

Appendices

Trans-generational changes and rural-urban disparity in household fuel use and cookstove ventilation in China: a multi-region study of 0.5 million adults

Running title: Patterns and trends in solid fuel use in China

Ka Hung Chan¹, Kin Bong Hubert Lam¹, Om P Kurmi¹, Yu Guo², Derrick Bennett¹, Zheng Bian², Paul Sherliker¹, Junshi Chen³, Liming Li^{2,4}, Zhengming Chen¹ on behalf of the China Kadoorie Biobank collaborative group

- ¹ Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK
- ² Chinese Academy of Medical Sciences, Beijing, China
- ³ China National Center for Food Safety Risk Assessment, Beijing, China
- ⁴ Department of Epidemiology, School of Public Health, Peking University Health Science Center, Beijing, China

Corresponding author:

Kin Bong Hubert Lam

Nuffield Department of Population Health, Big Data Institute Building, Old Road Campus, Oxford OX3 7LF, UK

Email: hubert.lam@ndph.ox.ac.uk

Tel: +44 (0)1865 743621

Table of Contents

Supplementary method 1. Detailed summary of study design of China Kadoorie Biobank.	3
Supplementary method 2. Example illustration of estimation of person-time of exposure to household air pollution using an age-at-risk approach	4
Figure A.1. Distribution of time started living in each house (blue) and time of resurvey interview (red)	7
Figure A.2 (a). Flow chart of number of participants at each stage of analysis—recall period of baseline survey	8
Figure A.2 (b). Flow chart of number of participants at each stage of analysis—first and second resurveys	9
Figure A.3. Regional variation of types of (a) cooking fuel and (b) heating fuel; and (c) presence of cookstove ventilation.....	10
Figure A.4. Temporal pattern of self-reported primary fuel type for cooking and presence of cookstove ventilation (combined) by study time in rural and urban areas.....	11
Figure A.5. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by education level in rural areas.....	12
Figure A.6. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by education level in urban areas	13
Figure A.7. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by household income level in rural areas	14
Figure A.8. Temporal pattern of self-reported fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by household income level in urban areas	15
Figure A.9. Temporal pattern of self-reported primary cooking fuel type in 10 study regions by time	16
Figure A.10. Temporal pattern of self-reported primary heating fuel in eight study regions by.....	17
Figure A.11. Temporal pattern of cookstove ventilation availability in 10 study regions by time	18

Supplementary method 1. Summary of study design of the China Kadoorie Biobank.

In 2004-08, the China Kadoorie Biobank (CKB) recruited 512,891 adults aged 30-79 years from five rural counties and five urban cities of China. Each of the 10 study sites chosen has a typical population of 0.5 million, and they were selected from China's nationally representative Disease Surveillance Points system¹ based on population stability, quality of death and disease registries, and local commitment and capacity as well as to represent a diverse range of socioeconomic levels, risk exposures and disease patterns. However, the study is not designed to be representative to the general population in China.²

For the urban study sites, recruitment was carried out in major cities of the wider region (Qingdao of Shandong Province, Harbin of Heilongjiang Province, Haikou of Hainan Province, Suzhou of Jiangsu Province, and Liuzhou of Guangxi Zhuang Autonomous Region). On the other hand, recruitment in the rural study sites were undertaken in scattered rural villages surrounding a less well-known prefecture- or county-level city of the wider region (Pengzhou of Sichuan Province, Tianshui of Gansu Province, Huixian of Henan Province, Tongxiang of Zhejiang Province, and Liuyang of Hunan Province). For the ease of expression, the CKB studies address the urban study sites by the city and rural sites by the province. Of note, 10-15 years ago Suzhou was relatively rural but it has since urbanised rapidly. Nevertheless, many people still maintained a traditional way of life, with some baseline characteristics (e.g. education level, smoking patterns) similar to those in other rural study sites. Therefore, although Suzhou is classified as one of the urban sites in CKB, it can be considered as semi-urban in a wider context.

At each study site, 100-150 administrative units (village or street committee) were randomly selected for participant recruitment. At each administrative unit selected, all permanent residents aged 35-74 and without major disability were identified through official residential records, and invitation letters were delivered by local community leaders or health workers to encourage participation in the study. After contacting 1,801,200 registered residents, 512,869 participated, including 12,665 slightly outside the target age range, but they were not turned away so as to enhance community commitment (resulting in the 30-79 years actual baseline age range).

After obtaining informed consent, each participant undertook a range of physical measurements (e.g. blood pressure, lung function) and a questionnaire interview administered by trained health workers, and a 10-ml blood sample was collected following standardised procedures. The questionnaire consisted of 10 different sections assessing socio-demographic characteristics, lifestyle behaviours (e.g. diet, physical activity, smoking, household air pollution), medical history, reproductive history and so on.

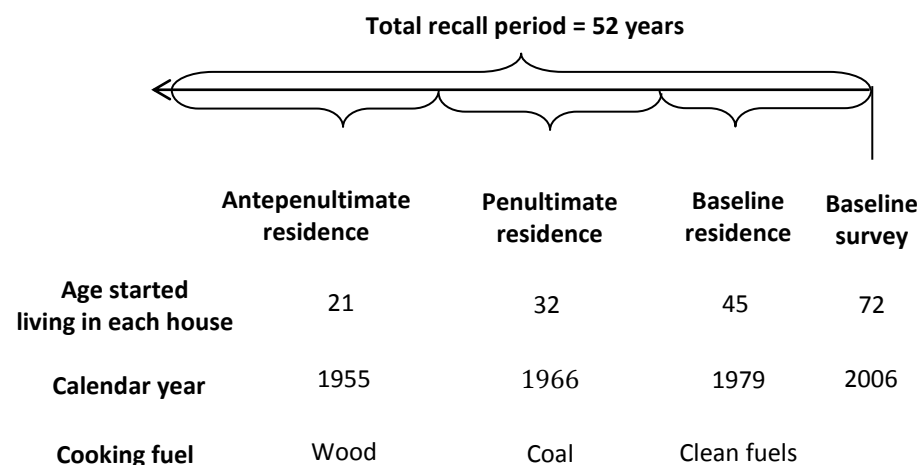
After the completion of the baseline survey, two resurveys were undertaken in 2008 and 2013-14, involving 4-5% of randomly selected surviving participants. The resurveys were designed to repeat the assessment (questionnaire and physical measurements) made at baseline, primarily to enable correction for "regression dilution bias"³ for less biased assessment of prospective associations between long-term usual levels of particular risk exposures and disease, but also allows investigation into the changing pattern of exposure such as household air pollution. Additional data (e.g. more detailed measurement of dietary patterns) were also collected in the resurveys to calibrate and enhance the baseline data.

For further information about CKB, please visit <http://www.ckbiobank.org/>.

Supplementary method 2. Example illustration of estimation of person-time of exposure to household air pollution using an age-at-risk approach

a) Scenario:

A participant aged 72 at baseline survey recalled 27 years of residence in the baseline residence, 14 years in the penultimate residence, 11 years in the antepenultimate residence (i.e. total recall period = 52 years). For cooking fuel, the participant reported using gas (i.e. clean fuels) in baseline residence, coal in penultimate residence and wood in the antepenultimate residence.



b) Summary of age lived through and exposure to cooking fuel in the three most-recent houses:

Houses	Age lived through each house(year-old)	Primary cooking fuel in each house*		
		Wood	Coal	Clean fuels
Antepenultimate house	21-31	1	0	0
Penultimate house	32-44	0	1	0
Baseline house	45-72	0	0	1

*For primary cooking fuel, 1 and 0 indicate EXPOSED and UNEXPOSED, respectively.

c) Proportion of age band that was informative:

A series of 5-year age bands was created from birth to 74 year-old. For the hypothetical scenario, the recall period covered 52/72 years of the participant's lifetime (i.e. 21-72 years old). The age bands where information of HAP exposure is available are the 'informative age bands'. The proportion of age band that was informative was calculated as illustrated below.

