Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Age-specific incidence trends and future burden of hepatitis E in China: a nationwide 16-year study with vaccination scenarios

Fei Wang
First Affiliated Hospital of Xi'an Jiaotong University
Yajuan He
First Affiliated Hospital of Xi'an Jiaotong University
Yuanyuan Duan
Xian Center for Disease Control and Prevention
Mi Ke
First Affiliated Hospital of Xi'an Jiaotong University
Yuxin Liu
First Affiliated Hospital of Xi'an Jiaotong University
Tianyan Chen
First Affiliated Hospital of Xi'an Jiaotong University
Zhen Tian
tianzhen@xjtufh.edu.cn

First Affiliated Hospital of Xi'an Jiaotong University

Research Article

Keywords: hepatitis E, age-specific incidence, population ageing, temperature, disease burden projection, vaccination, China

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11615986/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background:

Reported hepatitis E incidence has increased in China, but the relative contributions of age-specific rates, population ageing and temperature remain incompletely quantified. We assessed these contributions and explored the potential direct impact of vaccination scenarios.

Methods:

We analysed 388,867 cases reported in 31 mainland provincial-level divisions during 2005–2020 by age band (0–19, 20–59, ≥ 60 years). Trend, decomposition and temperature analyses ended in 2019; 2020 was used for description and sensitivity analyses. Kitagawa decomposition separated age-structure and age-specific-rate contributions. Province-specific distributed-lag non-linear models were pooled by multivariate meta-analysis. Projections combined the 2015–2019 baseline, United Nations population scenarios and temperature-derived climate multipliers. Fixed rates were compared with log-linear continuation using 2013–2019 and 2005–2019 windows. Vaccination scenarios estimated direct protection against reported clinical disease.

Results:

Between 2005 and 2019, crude incidence rose from 1.21 to 2.01 per 100,000; 78.9% of the increase was attributed to age-specific rates and 21.1% to age structure. At 27.37°C versus 14.178°C, the pooled relative risk among adults aged ≥ 60 years was 1.36 (95% CI 1.03–1.82), with I^2=0%; younger-band associations were non-significant. With age-specific rates fixed, annual cases in the 2050s were 34,365 under SSP2-4.5 and 34,667 under SSP5-8.5, respectively 22.0% and 23.1% above baseline; older-adult cases approximately doubled. Recent-trend scenarios ranged from 27,204 to 31,657 under SSP2-4.5, reflecting a fitted annual decline of 1.70% in reported incidence among adults aged ≥ 60 years during 2013–2019; long-term trend scenarios were substantially higher. Under fixed-rate SSP2-4.5 in the 2050s, 60% effective full-course coverage among adults aged ≥ 60 years could potentially avert 10,290 reported cases annually, requiring 26,940 full courses per case averted.

Conclusions:

Historical increases were driven mainly by age-specific rates, whereas ageing dominates fixed-rate projections. Future burden depends on trend window and damping. Older adults combine high reported incidence with a modest warm-tail association whose magnitude was sensitive to inclusion of 2020. Vaccination estimates illustrate targeting efficiency, not cost-effectiveness or a recommendation for routine older-adult immunisation.

Background

Hepatitis E virus (HEV) is a major cause of acute viral hepatitis worldwide, with transmission pathways and clinical severity varying by setting and host susceptibility [1, 2]. In China, the reported epidemiology has shifted from large waterborne outbreaks towards predominantly sporadic disease, while national notifications have increased even as several other notifiable viral hepatitides declined [3–5]. This divergence makes hepatitis E an important test of whether changes in reported burden reflect population structure, age-specific risk, environmental conditions, or changes in ascertainment.

Three features motivate an age-centred analysis. First, existing burden projections have largely extrapolated aggregate trends without jointly separating demographic structure, age-specific rates and vaccination scenarios [6]. Second, national counts can rise because age-specific rates increase, because the population becomes older, or both; treating these as interchangeable obscures different prevention implications. Third, older adults bear a high reported incidence and can experience severe outcomes, particularly when HEV superinfection occurs in people with chronic liver disease [7], while a licensed recombinant vaccine provides a potential prevention tool whose population-level targeting requires separate evaluation [8, 9].

We therefore analysed nationwide, province-month-age-band surveillance data for 2005–2020. Our objectives were to: (1) describe age-specific reported incidence and decompose the long-term change in crude incidence; (2) estimate temperature–incidence associations by age band; (3) compare fixed-rate and rate-continuation burden scenarios for the 2030s and 2050s; and (4) estimate the direct reported-clinical-case impact and resource requirements of illustrative vaccination strategies. We treated 2020 separately because pandemic-related disruption could distort long-term trend and exposure–response estimates.

