**National and subnational levels and causes of mortality among 5-19-year olds in China in 2004-2019 – a systematic analysis of evidence from the Disease Surveillance Points System**

Yunning Liu1\*, Yue Chu\*2,3, Diana Yeung4, Wei Wang1, Lijun Wang1, Peng Yin1, Jiangmei Liu1, Maigeng Zhou1‡, Li Liu4,5‡

1 National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

2 Department of Sociology, The Ohio State University, Columbus, Ohio

3 Institute for Population Research, The Ohio State University, Columbus, Ohio

4 The Institute for International Programs, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

5 Department of Population Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

\*indicates joint first-authorship

‡indicates joint corresponding and senior authorship

Appendix 1. DSP design, sampling, and data collection

Appendix 2. Additional details with analysis

Appendix 3. ICD-mapping of IMPROVE causes of deaths categories

Appendix 4. GATHER checklist

Appendix 5. Additional tables and figures (available in a separate file)

**Appendix 1. DSP design, sampling, and data collection**

The design and sampling of the DSP system evolved over time by roughly 4 stages: pre-1990, 1990-2003, 2004-2012, 2013-present.1-4

Graphical user interface, text

Description automatically generated

Abbreviation: Pop: population; U: urban; R: rural; m: million; VR: vital registration system.

The system first started with pilot sites of Dongcheng district and Tong county of Beijing in 1978. In 1989, 71 surveillance points in big cities and wealth rural areas from 29 out of 31 provinces volunteered to participate in disease surveillance work.1

In 1990, to gain national coverage and representativeness, the DSP system selected 145 surveillance points from all 31 provinces, using multi-stage stratified cluster sampling approach. All counties and cities were divided into 13 geographic regions, and then within each region, further stratified into 3 levels of urban and 4 levels of rural areas based on their economic and health status. Then urban communities and rural townships were selected as the surveillance points using probability proportionate to population size sampling (PPS) and systematic sampling in general, though with a few sites replaced considering the local capacity. A total of 58 urban points and 87 rural points were selected included in the DSP system, covering 10 million population. Since then, DSP has been generating estimates on levels and causes of mortality at national level and the six-region-residency-strata level (namely west rural, west urban, central rural, central urban, east rural, east urban). 2

In 2004 and 2005, the DSP system were expanded again to adapt to the rapidly changing mortality rate and regain representativeness. The selection of new surveillance points followed a similar multi-stage stratified cluster sampling approach as the previous expansion. All rural counties and urban cities were divided into first-level strata based on region (east, central, west). Then within each first-level stratum, counties/cities were further stratified into 9 second-level strata by tertiles of economic status (GDP for rural counties and urbanization for urban cities) and tertiles of population size. Existing DSP sites were given priority to remain in the system, and new sites were purposively selected considering local capacity and ge3161 points (63 urban points and 98 rural points) were included in DSP covering 73 million population. Evaluation of the system showed decent general representativeness at national, regional, and urban/rural level as measured by a few indicators including crude birth and death rates and population composition. 3

In 2013, the National Health and Family Planning Commission (NHFPC, previously known as Ministry of Health (MOH)) combined the DSPs and the vital registration (VR) system (hosted by previously MOH) to create an integrated national mortality surveillance system. 4 In this round of integration/expansion, disease surveillance points were selected using multi-stage stratified cluster sampling within each province. Within each of the 31 provinces in China (first-level strata), rural counties and urban districts were stratified into 8 second-level strata based on the following hierarchical structure:

1. high vs low urbanization, based on median urbanization index
2. high vs low population, based on median population size within each urbanization stratum
3. high vs low mortality, based on median total mortality rate within each urbanization-population stratum

The population indicators used in determination of the strata were from 2013 Statistical Yearbook of the Statistical Bureau of China. 7 A total of 248 province-residency-size-mortality strata were identified.

Then within each stratum, the number of surveillance sites needed for the aggregated system were sampled based on probability proportionate to population size. Conventional DSP sites were given priorities to be kept in the system, with 3 sites excluded due to low data quality and low local capacity to ensure data quality. Then 113 out of 319 sites from mortality surveillance sites in vital registration system were included. Factors like geographical balancing, not too advanced or lacking behind in human and financial resources, and local capability to manage and maintain long-term mortality surveillance work, were also taken into consideration in the final selection of surveillance sites. Along with 334 more newly selected surveillance points, current DSP dramatically increased its coverage to nearly a quarter of the population (24.3%, 323.8 million) with a total of 605 surveillance points (208 urban and 397 rural points), and achieved provincial representativeness for all estimates. 8