	Non-informative age bands				Informative age bands										
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Proportion of age band that was informative	0	0	0	0	0.8	1	1	1	1	1	1	1	1	1	0.6

There was no information provided for 0-19 year-old; With recall information from 21-72 year-old, 80% of the age band 20-24 and 60% of the age band 70-74, and all the age bands in between were considered informative.

d) Estimating duration of exposure using informative age bands and exposure recall data:

Then, the crude proportion of informative age band exposed to each category of cooking fuel was estimated. For example, from 21 to 31 year-old, the hypothetical participant exposed to wood as primary cooking fuel and therefore the crude proportion of exposure to wood in age band 20-24 was 0.8 (4 out of 5 years).

	Non-informative age bands				Informative age bands										
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Proportion of age band that was informative	0	0	0	0	0.8	1	1	1	1	1	1	1	1	1	0.6
Exposure to wood	0	0	0	0	0.8	1	0.4	0	0	0	0	0	0	0	0
Exposure to coal	0	0	0	0	0	0	0.6	1	0	0	0	0	0	0	0
Exposure to clean fuels	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1

e) Calculating percentage of length of informative age band being exposed to each HAP category:

The length of the first and last informative age band varied and ranged from 1-5 years (as shown in (d)). To allow valid comparison across individuals, a method of standardisation of the period of exposure to each HAP category was used to calculate the percentage of length of informative age band being exposed to each HAP category with the formula below.

Percentage of length of informative age band exposed to each HAP category = $\frac{\text{Crude proportion exposed}}{\text{Proportion of age band that was informative}} \times 100\%$

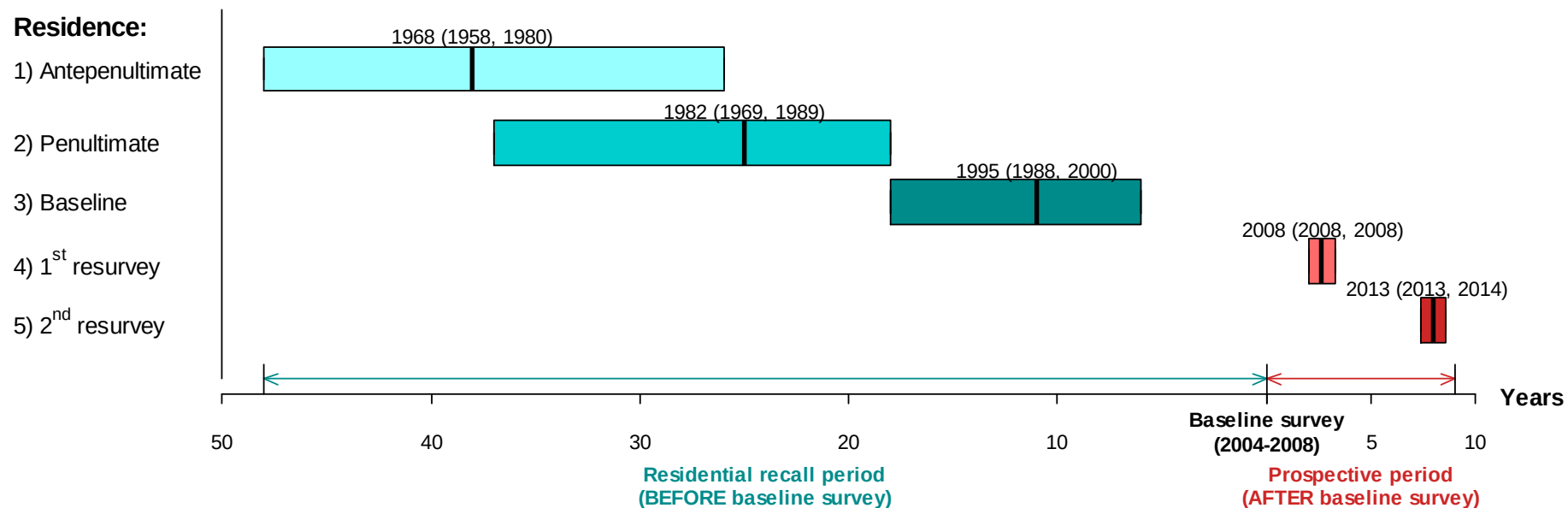
For the hypothetical participant, the percentage of length of age band 20-24 year-old exposed to wood = $0.8/0.8 \times 100\% = 100\%$; whereas that of age band 30-34 = $0.4/1 \times 100\% = 40\%$ (the remaining 60% of this complete age band was exposed to coal).

	Non-informative age bands				Informative age bands											
	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	
Proportion of age band that was informative	0	0	0	0	0.8	1	1	1	1	1	1	1	1	1	0.6	
Percentage of time exposed to wood	0	0	0	0	100%	100%	40%	0	0	0	0	0	0	0	0	
Percentage of time exposed to coal	0	0	0	0	0	0	60%	100%	0	0	0	0	0	0	0	
Percentage of time exposed to clean fuels	0	0	0	0	0	0	0	0	100%	100%	100%	100%	100%	100%	100%	