Methods

Study design and data sources

We conducted a nationwide ecological time-series and scenario study covering 31 provincial-level divisions of mainland China. Monthly clinically and laboratory-confirmed hepatitis E cases and deaths by province and single-year age were obtained from the Chinese Public Health Science Data Center, derived from the National Notifiable Disease Reporting System, for January 2005 through December 2020 [10]. Counts were aggregated into 0–19, 20–59 and ≥ 60-year bands. Five cases with unknown age were included in all-age descriptive totals but excluded from age-band rates, decomposition and projections.

Provincial monthly mean temperature and precipitation were derived from the CN05.1 gridded observational dataset at 0.25° resolution [11], and relative humidity was obtained from NASA POWER. Annual provincial populations by age band were compiled from national statistical yearbooks [12]. Future total population and age shares used the United Nations World Population Prospects 2024

medium variant [13]. Climate increments followed CMIP6-consistent estimates for China relative to 1995–2014 [14].

The full 2005–2020 period was used for descriptive burden. Primary trend decomposition and temperature models ended in 2019; 2020 was used only in descriptive and sensitivity analyses. The projection baseline was the mean of 2015–2019. This separation was specified to avoid interpreting the 2020 reporting contraction as a long-term epidemiological trend.

Using a common national age-band denominator allowed case counts and rates to answer different questions: counts indicate service burden, whereas rates permit comparison across age groups and years. We did not age-standardise the primary crude-rate decomposition because its purpose was precisely to quantify the contribution of changing population composition. Deaths were summarised descriptively because the small annual numbers did not support parallel age-specific projection or vaccination models. No correction factor was applied to the 2020 counts, because the available data cannot distinguish reduced transmission from reduced detection or reporting.

Observed burden and Kitagawa decomposition

Annual age-band incidence was calculated per 100,000 population. Kitagawa decomposition separated the change in crude incidence between 2005 and 2019 into age-structure and age-specific-rate components, using average population shares and average rates at the two endpoints [15]. The components sum exactly to the observed change. A 2020 endpoint calculation was treated as sensitivity analysis rather than as a description of the underlying long-term trend.

Temperature–incidence analysis

Province-specific quasi-Poisson distributed-lag non-linear models related monthly age-band counts to monthly mean temperature using a natural cubic spline cross-basis and lags of 0–3 months [16]. Models adjusted for calendar-month seasonality, long-term trend, precipitation and relative humidity, with log population as an offset. Province-band coefficient vectors were pooled using multivariate random-effects meta-analysis, and between-province heterogeneity was summarised using I^2 [17]. The primary analysis excluded 2020; inclusion of 2020 was a sensitivity analysis. We report relative risks (RRs) at 27.37°C and − 5.92°C versus the 14.178°C reference. Sparse paediatric series reduced the number of contributing provinces for the 0–19-year band.

Conditional burden scenarios

The primary demographic scenarios held age-specific rates at their 2015–2019 means. Projected annual cases were calculated by multiplying these rates by future age-band populations. Total population was set to 1.39 billion in the 2030s and 1.32 billion in the 2050s; the share aged ≥ 60 years was 25% and 35%,

respectively [13]. Changes from baseline were decomposed sequentially into total-population scale, age structure, age-specific-rate change and climate components.

Rate-continuation scenarios fitted ordinary least-squares log-linear trends to annual age-band rates. The primary trend window was 2013–2019 and the sensitivity window was 2005–2019. Extrapolation from the 2015–2019 baseline used each fitted slope multiplied by a cumulative damped-time term. Damping coefficients of 1.00, 0.95 and 0.90 represented no damping, 5% annual attenuation and 10% annual attenuation of the cumulative log-rate trend, respectively. These scenario ranges are not statistical confidence intervals and are not probabilistic forecasts.

Climate multipliers were derived from the primary 2005–2019 exposure–response curves. For each age band, we calculated $RR(T + \Delta T)/RR(T)$ over the observed 2015–2019 province-month temperature distribution and averaged the ratios using baseline reported cases as weights. Warming increments were + 0.4°C in the 2030s under both pathways, + 0.8°C in the 2050s under SSP2-4.5 and + 1.3°C under SSP5-8.5 [14]. The multipliers therefore account for the temperature distribution rather than applying a single warm-tail RR to all cases.

Vaccination scenarios

We evaluated two target populations: adults aged ≥ 60 years and all adults aged ≥ 20 years. Effective full-course coverage was set to 40%, 60% or 80%, combining completion and effective delivery rather than representing first-dose uptake. The base vaccine efficacy (VE) was 0.868, the modified intention-to-treat estimate over 4.5 years; sensitivity analyses used 0.70 and 0.95 [8, 9]. The original trial enrolled healthy adults aged 16–65 years, so application beyond age 65 requires extrapolation.