**Appendix 2. ICD-mapping of IMPROVE causes of death categories**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Level** | **Category code** | **Cause name** | **ICD-10** | **ICD-9** |
| 1 | AD1 | Communicable, maternal, perinatal and nutritional conditions | A00-B99, D50-D53, D64.9, E00-E02, E40-E46, E50-E68, G00, G03-G04, H65-H66, J00-J22, N70-N73, O00-O99, P00-P96 (minus P23, P37.3, P37.4), U04 | 001-041, 045-139, 243, 260–269, 278, 279.5, 280-281, 285.9, 320-323, 381-382, 460-466, 480-487, 614-616, 630–676, 760-779, 783 |
| 2 | AD2 | HIV/AIDS | B20-B24 | 042-044,279.5 |
| 2 | AD3 | Diarrhea | A00, A03, A04, A06-A09 | 001, 004, 006-009 |
| 2 | AD4 | Measles | B05 | 055 |
| 2 | AD5 | Pertussis | A37 | 033 |
| 2 | AD6 | Typhoid | A01 | 002 |
| 2 | AD7 | Meningitis/encephalitis | A39, A83-A86, B94.1, G00, G03, G04 | 036, 049.0-.049.9, 062-064, 139.1, 320-323 |
| 2 | AD8 | Malaria | B50-B54 | 084 |
| 2 | AD9 | Lower Respiratory Infections | J09-J22, P23, U04 | 466, 480-487 |
| 2 | AD10 | Tuberculosis | A15-A19, B90 | 010-018, 137 |
| 2 | AD11 | Maternal causes | O00-O99 | 630–676 |
| 2 | AD12 | Other communicable, maternal, perinatal and nutritional conditions | Remainder of AD1 | Remainder of Group AD1 |
| 1 | AD13 | Non-communicable diseases | C00-C97, D00-D48, D55-D89 (except D64.9), E03-E07, E10-E34, E65-E88, F01-F99, G06-G98 , H00-H61, H68-H93, I00-I99, J30-J98, K00-K92, L00-L98, M00-M99, N00-N64, N75-N98, Q00-Q99 | 140-242, 244-259, 270-279 (minus 279.5), 282-319 (minus 285.9), 324-380, 383-459, 470-478, 490-611, 617-629, 680-759 |
| 2 | AD14 | Congenital anomalies | Q00-Q99 | 740–759 |
| 2 | AD15 | Neoplasms | C00-D48 | 140–239 |
| 2 | AD16 | Cardiovascular disease | I00-I99 | 390-459 |
| 2 | AD17 | Endocrine, immune, and blood diseases | D55-D89 (except D64.9), E03-E07, E15-E34, E65-E88 | 240–242, 244-246, 251-259, 270-279 (except 274 and 279.5), 282-289 (except 285.9), 325, 415.1, 444, 451–453 |
| 2 | AD18 | Digestive (Gastrointestinal) disease | K20-K92 | 530-579 |
| 2 | AD19 | Other NCDs | Remainder of AD13 | Remainder of AD13 |
| 1 | AD20 | Injuries | V01-Y09, Y35-Y36, Y40-Y86, Y88-Y89, Y87.0, Y87.1 | E800-E978, E990-E999 |
| 2 | AD21 | Road traffic accidents | V01-V04, V06, V09-V80, V87, V89, V99 | E800–E819, E826-E829, E929.0-E929.1 |
| 2 | AD22 | Drownings | W65-W74 | E910 |
| 2 | AD23 | Natural disaster | X30-X39 | E900-E909 |
| 2 | AD24 | Interpersonal violence | X85-Y09, Y87.1 | E960-E969 |
| 2 | AD25 | Collective violence: legal intervention | Y35 | E970-E978 |
| 2 | AD26 | Collective violence: war | Y36 | E990–E999 |
| 2 | AD27 | Self-Harm | X60-X8, Y87.0 | E950-E959 |
| 2 | AD28 | Other injuries | Remainder of AD20 | Remainder of AD20 |
| 1 | AD29\* | Ill-defined | R00-R99, Y10-Y34, Y87.2 | 780-799, 980-989 |
| 2 | AD30\* | Injury intent not determined | Y10-Y34, Y87.2 | 980-989 |
| 2 | AD31\* | Other ill-defined | R00-R99 | 780-799 |

\*Footnote: AD29 Ill-defined causes, and its subcategories AD30-AD31 are considered garbage codes in the study. Deaths with garbage codes were essentially proportionally redistributed to known causes excluding natural disaster and collective violence.