f) Summary

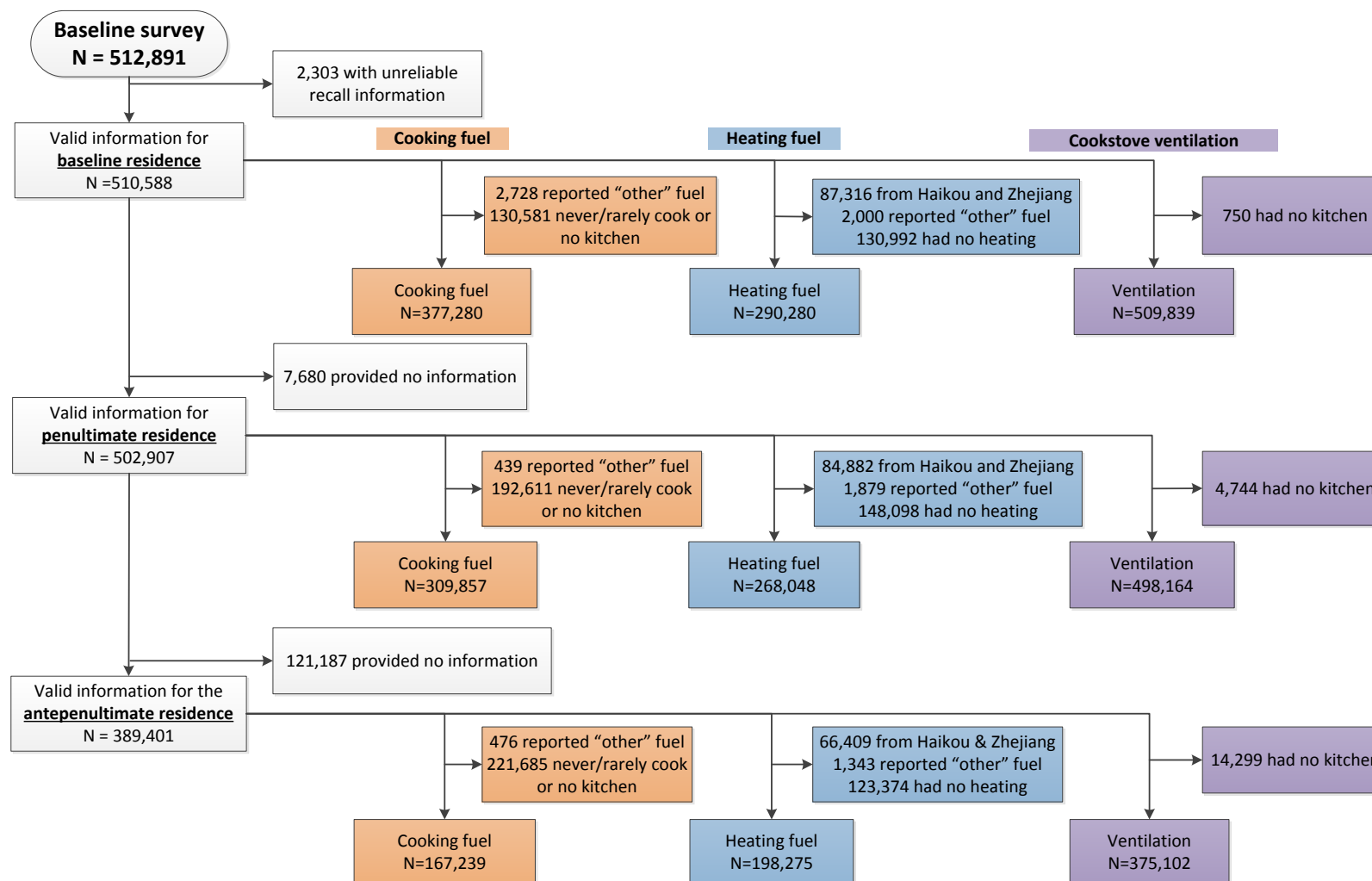
Procedures c) to e) were applied to all participants for each HAP exposure indicator (primary cooking fuel, primary heating fuel and cookstove ventilation). The rural and urban participants were then divided into five birth cohorts according to year of birth: before 1939, 1940-49, 1950-59, 1960-69 and after 1970 to depict the overall historical development of HAP exposure across multiple generations in rural and urban areas.

Figure A.1. Distribution of time started living in each house (blue) and time of resurvey interview (red) ^a



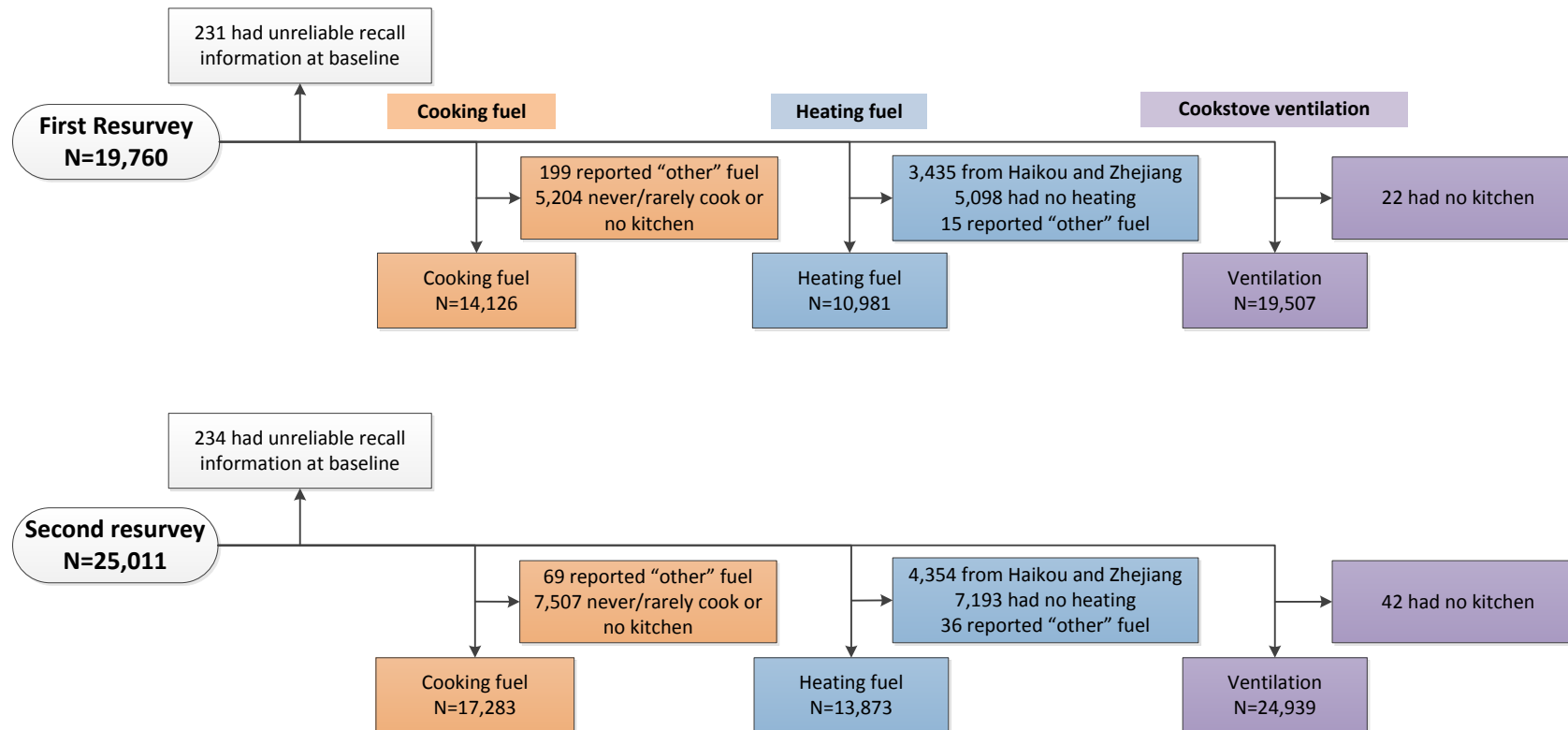
^a Box plots displaying hinges at 75th and 25th percentiles and vertical bold-line at median; four-digit numbers above median lines are equivalent calendar years of median (interquartile range); numbers of two-headed arrows are overall mean $\pm$ standard deviation of total recall and prospective periods in years counted from the date of the baseline survey (conducted in 2004-08)

Figure A.2 (a). Flow chart of number of participants at each stage of analysis—recall period of baseline survey ^a