Annual potentially averted reported clinical cases equalled projected target-band reported cases multiplied by effective coverage and VE. We also calculated vaccinated population, full courses per annual case averted and doses per annual case averted assuming a three-dose schedule. Base results describe annual impact while effective coverage is maintained. Five- and 10-year protection-window calculations multiplied annual impact by the window under constant incidence and no waning; they are arithmetic sensitivity scenarios, not evidence of protection duration. Only direct protection was included. We did not model herd immunity, infection occurrence, under-ascertainment, natural immunity, revaccination, implementation costs or cost-effectiveness.

Statistical analysis

Analyses used Python with pandas, NumPy, SciPy and statsmodels; distributed-lag and multivariate meta-analysis procedures followed the dlnm and mvmeta frameworks [16, 17]. Confidence intervals for temperature estimates were two-sided 95% intervals. Trend-window p values describe historical log-linear fits and were not used to select a preferred future scenario. Projection decomposition closure and

missingness were checked programmatically. Tables and figures present rounded values; calculations used unrounded outputs.

Quality-control checks compared all-age totals with the sum of known age bands plus unknown-age cases and confirmed that baseline counts reconciled with the projection input. For every projection, the baseline plus the scale, age-structure, rate and climate components was required to equal the final total within numerical tolerance. We generated vaccination outputs for every fixed-rate and rate-continuation scenario rather than restricting sensitivity analyses to the most favourable projection. The complete projection scenario grid is supplied in Additional file 1; other deterministic outputs are available on reasonable request, allowing readers to distinguish sampling estimates in the observed models from deterministic choices in the future scenarios.

Kimi (Moonshot AI) was used to assist statistical code development and language editing. The authors reviewed all generated code, analytical outputs and text and take full responsibility for the content. Figures were generated programmatically from numerical results using matplotlib; no generative AI was used to create or alter scientific images.

Results

Nationwide and age-specific reported burden

Between 2005 and 2020, 388,867 hepatitis E cases and 423 deaths were reported, including 388,862 cases in the three known age bands and five with unknown age (Table 1). The 2015–2019 annual mean was 28,172.6 cases across all ages and 28,172.2 in known age bands. Baseline rates were 0.106 per 100,000 at 0–19 years, 2.173 at 20–59 years and 4.184 at ≥ 60 years; the older-adult rate was therefore approximately 1.9 times that of younger adults. In 2020, reported cases fell to 19,034 and the crude rate to 1.35 per 100,000, with declines in every age band (Table 1; Fig. 1a).

Table 1
Reported hepatitis E burden by age band and selected period, mainland China, 2005–2020

Characteristic	2005	2010	2015–2019 annual mean	2019	2020
0–19 years: Cases, n/year	364	453	345	345	216
Deaths, n/year	0	0	0	0	0
Rate per 100,000	0.10	0.14	0.11	0.11	0.07
20–59 years: Cases, n/year	11,060	16,110	17,894	17,864	12,498
Deaths, n/year	21	13	6	5	7
Rate per 100,000	1.42	1.93	2.17	2.18	1.52
≥ 60 years: Cases, n/year	4,114	7,119	9,933	9,946	6,320
Deaths, n/year	27	22	10	11	5
Rate per 100,000	2.69	4.01	4.18	3.90	2.39
Unknown age: Cases, n/year	3	0	0	0	0
All ages: Cases, n/year	15,541	23,682	28,173	28,155	19,034
Deaths, n/year	48	35	15	16	12
Crude rate per 100,000	1.21	1.78	2.03	2.01	1.35

Note: Rates are per 100,000 population. Values for 2015–2019 are annual means. Unknown-age population and rates are not calculable. The five unknown-age cases across 2005–2020 explain small differences between all-age and known-age-band totals. The known-age-band baseline used for projections is 28,172.2 cases/year; the corresponding all-age mean is 28,172.6.

The age-specific curves provide context for these summary rates. Reported incidence at 0–19 years remained below 0.2 per 100,000 throughout the study period. Adults aged 20–59 years accounted for the largest share of cases because their population denominator was largest, whereas adults aged ≥ 60 years had the highest rate in every year. The older-adult rate rose markedly during the earlier part of the study and remained above its 2005 level through 2019 despite lower values after 2011. The parallel downward movement in all three bands in 2020 contrasts with their different longer-term trajectories and supports treating that year separately rather than as the next point in a smooth national trend.

Decomposition of the 2005–2019 increase

Among known age bands, crude incidence increased from 1.208 per 100,000 in 2005 to 2.009 in 2019, an absolute change of 0.801 per 100,000. Age-structure change contributed 0.169 per 100,000 (21.1%) and age-specific-rate change contributed 0.632 (78.9%; Fig. 1b). Thus, most of the pre-pandemic increase was not explained by population ageing alone. Historical trend fits were nevertheless window-

dependent: during 2013–2019, the annual rate change was − 1.41% at 0–19 years, + 0.23% at 20–59 years and − 1.70% at ≥ 60 years, whereas the corresponding 2005–2019 estimates were − 1.48%, + 2.73% and + 2.58% (Additional file 1: Table S1).

