recall bias and non-disclosure. Third, the cross-sectional design precludes causal inference. Furthermore, key behavioral, psychosocial, and healthcare access–related factors were not captured in the surveillance dataset and therefore could not be assessed. Despite these

limitations, the use of nationwide surveillance data and complementary analytic approaches strengthens the robustness of the findings. In conclusion, this study identified substantial heterogeneity in late HIV diagnosis among older adults in China, providing epidemiologically grounded evidence to inform targeted early testing strategies and prioritize interventions for high-risk subgroups.

Conflicts of interest: The authors declare no conflicts of interest.

Ethical statement: Approval by the Ethics Committee of the National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention (approval number:
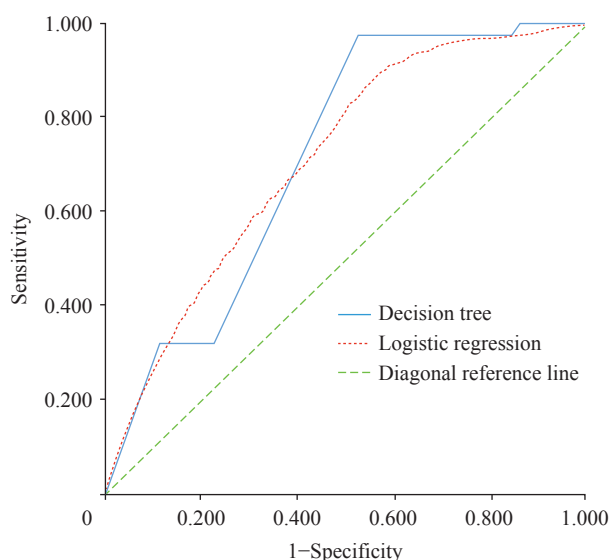
FIGURE 2. ROC curves comparing the predictive performance of the two models for late HIV diagnosis among adults aged ≥50 years, 2022–2024.
Abbreviation: ROC=receiver operating characteristic.

KX250108838).

Corresponding author: Yufen Liu, yufenliu69@chinaaids.cn.

1 National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.

REFERENCES

1. He N. Research progress in the epidemiology of HIV/AIDS in China. China CDC Wkly 2021;3(48):1022 – 30. https://doi.org/10.46234/ccdcw2021.249.

2. Croxford S, Stengaard AR, Brännström J, Combs L, Dedes N, Girardi E, et al. Late diagnosis of HIV: an updated consensus definition. HIV Med 2022;23(11):1202 – 8. https://doi.org/10.1111/hiv.13425.

3. Ma KF, Zhang XT, Ge L, Chen FF, Cai C, Qin QQ, et al. Analysis on the late-diagnosis among newly detected HIV/AIDS cases aged 50 years or older in China from 2015 to 2019. Chin J AIDS STD 2022;28(1): 16 – 20. https://doi.org/10.13419/j.cnki.aids.2022.01.04.

4. Moreno S, Mocroft A, Monforte AD. Medical and societal consequences of late presentation. Antivir Ther 2010;15 Suppl 1:9-15. http://dx.doi.org/10.3851/IMP1523.

5. Jin X, Xiong R, Wang LY, Mao YR. Analysis on the 'late diagnosis' (LD) phenomena among newly identified HIV/AIDS cases in China, 2010-2014. Chin J Epidemiol 2016;37(2):218 – 21. https://doi.org/10.3760/cma.j.issn.0254-6450.2016.02.014.

6. Jin J, Pan SN, Chen J, Yin JL, Ba HH, Hou HH, et al. Prevalence, risk factors, and clinical outcomes with advanced HIV disease among people with newly diagnosed HIV during the "Treat-All" era: a retrospective cohort study from Xi'an City, China. Infect Drug Resist 2025;18:2427 – 38. https://doi.org/10.2147/IDR.S518809.

7. Hu X, Liang BY, Zhou CX, Jiang JJ, Huang JG, Ning CY, et al. HIV late presentation and advanced HIV disease among patients with newly diagnosed HIV/AIDS in southwestern China: a large-scale cross-sectional study. AIDS Res Ther 2019;16(1):6. https://doi.org/10.1186/s12981-019-0221-7.

8. Hou YS, Jin YC, Cai C, Tang HL, Qin QQ, Lyu F. Characteristics of the HIV/AIDS epidemic among people aged ≥ 50 years in China during 2018-2021. Biomed Environ Sci 2024;37(4):399 – 405. https://doi.org/10.3967/bes2024.044.

9. Sun CQ, Li JJ, Liu XY, Zhang Z, Qiu T, Hu HY, et al. HIV/AIDS late presentation and its associated factors in China from 2010 to 2020: a systematic review and meta-analysis. AIDS Res Ther 2021;18(1):96. https://doi.org/10.1186/s12981-021-00415-2.

10. Cai C, Tang HL, Li DM, Qin QQ, Chen FF, Jin YC, et al. Evolution of HIV epidemic and emerging challenges—China, 1989-2023. China CDC Weekly 2024;6(48):1251 – 6. https://doi.org/10.46234/ccdcw2024.251.