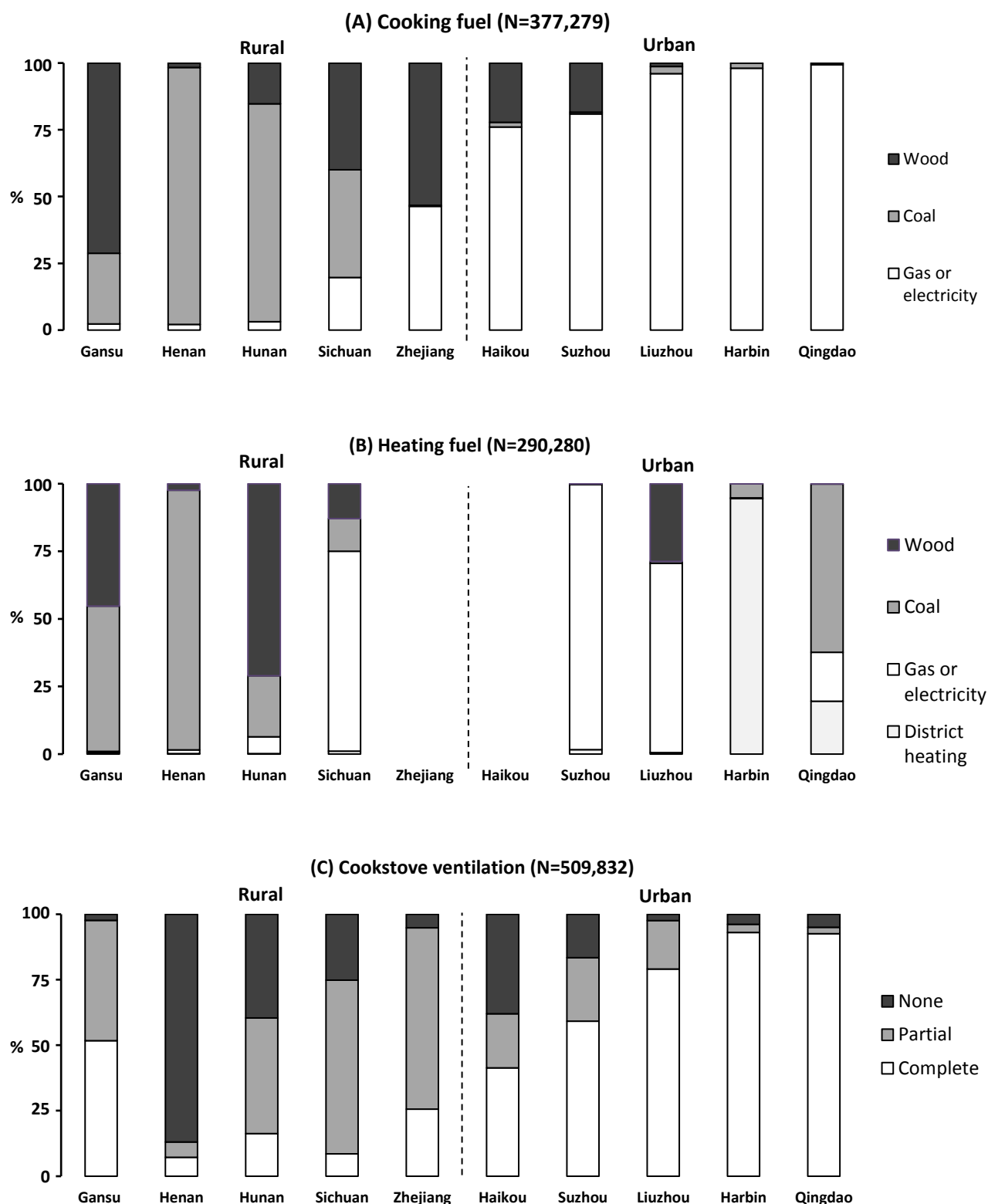
^a Recall period covered the three most recent residences counted from the time of baseline survey. Participants whose recall period was greater than their age at baseline were first excluded from all analyses; for cooking fuel, participants reported using "other" fuel or had never/ rarely cook or no kitchen were excluded; for heating fuel, participants from Haikou and Zhejiang those reported using "other" fuel or had no heating (in the rest of the regions) were excluded; for cookstove ventilation, participants reported having no kitchen were excluded.

Figure A.2 (b). Flow chart of number of participants at each stage of analysis—first and second resurveys ^a



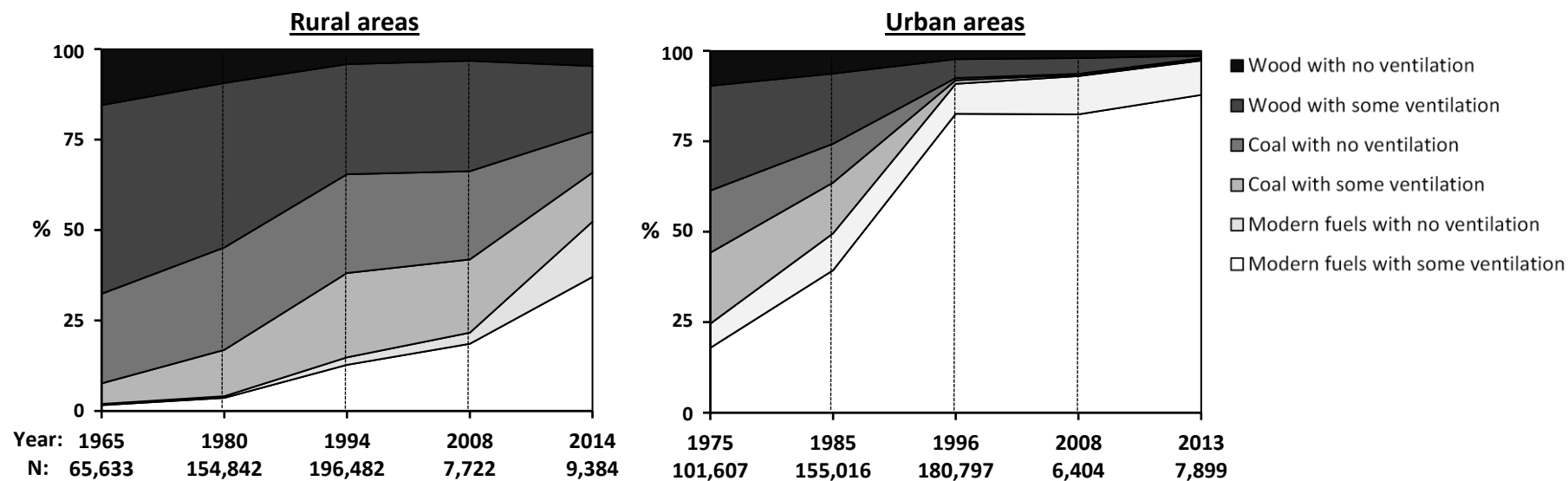
^a The first and second resurvey were conducted in 4-5% randomly sampled participants of baseline survey. Participants whose recall period was greater than their age at baseline survey were first excluded from all analyses; for cooking fuel, participants reported using "other" fuel or had never/ rarely cook or no facility were excluded; for heating fuel, participants from Haikou and Zhejiang those reported using "other" fuel or had no heating (in the rest of the regions) were excluded; for cookstove ventilation, participants reported having no kitchen were excluded.

Figure A.3. Regional variation of types of (a) cooking fuel and (b) heating fuel; and (c) presence of cookstove ventilation ^a