Temperature associations by age

In the primary 2005–2019 analysis, the RR at 27.37°C versus 14.178°C was 2.01 (95% CI 0.75–5.41) at 0–19 years, 1.00 (0.79–1.29) at 20–59 years and 1.36 (1.03–1.82) at ≥ 60 years (Additional file 1: Table S2; Additional file 1: Fig. S1). Only the older-adult warm-tail association was statistically significant. At − 5.92°C, the older-adult RR was 0.49 (0.32–0.77). Between-province I^2 was 0% in all bands, with 24 provinces contributing to the paediatric pooled estimate and 30 to each adult estimate. Including 2020 strengthened the older-adult warm-tail estimate to 1.61 (1.22–2.12), supporting sensitivity to the pandemic-affected year rather than changing the direction of the primary finding.

Fixed-rate and rate-continuation burden scenarios

With age-specific rates fixed at baseline, annual cases increased to 32,359 in the 2030s (+ 14.9%) under either pathway (Fig. 2a). In the 2050s, totals were 34,365 under SSP2-4.5 (+ 22.0%) and 34,667 under SSP5-8.5 (+ 23.1%). Under SSP2-4.5, the change comprised + 7,039 cases from age structure, − 1,332 from total-population scale and + 486 from the climate multiplier. Under SSP5-8.5, the climate component was + 788 cases. These climate increments were small relative to the demographic component, despite the significant warm-tail association in older adults.

Cases at ≥ 60 years rose from 9,932.8 at baseline to 14,700 in the 2030s and to 19,757 (SSP2-4.5) or 20,022 (SSP5-8.5) in the 2050s, approximately a 99–102% increase (Fig. 2b). This near doubling is conditional on fixed age-specific rates and the assumed future age composition.

Rate-continuation scenarios did not uniformly exceed the fixed-rate projection. Across damping coefficients based on 2013–2019 trends, 2050s totals ranged from 27,204 to 31,657 under SSP2-4.5 and from 27,397 to 31,918 under SSP5-8.5 (Additional file 1: Table S3). By contrast, 2005–2019 trend scenarios under SSP2-4.5 ranged from 44,090 to 76,898 across the same damping grid (Additional file 1: Fig. S2). The recent window implies declining older-adult rates, whereas the long window implies rising adult rates; the future total is therefore dominated by trend-window and damping assumptions as well as demography.

Potential direct impact of vaccination scenarios

Under fixed-rate SSP2-4.5, 60% effective full-course coverage at ≥ 60 years could potentially avert 7,656 reported cases annually in the 2030s and 10,290 in the 2050s (Table 2). At 40% and 80% coverage, corresponding 2050s estimates were 6,860 and 13,719. The programme would require 27,234 full courses per annual case averted in the 2030s and 26,940 in the 2050s; at three doses per course, the latter equals 80,820 doses per annual case averted. For all adults aged ≥ 20 years, 60% coverage in the 2050s could potentially avert 17,783 reported cases annually but would require 37,857 full courses per case averted. Older-adult targeting therefore requires fewer courses per case averted, but the absolute

resource requirement remains substantial. Under SSP5-8.5, the corresponding older-adult estimate was 10,427 cases and 26,584 courses per case averted. VE sensitivity changed averted-case totals proportionally; vaccine-efficacy and protection-window results are reported in Additional file 1: Table S4 and Additional file 1: Table S5, respectively.

Table 2
Potential direct impact of vaccination under fixed-rate SSP2-4.5 scenarios

Period	Target	Effective coverage	Reported cases averted/year	Full courses per case averted	Doses per case averted
2030s	≥ 60 years	40%	5,104	27,234	81,702
2030s	≥ 60 years	60%	7,656	27,234	81,702
2030s	≥ 60 years	80%	10,208	27,234	81,702
2030s	≥ 20 years	40%	11,143	41,165	123,494
2030s	≥ 20 years	60%	16,715	41,165	123,494
2030s	≥ 20 years	80%	22,286	41,165	123,494
2050s	≥ 60 years	40%	6,860	26,940	80,820
2050s	≥ 60 years	60%	10,290	26,940	80,820
2050s	≥ 60 years	80%	13,719	26,940	80,820
2050s	≥ 20 years	40%	11,855	37,857	113,571
2050s	≥ 20 years	60%	17,783	37,857	113,571
2050s	≥ 20 years	80%	23,710	37,857	113,571

Note: Base vaccine efficacy is 0.868. Effective coverage refers to completion of an effective full course. Estimates represent direct protection against reported clinical cases while coverage is maintained, not all HEV infections or cost-effectiveness. Doses assume a three-dose course. Results are conditional on fixed age-specific rates and SSP2-4.5 population and climate inputs.