**Appendix 3. Additional details with analysis**

**3.1 Underreporting adjustments**

**3.1.1 Underreporting survey**

Apart from the double-checking and correction during the routine data reporting in local and central CDC, DSP have been conducting retrospective quality assessment and independent underreporting survey every 3 years since 2009 in all DSP sites, to review the completeness of reported deaths and causes of death for the 3 years prior to the survey. Townships and districts were selected following certain criteria, and all households in the selected areas were included in the household survey. All deaths occurred in 3 years prior to the survey year were identified and recorded by DSP staffs from multiple data sources: household interviews, local health facilities, Family Planning Offices, Maternal and Child Health Offices, Civil Affairs Departments, and Public Security Bureaus etc.. Cases not originally captured in routine DSP reporting were defined as missed cases, and causes of death and demographic information were collected for missed cases during the household interviews. Data were then electronically reported via the underreporting survey system and underlying causes were ICD-coded by local CDC.

**3.1.2 Underreporting adjustment**

The underreporting rates (URR) were calculated at provincial level annually for year 2006-2017 using the equation below.

No clear evidence showed that the data completeness and overall quality varied across causes, thus same URRs were applied to adjust all-cause and cause-specific number of deaths.

**3.2 Sampling probability calculation and data adjustment**

We calculated the stratum-specific weights used to adjust the population and get subnational and national representative estimates for number of deaths by cause.

Table S1. Stratification for calculating sampling probabilities

|  |  |  |
| --- | --- | --- |
| Time period | 1st-level strata | 2nd-level strata |
| 2004-2012 | 31 provinces | 9-level based on GDP/urbanization and population size |
| 2013-2019 | 31 provinces | 8-level based on urbanization, population size and mortality level |

Likelihood for being sampled for people from 2nd-level stratum j (Likelihood j) for each age and gender group could be calculated using:

Where i denotes the ith surveillance point in stratum j, mj is the total number of surveillance points in stratum j, i ∈ (1,…,mj).

The total number of deaths within each stratum for each age and gender group could be calculated using:

Where ATNDj is adjusted total number of deaths for stratum j, TNDi is the total number of deaths for each surveillance point after adjusting for under-reporting.

**3.3 Crisis estimates**

**3.3.1 Death envelope and crisis**

All-cause death envelope is equivalent to all-cause number of death estimates, and is used interchangeably in this manuscript. It’s usually used when fitting the cause-specific estimates “within” the all-cause estimates’ boundaries, so the number from single causes would add up to match the all-cause totals. “Crisis-free” envelop would thus mean the total number of all causes except crisis.

Natural disasters and collective violence that severely impacted the population on a large scale are considered crisis in WHO-IGME estimates. The identification of crisis was case-by-case. 2008 Sichuan earthquake was identified as the only national natural disaster crisis during the period of 2004-2019 in China. Death estimates due to crisis are usually hard to capture, especially for countries relying on population-based surveys or sample registration system, where national estimates are inferred from sample of a small portion of the population. Therefore, crisis estimates directly from non-vital-registration data sources need to be used with cautious.

**3.3.2 Subnational estimates for 2008 Sichuan earthquake**

Given that DSP wasn’t designed to fully cover all urban and rural areas that was affected by earthquake, and considering the challenges in cause of death data collection and verbal autopsy survey post-disaster via surveillance system, we derived natural disaster estimates for four provinces affected by this crisis – Sichuan, Shaanxi, Gansu and Chongqing, from IGME age-specific crisis estimates, for year 2008.

IGME total deaths exceeds the non-crisis deaths envelope were identified as national age-specific crisis death for year 2008. Then these deaths were redistributed to Sichuan, Shaanxi, Gansu and Chongqing provinces proportional to total deaths reported in latest available government reports and mainstream news in these provinces. 5-9 Yunnan, Henan, Hubei, Guizhou and Hunan had reported deaths due to earthquakes, but was either not in age groups of interest, 10-12 or was thus were not included in the crisis estimation.

Then for urban and rural split of the deaths, we imputed percentage of urban earthquake deaths from sub-regional reporting when available, 6 otherwise used percentage of urban population instead. For Sichuan, we treated earthquake deaths in cities as urban deaths. 5 For Shaanxi, deaths in cities with top 5 GDP per capita in province as urban deaths. 6 For Chongqing, we calculated the proportion based on deaths with location information available. 8 For Gansu, we used percentage of urban population for 2008. 13 Then we used the 5-year-moving-average smoothed gender splits to further breakdown the estimates by gender.