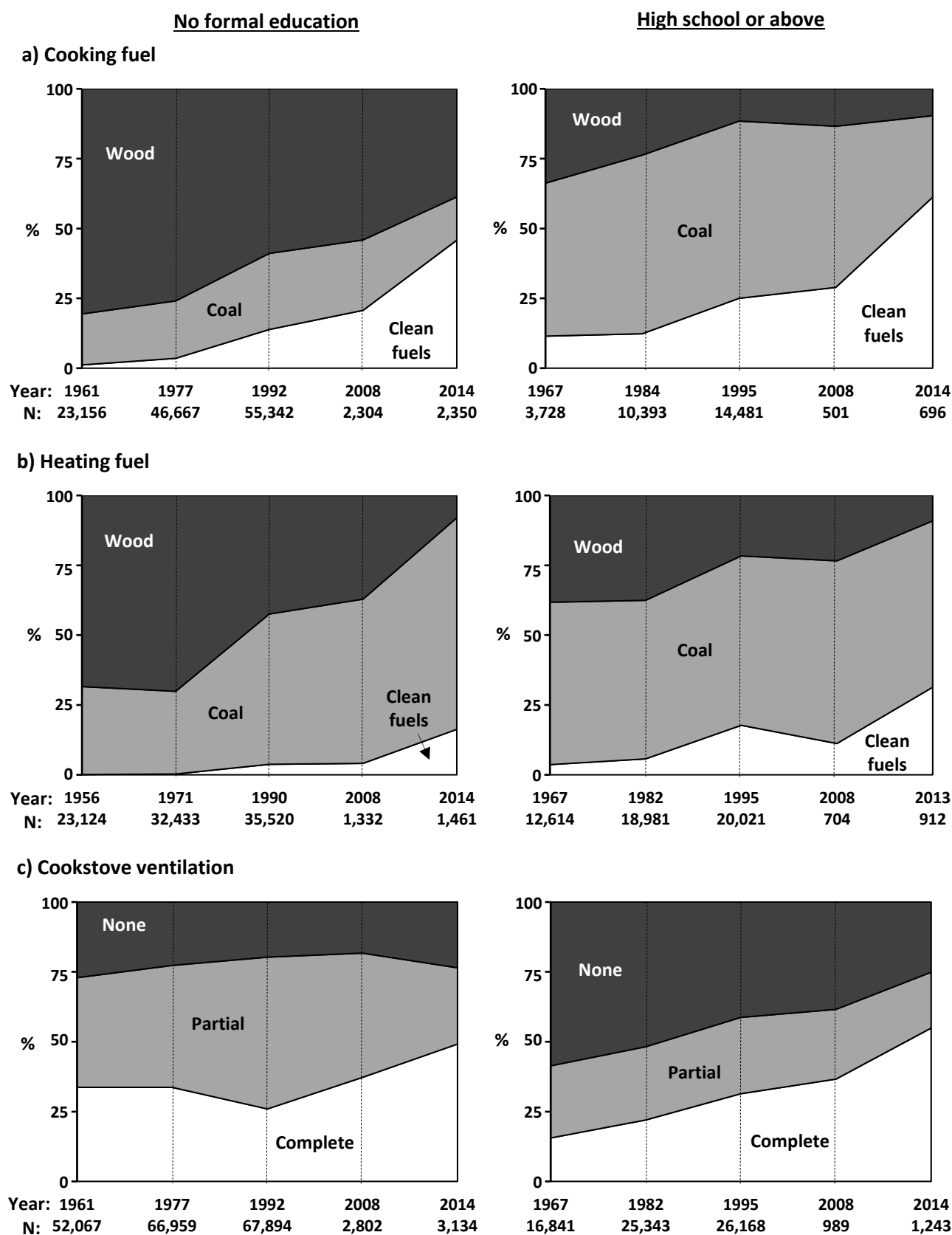
^a Prevalence estimates are adjusted for age and gender. Each chart is divided into two sections, with rural regions on the left and urban regions on the right. Regions are ordered by the percentages of solid fuel used for cooking in descending order from left to right. Participants from Haikou and Zhejiang were excluded from the analysis on heating fuel.

Figure A.4. Temporal pattern of self-reported primary fuel type for cooking and presence of cookstove ventilation (combined) by study time in rural and urban areas ^a



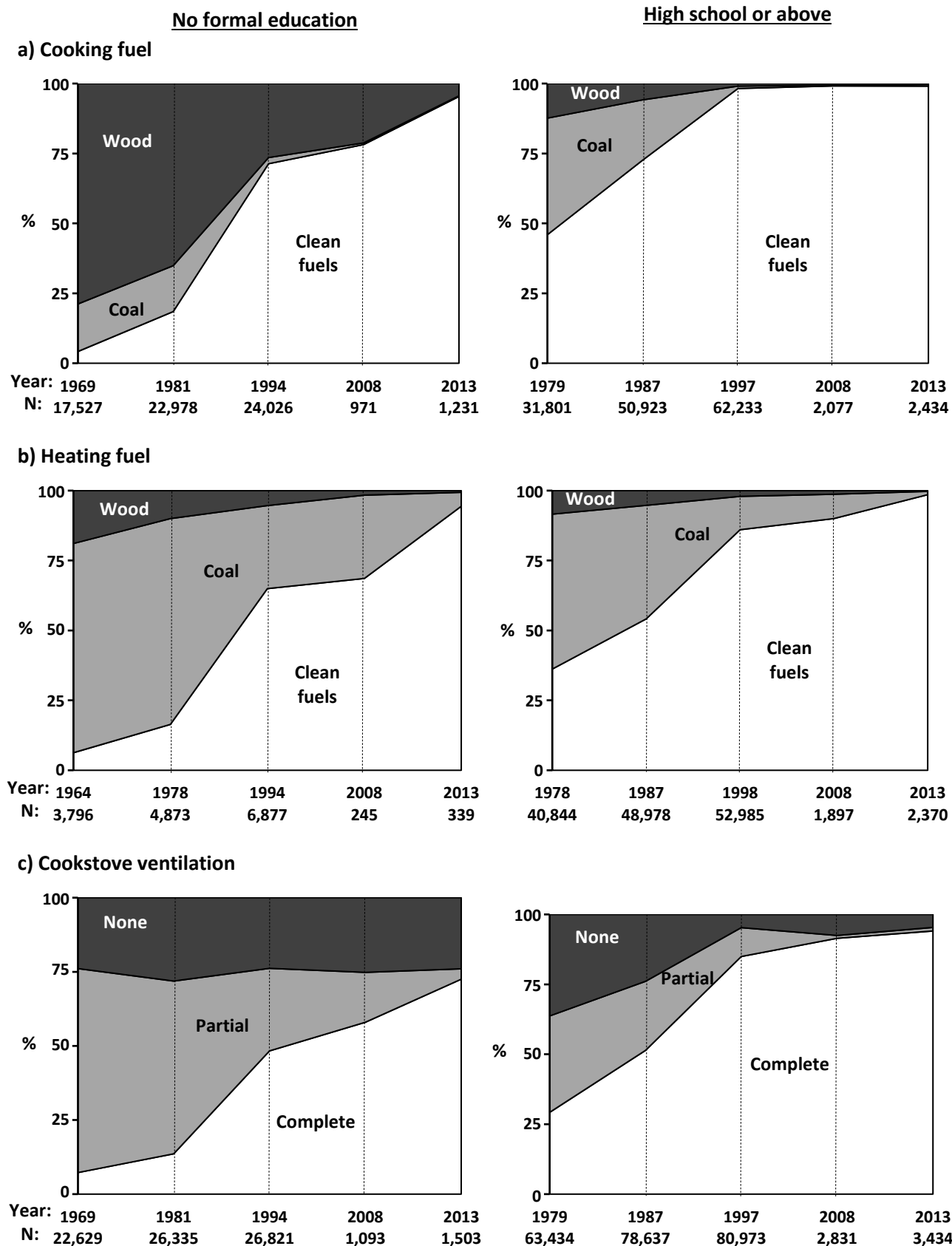
^a Percentage estimates were adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and the calendar year when 1st or 2nd resurvey was conducted; and 2) number of participants that provided information at respective time period/point. Clean fuels include gas and electricity. Participants who never/rarely cooked, reported using 'other' fuel type, or had no kitchen were excluded.