Discussion

This nationwide analysis yields four main findings. First, 78.9% of the 2005–2019 rise in crude reported incidence was attributable to age-specific-rate changes and 21.1% to age structure. Second, a positive warm-tail temperature association persisted among adults aged ≥ 60 years after exclusion of 2020, although it was weaker than in the full-period sensitivity analysis. Third, population ageing approximately doubles older-adult cases under fixed rates, but rate-continuation scenarios range from below baseline to markedly above it depending on the historical window. Fourth, vaccination scenarios show that older-adult targeting averts fewer total cases than vaccinating all adults but requires fewer full courses per reported case averted.

The decomposition distinguishes two processes that are sometimes combined under the label of ‘ageing’. A growing older population increases the number exposed to high age-specific rates, whereas increasing rates within age bands imply changes in ascertainment, exposure, susceptibility or their combination. Surveillance data cannot separate improved testing and reporting, care-seeking, diagnostic reclassification, dietary or exposure change, and birth-cohort effects. The long-term rise and the 2020 contraction should likewise be interpreted as distinct processes: the former occurred over 15 years, whereas the latter coincided with pandemic-related disruption to health care and notifications [10]. The recent decline in fitted older-adult rates cautions against assuming that earlier increases will continue indefinitely.

The temperature result complements, rather than independently proves, the rationale for paying attention to older adults: high reported incidence and temperature-related relative risk are different quantities and may share reporting or behavioural influences. The direction of the older-adult warm-tail association was consistent across analysis periods, but its magnitude was sensitive to inclusion of 2020 (RR 1.36 in the primary analysis versus 1.61 when 2020 was included) and should not be interpreted as precisely determined. A warm-tail RR does not imply that long-term warming explains most of the historical or future burden. The climate multipliers average modest temperature shifts over the full observed temperature distribution, while demographic ageing changes the size of the high-incidence group. This helps reconcile local positive or heterogeneous meteorological findings with a comparatively small national warming contribution [18–21].

The projection contrast is central. Fixed rates provide a transparent demographic benchmark, not a forecast that incidence will remain constant. The 2013–2019 window produces lower future cases because recent older-adult rates declined, while the 2005–2019 window produces higher cases because adult rates increased over the longer period. Damping further changes the totals. Presenting these as conditional scenario ranges avoids an internally inconsistent claim that rates rose historically but will necessarily remain fixed, and avoids selecting the window that produces the most dramatic headline. Updated post-2020 age-specific data will be essential before these scenarios can be treated as planning forecasts [6, 22].

Vaccination estimates should be read as decision-oriented boundaries rather than implementation advice. At equal effective coverage, vaccinating all adults averts more reported cases, whereas older-adult targeting uses fewer courses per case averted because incidence is concentrated and the target denominator is smaller. Neither metric establishes cost-effectiveness, affordability or acceptability. Trial participants were 16–65 years old, so efficacy above 65 is extrapolated; effective coverage, duration of protection, background natural immunity and the relation between reported clinical disease and all HEV infections add further uncertainty [8, 9]. The very large numbers of courses per reported case averted also show why high trial efficacy alone is insufficient to justify routine population-wide vaccination.

For surveillance and service planning, the findings support reporting both counts and age-specific rates. Counts alone would emphasise the larger adult population and could obscure the higher individual-level

reported rate in older adults; rates alone would understate the absolute service burden in younger adults. Maintaining single-year-age notifications, documenting diagnostic changes and retaining comparable denominators would allow future analyses to distinguish within-band ageing from genuine rate change. Projection updates should be triggered not only by new population estimates but also by material changes in age-specific notifications. A planning range that includes fixed and damped-trend scenarios is more defensible than a single extrapolated total when the recent and long-term fitted slopes differ in direction.

For prevention research, the results identify several questions that can be tested with more informative designs. Individual-level studies could determine whether occupational, dietary, clinical or care-seeking factors explain the age pattern. Linkage of notifications with diagnostic-test volumes could help separate ascertainment from incidence change. Vaccine studies in adults older than 65, including people with chronic liver disease, would reduce reliance on extrapolated efficacy. Economic evaluations should include vaccine price, delivery completion, hospitalisation averted and the proportion of infections that become reported clinical disease. Such evidence would convert the present direct-effect arithmetic into a genuine comparison of strategies without assuming that the age group with the highest rate is automatically the optimal target.

The study also has methodological strengths. It uses a common age structure across 31 provincial-level divisions over 192 months, preserves a national decomposition with exact closure, and pools temperature curves while reporting the number of contributing provinces. The same primary temperature curves feed the climate multipliers, preventing a mismatch between the exposure–response analysis and projection. Most importantly, the analysis separates observed statistical associations from deterministic scenario choices: confidence intervals are used for the former, whereas fixed rates, trend windows, damping, effective coverage and protection windows are explicitly labelled as assumptions. This distinction makes the outputs easier to update and less vulnerable to overinterpretation.