Table S2. Subnational breakdowns for excess deaths due to 2008 Sichuan Earthquake

|  |  |  |  |
| --- | --- | --- | --- |
| **Province** | **Total earthquake death** | **Percent in national crisis death** | **Percent deaths in urban area** |
| Sichuan | 68660 | 99.27% | 45.56% |
| Shaanxi | 122 | 0.18% | 48.36% |
| Gansu | 364 | 0.53% | 60.95% |
| Chongqing | 16 | 0.02% | 45.45% |

**3.4 Annual Rate of Reduction (ARR)**

The annual rate of reduction (ARR) was calculated by:

**Appendix 4. GATHER checklist**

|  |  |  |
| --- | --- | --- |
| **Item** | **Checklist item\*** | **Section(s) or sources providing information** |
| **Objectives and funding** | | |
| 1 | Estimated indicator and population | Methods |
| 2 | Funding sources | Funding section of abstract |
| Data Inputs | | |
| For all data inputs from multiple sources that are synthesized as part of the study: | | |
| 3 | Data identification | Methods and appendix 1 |
| 4 | Inclusion/exclusion criteria | Not applicable |
| 5 | Included data sources and their main characteristics | Methods and appendix 1 |
| 6 | Potential important biases of input data | Discussion and appendix 1,2 |
| For data inputs that contribute to the analysis but were not synthesized as part of the study: | | |
| 7 | Source of other data inputs | Methods |
| For all data inputs: | | |
| 8 | Accessible input data files | Open access databases\*\* |
| Data analysis | | |
| 9 | Conceptual overview of the data analysis method | Methods and open access databases\*\* |
| 10 | Description of all steps of the analysis | Methods and appendix 1-3 |
| 11 | Model selection methods | Not applicable |
| 12 | Model performance and/or sensitivity analysis. | Not applicable |
| 13 | Uncertainty estimation methods | Methods and discussion |
| 14 | Statistical code | Open access databases\*\* |
| Results and discussion | | |
| 15 | Accessible estimates data files | Open access databases\*\* |
| 16 | Uncertainty of the estimates | Results and discussion |
| 17 | Results interpretation in light of existing evidence | Results and discussion |
| 18 | Limitations of the estimates | Discussion |

\*Detailed GATHER statement with explanation and elaboration of the items can be found on gather‐statement.org

\*\* Open access databases with input files and analytical code of the study can be accessed on github <https://github.com/y-chu/China_5-19COD>

**Reference**

1. Yang G. Selection of DSP points in second stage and their presentation. Chinese Journal of Epidemiology [Zhonghua liuxingbingxue zazhi]. 1992;13(4):197.

2. Yang G, Hu J, Rao KQ, Ma J, Rao C, Lopez AD. Mortality registration and surveillance in China: history, current situation and challenges. Popul Health Metr. BioMed Central; 2005;3(1):1–9.

3. Zhou M, Jiang Y, Huang Z, Wu F. Adjustment and representativeness evaluation of national disease surveillance points system. Disease Surveillance. Editorial Board of Disease Surveillance; 2010;25(3):239–44.

4. Liu S, Wu X, Lopez AD, Wang L, Cai Y, Page A, et al. An integrated national mortality surveillance system for death registration and mortality surveillance, China. Bull World Health Organ. 2016;94(1):46–57.

5. Tencent news. Summary of Sichuan earthquake deaths [Internet]. 2008 [cited 2021 Apr 16]. Available from: https://news.qq.com/zt/2008/dizhen/newdata.htm

6. CCTV. 122 people died and 24,300 houses collapsed in rural areas in Shaanxi due to Sichuan earthquake [Internet]. 2008. Available from: <http://news.cctv.com/china/20080613/108871.shtml>. Accessed: Apr 16 2021

7. CCTV. Summary of Sichuan earthquake deaths. 2008. Available from: <http://news.cctv.com/china/20080513/100052.shtml>. Accessed: Apr 16 2021

8. The State Council Information Office of the Peoples Republic of China. Press Conference for earthquake damage of Chongqing [Internet]. 2008. Available from: <http://www.scio.gov.cn/xwfbh/gssxwfbh/xwfbh/chongqing/document/317451/317451.htm>. Accessed: Apr 16 2021

9. Phoenix Television. Summary of Sichuan earthquake deaths [Internet]. 2008. Available from: <http://news.ifeng.com/special/0512earthquake/> Accessed: Apr 16 2021

10. Xinhua News Agency. Wenchuan Earthquake affected Yunnan Zhaotong, 1 dead and 9 injured [Internet]. 2008 [cited 2021 Apr 16]. Available from: https://news.ifeng.com/c/7fYp4h3wTwU

11. China News Service. Sichuan earthquake in Henan, 2 died and 5 injured [Internet]. 2008. Available from: <http://news.sina.com.cn/c/2008-05-13/130215529343.shtml> Accessed: Apr 16 2021

12. Xiaoxiang Morning News. People who died in Changde earthquake received 80,000 yuan in funeral expenses [Internet]. 2008. Available from: <http://epaper.xxcb.cn/xxcba/html/2008-05/20/content_33810.htm> Accessed: Apr 16 2021

13. Bureau of Statistics. Demographic characteristics of Gansu in 2008 [Internet]. 2009. Available from: <http://www.stats.gov.cn/ztjc/ztfx/dfxx/200903/t20090323_34708.html> Accessed: Apr 16 2021