Figure A.5. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by education level in rural areas ^a



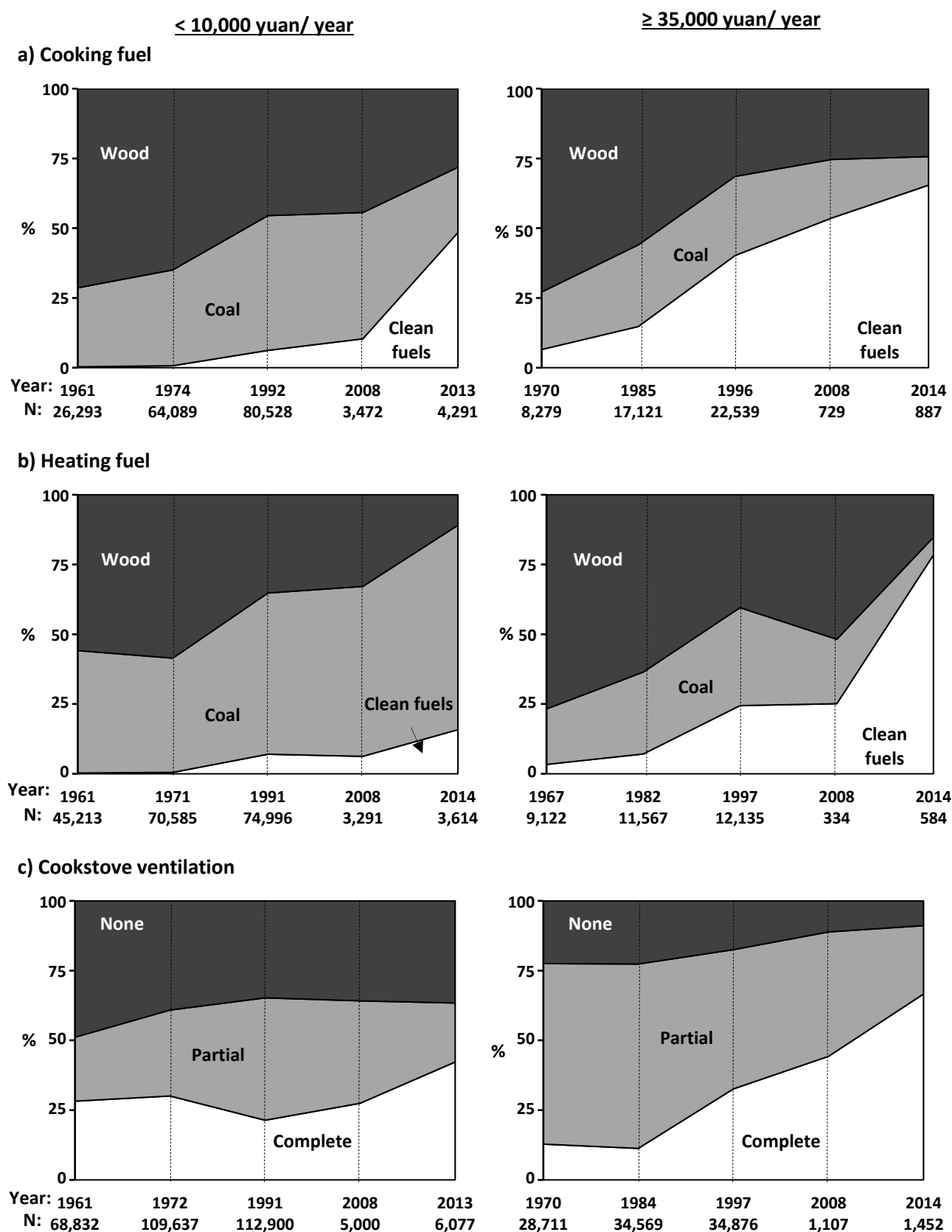
^a Percentage estimates were adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and the calendar year when 1st or 2nd resurvey was conducted; and 2) number of participants that provided information at respective time period/point. Clean fuels include gas, electricity, and district heating (for heating only). Participants who never/rarely cooked, reported using 'other' fuel type, had no kitchen were excluded from (a); participants who had no heating or from Zhejiang and Haikou and those reported using 'other' fuel type were excluded from (b); participants who had no kitchen were excluded from (c).

Figure A.6. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by education level in urban areas ^a



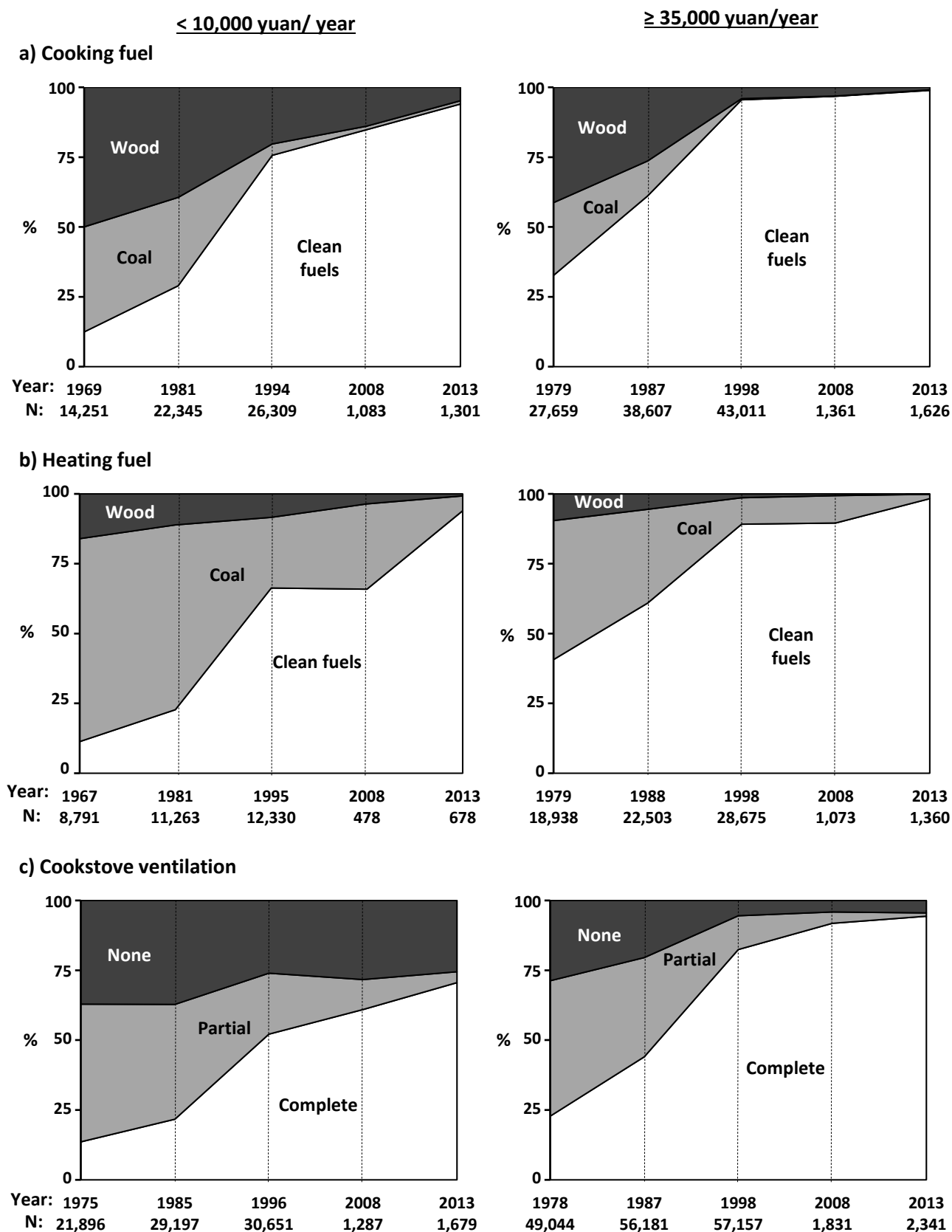
^a Percentage estimates were adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and the calendar year when 1st or 2nd resurvey was conducted; and 2) number of participants that provided information at respective time period/point. Clean fuels include gas, electricity, and district heating (for heating only). Participants who never/rarely cooked, reported using 'other' fuel type, had no kitchen were excluded from (a); participants who had no heating or from Zhejiang and Haikou and those reported using 'other' fuel type were excluded from (b); participants who had no kitchen were excluded from (c).