Limitations

This study has several limitations. Notifications underrepresent all infections and are vulnerable to changes in diagnostics, health-care access and reporting. Province-month exposure assignment is ecological and may misclassify individual exposure; no genotype or individual clinical data were available. The three broad age bands conceal within-band ageing and prevent direct estimation of single-year-age targeting. Future population totals and age shares are scenario inputs, and climate increments were applied nationally rather than downscaled. Excluding 2020 reduces pandemic distortion but does not remove earlier reporting changes, and post-2020 data were unavailable. Log-linear extrapolation is structural rather than mechanistic, and scenario ranges omit parameter and model uncertainty. Short-term food-supply and market disruptions were not modelled and cannot be excluded as influences on incidence in specific years. Vaccine calculations include direct protection only, assume simple multiplicative coverage and efficacy, and do not represent avoided infections, cost-effectiveness or an immunisation recommendation.

Conclusions

Most of China's pre-2020 increase in reported hepatitis E incidence reflected age-specific-rate changes rather than population ageing. Nevertheless, ageing produces a substantial increase in older-adult cases when rates are fixed, while alternative trend windows generate materially different future totals. Older adults have both high reported incidence and a modest warm-tail temperature association. Targeted vaccination scenarios quantify potential reported-case reductions and resource requirements, but clinical, economic and implementation evidence is needed before policy recommendations can be made.

Abbreviations

CI
confidence interval
CMIP6
Coupled Model Intercomparison Project Phase 6
HEV
hepatitis E virus
I^2
between-province heterogeneity statistic
NASA
National Aeronautics and Space Administration
RR
relative risk
SSP
Shared Socioeconomic Pathway
VE
vaccine efficacy.

Declarations

Declarations

Ethics approval and consent to participate.

The study used anonymised, aggregated surveillance counts and publicly available statistics; institutional review board approval and informed consent were not required.

Consent for publication.

Not applicable.

Competing interests.

The authors declare that they have no competing interests.

Additional files

Additional file 1.

File name: Additional file 1 (Additional_file_1.docx). File format: Microsoft Word document (.docx). Title: Supplementary analyses for age-specific hepatitis E trends and future burden. Description: Five supplementary tables covering historical age-specific rate trends, temperature–incidence estimates, all conditional projection scenarios, vaccine-efficacy sensitivity and protection-window sensitivity, plus two supplementary figures.

Funding.

This work was supported by the Natural Science Basic Research Program Youth Project of the Shaanxi Provincial Science and Technology Department [grant number 2023-JC-QN-0830]. The funding body played no role in study design, data collection, analysis or interpretation, or manuscript writing.

Author Contribution

FW: conceptualisation, data curation, formal analysis, visualisation and writing—original draft. YH: conceptualisation, data curation, methodology, validation and writing—original draft. YD: investigation, resources and validation. MK: validation and writing—review and editing. YL: data curation, methodology and validation. TC: supervision and writing—review and editing. ZT: conceptualisation, funding acquisition, project administration, supervision and writing—review and editing. FW and YH contributed equally. All authors read and approved the final manuscript.

Acknowledgement

We thank the Chinese Public Health Science Data Center and the Arid Land Ecology and Resources Science Data Center for data access.

Data Availability

Provincial monthly case counts were obtained from the Chinese Public Health Science Data Center and are available subject to its data-use agreements. CN05.1 meteorological data are available from the Arid Land Ecology and Resources Science Data Center , and NASA POWER data are publicly available.

Derived analysis datasets and analysis code are available from the corresponding author on reasonable request.

References

1. Kamar N, Bendall R, Legrand-Abravanel F, et al. Hepat E Lancet. 2012;379:2477–88.
2. World Health Organization. Hepatitis E fact sheet. Geneva: WHO. 2024. https://www.who.int/news-room/fact-sheets/detail/hepatitis-e. Accessed 4 Oct 2026.
3. Ren X, Wu P, Wang L, et al. Changing epidemiology of hepatitis A and hepatitis E viruses in China, 1990–2014. Emerg Infect Dis. 2017;23:276–9.
4. Jia Z, Yi Y, Liu J, et al. Epidemiology of hepatitis E virus in China: results from the third national viral hepatitis prevalence survey, 2005–2006. PLoS ONE. 2014;9:e110807.
5. Yang S, Wu J, Ding C, et al. Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study. Lancet Infect Dis. 2017;17:716–25.
6. Ouyang G, Pan G, Li Q, et al. Global burden of acute hepatitis E between 1990 and 2019 and projections until 2030. Liver Int. 2024;44:329–42.
7. Chen C, Zhang SY, Zhang DD, et al. Clinical features of acute hepatitis E super-infections on chronic hepatitis B. World J Gastroenterol. 2016;22:10388–98.
8. Zhu FC, Zhang J, Zhang XF, et al. Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale, randomised, double-blind placebo-controlled, phase 3 trial. Lancet. 2010;376:895–902.
9. Zhang J, Zhang XF, Huang SJ, et al. Long-term efficacy of a hepatitis E vaccine. N Engl J Med. 2015;372:914–22.
10. National Health Commission of the People's Republic of China. Annual notifiable infectious disease situation reports, 2005–2023. Beijing: NHC; 2006–24 (in Chinese).
11. Wu J, Gao XJ. A gridded daily observation dataset over China region and comparison with the other datasets. Chin J Geophys. 2013;56:1102–11. (in Chinese).
12. National Bureau of Statistics of China. China Statistical Yearbook (2006–2021 editions). Beijing: China Statistics (in Chinese).
13. United Nations. Department of Economic and Social Affairs, Population Division. World Population Prospects 2024. New York: United Nations; 2024.
14. IPCC. Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report. Cambridge: Cambridge University Press; 2021.
15. Kitagawa EM. Components of a difference between two rates. J Am Stat Assoc. 1955;50:1168–94.
16. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. Stat Med. 2010;29:2224–34.

17. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med. 2012;31:3821–39.
18. Yang L, Wang J, Peng J, Zhang L, Cheng X. Associations of ambient temperature and heat waves with risks of hepatitis E in Jiangsu, China (2010–2023): a time-stratified case-crossover study. Infect Dis Model. 2026;11:268–77.
19. Zhao X, Li M, Haihambo N, et al. Periodic characteristics of hepatitis virus infections from 2013 to 2020 and their association with meteorological factors in Guangdong, China: surveillance study. JMIR Public Health Surveill. 2023;9:e45199.
20. Huang X, Zhang X, Chen L, et al. Active surveillance of hepatitis E: a 10-year epidemiological analysis in a city in eastern China. Emerg Microbes Infect. 2024;13:2373315.
21. Zhu B, Liu J, Fu Y, Zhang B, Mao Y. Spatio-temporal epidemiology of viral hepatitis in China (2003–2015): implications for prevention and control policies. Int J Environ Res Public Health. 2018;15:661.
22. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability, disability-adjusted life-years, and healthy life expectancy for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403:2133–61.
23. Arid Land Ecology and Resources Science Data Center. CN05.1 meteorological data. https://data.dcxjegi.cn. Accessed 4 Nov 2026.
24. NASA Langley Research Center. POWER Data Access Viewer. https://power.larc.nasa.gov. Accessed 4 Nov 2026.

Figures

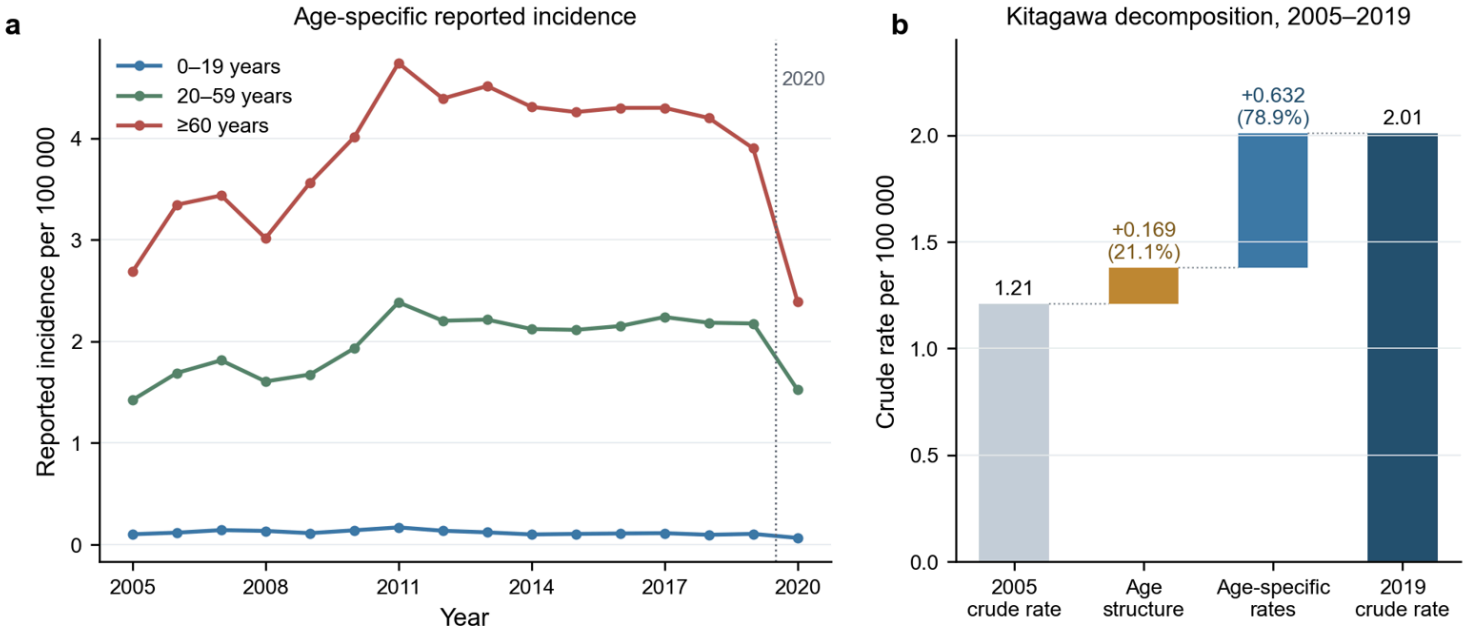


Figure 1

Age-specific reported incidence and Kitagawa decomposition.