Figure A.7. Temporal pattern of self-reported primary fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by household income level in rural areas ^a



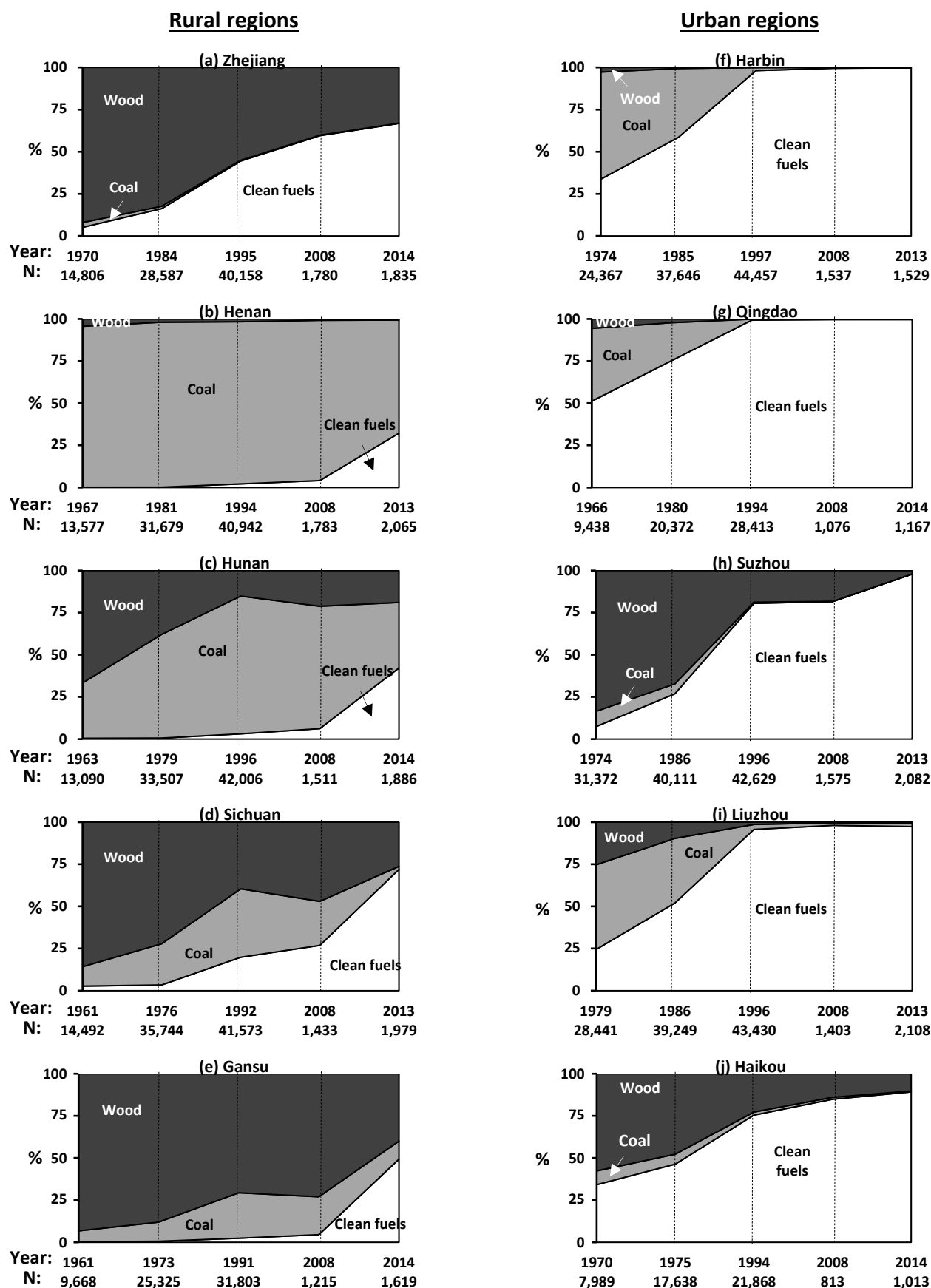
^a Percentage estimates were adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and the calendar year when 1st or 2nd resurvey was conducted; and 2) number of participants that provided information at respective time period/point. Clean fuels include gas, electricity, and district heating (for heating only). Participants who never/rarely cooked, reported using 'other' fuel type, had no kitchen were excluded from (a); participants who had no heating or from Zhejiang and Haikou and those reported using 'other' fuel type were excluded from (b); participants who had no kitchen were excluded from (c).

Figure A.8. Temporal pattern of self-reported fuel type for a) cooking and b) heating; and c) presence of cookstove ventilation by study time by household income level in urban areas ^a



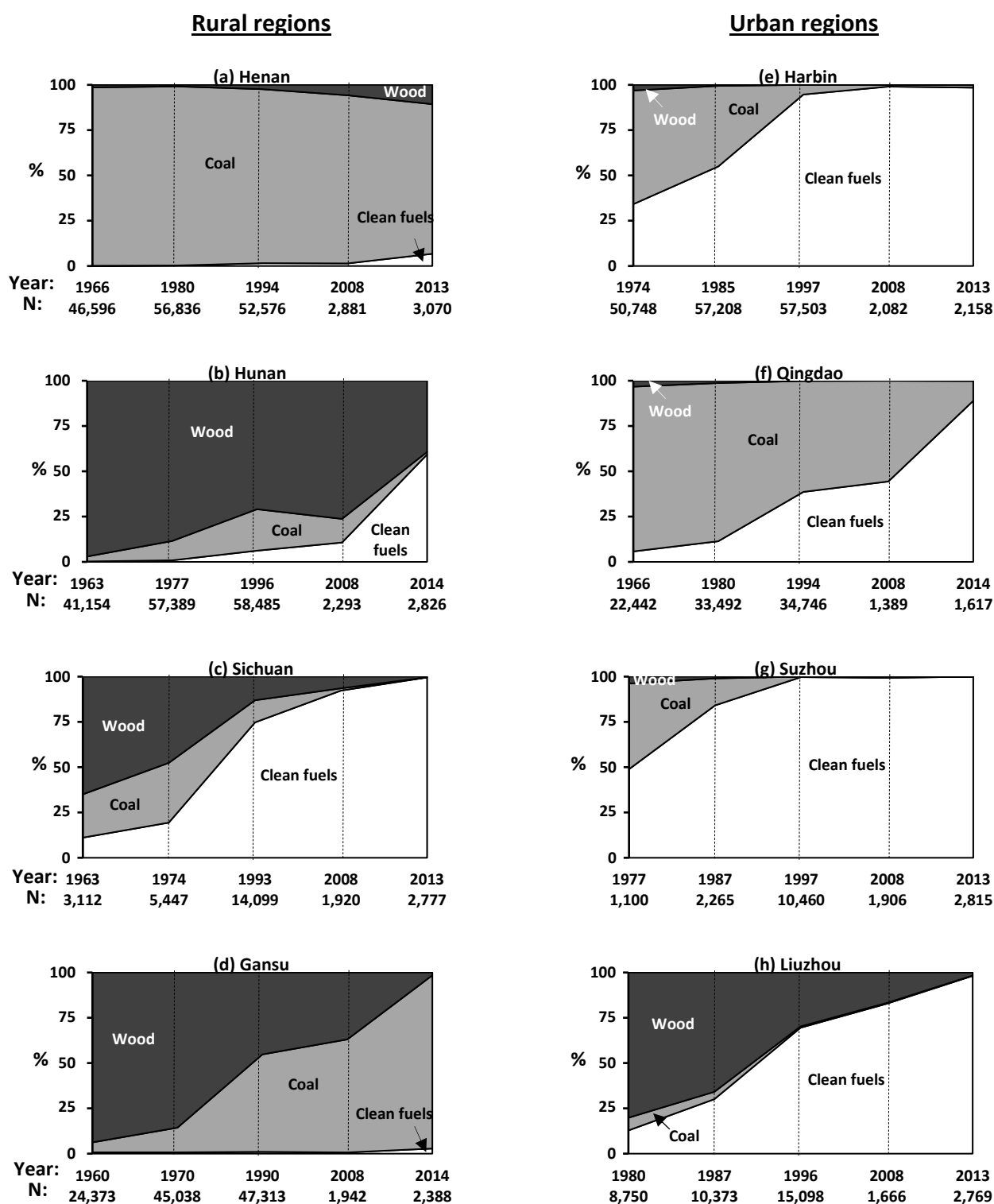
^a Percentage estimates were adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and the calendar year when 1st or 2nd resurvey was conducted; and 2) number of participants that provided information at respective time period/point. Clean fuels include gas, electricity, and district heating (for heating only). Participants who never/rarely cooked, reported using 'other' fuel type, had no kitchen were excluded from (a); participants who had no heating or from Zhejiang and Haikou and those reported using 'other' fuel type were excluded from (b); participants who had no kitchen were excluded from (c).

Figure A.9. Temporal pattern of self-reported primary cooking fuel type in 10 study regions by time ^a



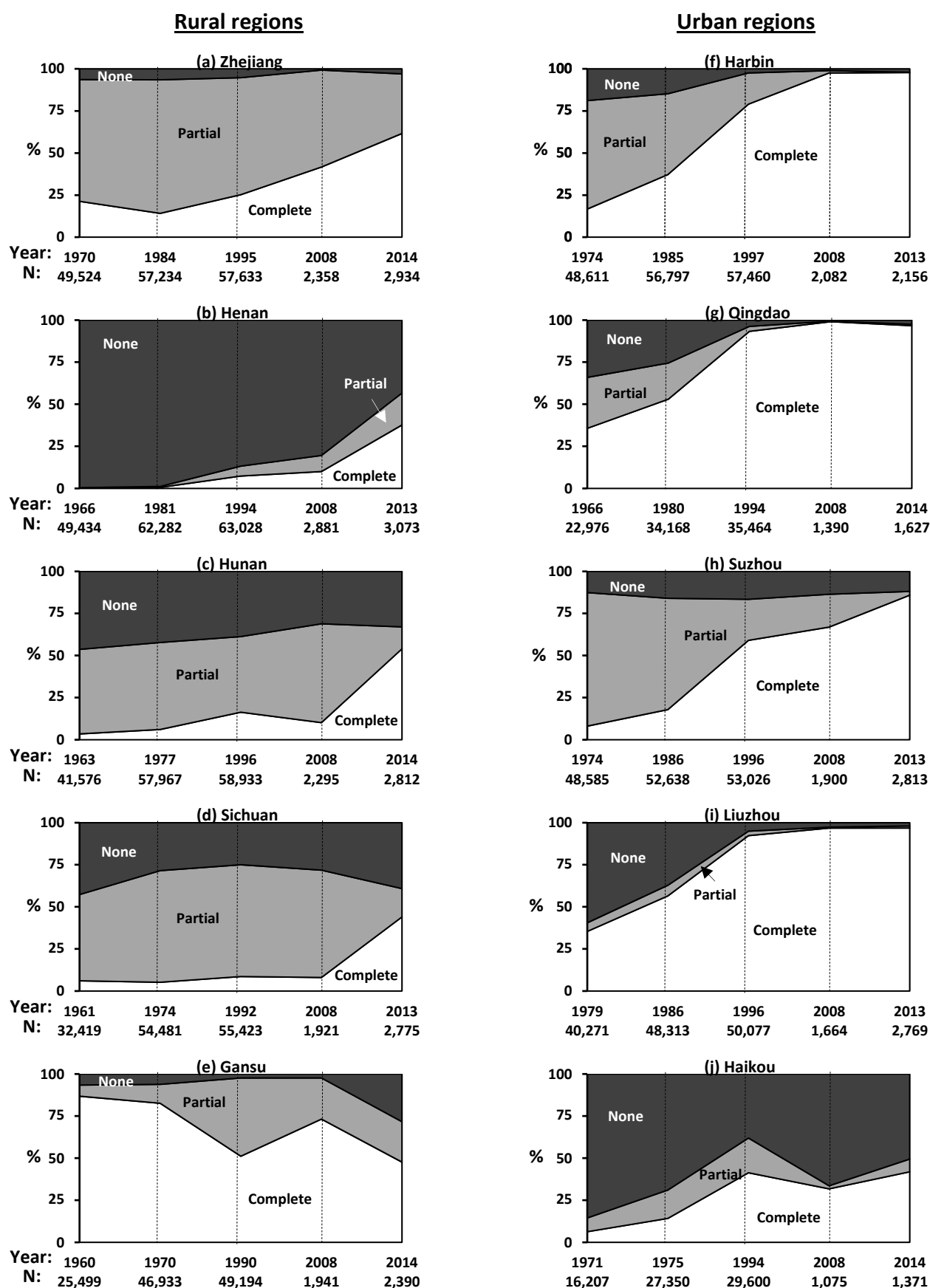
^a Percentage estimates are adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and year when 1st and 2nd resurvey is conducted; and 2) number of participants that provided the respective information. Clean fuels include gas and electricity. Rural regions are ordered from east to west; urban regions are from north to south.

Figure A.10. Temporal pattern of self-reported primary heating fuel in eight study regions by time ^a



^a Percentage estimates are adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and year when 1st and 2nd resurvey is conducted; and 2) number of participants that provided the respective information. Clean fuels include gas, electricity, and district heating. Haikou (an urban region) and Zhejiang (a rural region) were excluded because >99% of participants reported no heating at baseline. Rural regions are ordered from east to west; urban regions are from north to south.

Figure A.11. Temporal pattern of cookstove ventilation availability in 10 study regions by time ^a



^a Percentage estimates are adjusted for age, gender, and regions. X-axis labels are 1) median year started living in each of the three most recent residences recalled in baseline survey and year when 1st and 2nd resurvey is conducted; and 2) number of participants that provided the respective information. Participants who had no kitchen were excluded. Rural regions are ordered from east to west; urban regions are from north to south.

References:

Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. 2011. China kadoorie biobank of 0.5 million people: Survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 40:1652-1666.

Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. 1999. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *Am J Epidemiol* 150:341-353.

Yang GH, Stroup DF, Thacker SB. 1997. National public health surveillance in china: Implications for public health in china and the united states. *Biomedical and environmental sciences* : BES 10:1-13.