(a) National annual reported incidence per 100,000 by age band, 2005–2020. The dotted line separates the pandemic-affected 2020 observation; primary trend analyses ended in 2019. (b) Decomposition of the change in crude incidence between 2005 and 2019 into age-structure and age-specific-rate components. Values are absolute changes per 100,000, with percentages of the total change.

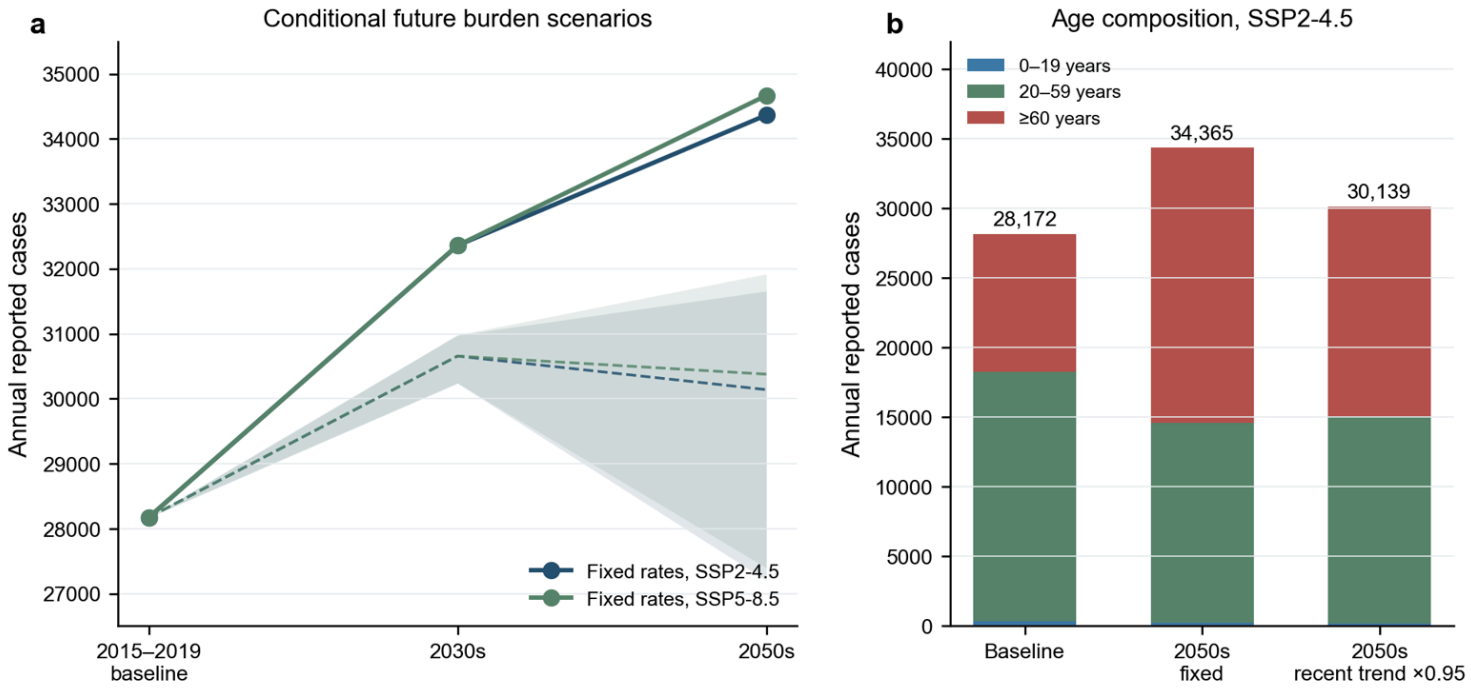


Figure 2

Conditional future burden scenarios.

(a) Solid lines show fixed-rate primary scenarios; dashed lines show the 2013–2019 rate-continuation scenario with damping 0.95; shaded areas span the three damping assumptions (1.00, 0.95 and 0.90). Shaded ranges are scenario ranges, not confidence intervals. (b) Age composition at baseline and in the 2050s under SSP2-4.5 for fixed rates and the recent-trend scenario with damping 0.95. Values above bars are total annual reported cases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx