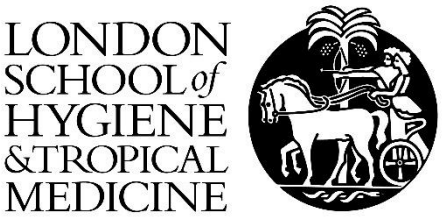


Influences of place of residence on risk factors for atherosclerotic cardiovascular diseases in South India

Tina Bonde Sørensen, MSc ANutr

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

May 2020

Department of Population Health

Faculty of Epidemiology and Population health

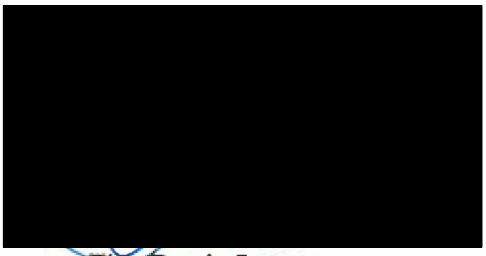
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE (LSHTM)

University of London

Funded by:	The Bloomsbury Colleges PhD studentship, Wellcome Trust, the Newton-Bhabha Fund and the LSHTM Research Degree Traveling Scholarship through the London School of Hygiene and Tropical Medicine
Research group affiliation:	The Andhra Pradesh Children and Parents Study (APCAPS) research group, India and United Kingdom

STATEMENT OF OWN WORK

I Tina Bonde Sørensen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis



Tina Bonde Sørensen

25 May 2020

ABSTRACT

There has been much debate about the role of place of residence (rural vs. urban) on non-communicable disease outcomes but the potential pathways are relatively poorly quantified. This PhD explores the links between place of residence and cardiovascular diseases (CVDs) in rapidly urbanising India in three connected research papers.

The first paper systematically reviewed the available published evidence on the associations of engaging in agriculture with CVD incidence and prevalence and CVD risk factors in low- and middle-income countries (LMICs). The review included 15 observational studies, and high-quality evidence was lacking. Thirteen studies from five LMICs suggested that agricultural workers living in rural areas had a lower prevalence of some important CVD risk factors (hypertension and high body mass index [BMI]) but higher prevalence of others (smoking and underweight) than non-agricultural workers mainly living in urban areas.

Building on these initial findings, the second paper estimated the association of urbanisation level with a range of CVD risk factors using data from the third wave of the Andhra Pradesh Children and Parents Study (APCAPS) (n=6236). Remote sensing night-time light intensity (NTLI) data (unitless digital numbers) provided a continuous proxy measure of levels of urbanisation for 27 APCAPS villages. Mixed-effects linear regression models with log-transformed NTLI were used in analysis. Increasing NTLI was associated with rises in mean BMI and systolic blood pressure (SBP), but not low-density lipoprotein (LDL) or fasting plasma glucose (FPG), after adjusting for confounders and these increases were greater among participants aged above 40 years.

Using this dataset in the third research paper, the total effect of increasing urbanisation level on mean SBP was decomposed into direct and indirect effects via hypothesised pathways. Mediation analysis was performed using mixed-effects linear regression models for SBP, log-transformed NTLI (that ranged from 4.1 to 7.0) and three composite mediators summarising (i) socio-demographic (e.g. occupation and education), (ii) lifestyle, mental health (e.g. diet and depression), and (iii) metabolic factors (e.g. BMI and LDL). All models were gender-stratified and adjusted for age and other confounders. Mean SBP was 122.7mmHg (±15.7) among men and 115.8mmHg (±14.2) among women. A one unit (integer) increase in log-NTLI was associated with a rise in SBP by 2.0mmHg (95% CI 0.4, 3.5) among men and 1.3mmHg (95% CI 0.006, 2.6) among women. A considerable indirect effect via the metabolic pathway elevated SBP among men by 4.6mmHg (95% CI 2.0, 7.3) and a smaller SBP rise among women by 0.7mmHg (95% CI 0.1, 1.3) per one log-NTLI increase. Among men, but not women, NTLI

acted indirectly via the lifestyle and mental health pathway to elevate SBP by 0.7mmHg (95% CI 0.1, 1.3) per one log-NTLI increase. The total effect among both genders and the indirect effect via metabolic factors among men approximately doubled among participants aged above 40 years.

This PhD identified night-time light intensity as a potentially important continuous proxy indicator of urbanisation levels in India and formally tested potential causal pathways linking urbanisation level with CVD risk factors. Increasing level of urbanisation was associated with greater mean SBP and BMI at early stages of urbanisation in South India. The findings offer new insights into possible pathways through which urbanisation may act on CVD risk factors. Pathways via metabolic factors independent of socio-demographic, lifestyle and mental health factors emerged as particularly interesting. These findings identified a need to understand better the indirect effects of urbanisation-related upstream determinants on CVD risk factors in India, independent of socio-demographic, lifestyle and mental health factors. Mediation analysis may be a useful approach to inform strategies to mitigate the expected large public health and economic impacts from continued, rapid urbanisation in India. I recommend implementing and scaling available, evidence-based, population-wide, primordial prevention strategies in India that target upstream determinants to modify behavioural risk factors for NCDs. In line with WHO's best buys for NCD prevention and recommendations of the WHO and the Climate and Clean Air Coalition, strategies could include (but are not limited to), advertisement restriction, health promotion campaigns; taxation of alcohol, petrol and food content; improved enforcement of the Cigarettes and Other Tobacco Products Act and shifts to low-emission transport options. Evaluation of the ongoing Smart City Mission could identify innovative and effective interventions for cleaner, greener, and more environmentally sustainable urban development throughout India. In order to tackle expected rises in CVDs with urbanisation in India (and elsewhere in LMICs), I further call for a new interdisciplinary urbanisation science and collaborations to scale-up national surveillance systems and periodical surveys on environment, demographics, phyco-social factors and NCDs.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the many advisors, colleagues, collaborators, friends, family and loved ones, to whom I owe the completion of my PhD.

Firstly, I am incredibly grateful to my primary supervisor, Professor Alan D Dangour (LSHTM), for taking me on, believing in me, and always making time for me. Thank you for your invaluable support, guidance, constructive criticism, warm encouragement, and for occasionally telling me to go home. Secondly, a great thanks to my secondary supervisor Professor Bhavani Shankar (SOAS), for your insightful inputs throughout my PhD process, and not least, your compassion, support and encouragement during periods of hardship.

I would like to extend my sincere gratitude to the brilliant statisticians I have had the pleasure of working with and learning so much from. A special thanks to Professor Stijn Vansteelandt (Ghent University and LSHTM), without whom I could not have done the mediation analysis. Your patience and willingness to share your wealth of knowledge on causal inference truly made a world of difference to my work and understanding of statistics. Secondly, a big thanks to Dr John Gregson (LSHTM) for supporting me throughout my PhD process, always taking time for my statistics questions and reassuring me when I felt overwhelmed. I would also like to express my gratitude to Associate Professor Karla Diaz-Ordaz (LSHTM) and Associate Professor Daniel Altmann (LSHTM) who kindly helped me get to grips with multiple imputations and matrix expressions. While on the topic of statistics, I also want to offer my warm thanks to Dr Ruth Farmer (LSHTM) and Stephen Nash (LSHTM) for hours of fantastically nerdy and incredibly helpful conversations and countless laughs.

I would like to express my appreciation to Professor Sanjay Kinra (LSHTM), for providing me the opportunity to do my PhD research as part of the Andhra Pradesh Children and Parents Study team. Your hard questions have pushed me to be a critical and rigorous researcher and think ‘outside of the box’. A huge thanks to Dr Robin Wilson (University of Southampton) for your invaluable work on extracting and preparing the night-time light intensity data for the APCAPS villages and your manuscript contributions; to Dr Mika Matsuzaki (LSHTM) for your contributions to the review; to Associate Professor Suneetha Kadiyala (LSHTM) for your advice on agriculture, nutrition and health, manuscript edits and support during my time in India; to Dr Shelby Yamamoto for your support during the early stages of my work; and to Dr Rosemary Green for lending me your skills on finite fixture modelling and not least, for lending me your wizzy computer.

I am deeply grateful to my funders, the Bloomsbury Colleges PhD studentship, the Wellcome Trust, the Newton-Bhabha Fund and the LSHTM research degree traveling scholarship.

A big thanks also to Jenny Fleming, Lauren Dalton and the IT team at LSHTM for your administrative and IT support, and not least your relentless chasing of wizzy computers on my behalf.

I owe important thanks to the APCAPS team in Hyderabad who welcomed me and looked after me during my time in India. It has been a pleasure working with you and getting to know you. I also owe a heartfelt thanks to the wonderful people I met in India outside of work, who made my time there very special. A special thanks to Rajit, Sonu, Karthik and co, Ankita, Kiruthika and Richa.

I feel incredibly lucky to have met and shared this journey with a number of kind and inspiring people in London. A special thanks to Ruth, Stephen, Phil, Sophie, Rebecca and Rohini. I also owe my deepest gratitude to the incredible and most patient people outside my PhD bubble, Lionel, Katrine, Giulia, Marion and Natalie, for your unwavering support, keeping me sane, celebrating my achievements (and my failures) and making me laugh.

Last, but not least, a huge thanks to my family -mum, dad, Christian, Karin and my fantastic nephews. My especially heartfelt gratitude goes to my brother, the calmest and kindest person I know, and the amazing Karin, without whom none of this would have happened. Tusind tak!

Update on corrected version (27/02/2020): I would also like to extend a big thank you to my examiners, Dr Anna Pearce and Dr Amitava Banerjee, for taking the time for my viva and for their insightful comments that helped me strengthen the thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APCAPS	Andhra Pradesh Children and Parents Study
BMI	body mass index
C	set of confounders
CHD	Coronary Heart Disease
CI	confidence interval
CRP	C-reactive protein
CVD	cardiovascular disease
d	day
DALYs	disability adjusted life years
DMSP-OLS	Defence Meteorological Satellite Programme's Operational Linescan System
FPG	fasting plasma glucose
h	hour(s)
HDL	High-density lipoprotein
HHDs	hypertensive heart diseases
HNT	Hyderabad Nutrition Trial
ICDS	the Integrated Child Development Service
IHD	ischemic heart disease
IQR	inter quartile range
LDL	low-density lipoprotein
LMIC	low- and middle-income country
LSHTM	London School of Hygiene and Tropical Medicine
M	composite mediator

MET	metabolic equivalent tasks
MI	myocardial infarction
n	sample size
na	not available/not applicable
NCD	non-communicable disease
NHP	National Health Policy
NPCDCS	The National Program for Prevention and Control of Cancer, Diabetes, CVD and Stroke
NTLI	night-time light intensity
OR	odds ratio
PAL	physical activity level
PAR	physical activity ratio
PR	prevalence ratio
SBP	systolic blood pressure
sd	standard deviation
SES	socio-economic status
UK	United Kingdom
UNPF	United Nations Population Fund
y	years
β	beta-coefficient

STRUCTURE OF THE THESIS

This research paper-style thesis is prepared in accordance with London School of Hygiene and Tropical Medicine regulations. The thesis is organised into six chapters, three of which describe submitted research papers.

Chapter 1 provides an introduction and rationale for the PhD research presented in this thesis.

Chapter 2 summarises the identified research gaps, describes my PhD journey, outlines the PhD aims and objectives and provides supplementary information on the study setting and data used in the primary research papers (presented in later chapters).

Chapter 3 describes the first submitted research paper. This is a systematic literature review addressing the first thesis objective by reviewing the published evidence on associations of engaging in agriculture compared to types of non-agricultural employment with cardiovascular risk factors in low-and middle-income countries.

Chapter 4 describes the second submitted research paper. This exploratory cross-sectional analysis addresses the second objective of the thesis by using a novel continuous proxy for urbanisation level to explore the association of place of residence with mean systolic blood pressure (SBP), body mass index (BMI), low-density lipoprotein (LDL) and fasting plasma glucose (FPG) among adults at various stages of urbanisation in Telangana, South India.

Chapter 5 describes the third submitted research paper that builds directly on the findings from Chapter 4 and addresses the third objective of this thesis. This is a mediation analysis decomposing the total effect of place of residence (measured by increasing urbanisation levels) into a direct effect and three indirect effects via groups of hypothesised mediators.

Chapter 6 summarises key findings of the PhD research, discusses related research and policy implications and provide conclusions.

The three submitted research papers were included without adaptation, and although repetition has been minimised where possible, some overlap should be expected between chapters. In most chapters, tables and figures were inserted into the main text following their citation. Tables, boxes and figures of the third research paper (Chapter 5) were inserted after the reference list. Additional tables, figures or appendices of research papers intended for online publication are included after the respective papers' reference lists.

CHAPTER 1 INTRODUCTION

1.1 Global epidemiological transitions

Significant gains in life expectancy have been achieved globally over the past several decades. This achievement reflects substantial commitment and investment into improving healthcare and reducing risk factors associated with child and maternal mortality.[1-3] While challenges remain to combatting preventable communicable, maternal and neonatal mortality in many low- and middle-income countries (LMICs), non-communicable diseases (NCDs), such as cardiovascular disease (CVD), diabetes and cancer, have emerged as the predominant threats to global human health and survival.[1-3] The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD)[2] compiles large amounts of data from a number of sources to track long-term trends and impacts of global, regional and country-specific health threats as well as identify opportunities for improving human survival, health and wellbeing in all regions of the world.[2, 4] In their 2017 global, regional, and national mortality review update an estimated 73.4% (95% confidence interval (CI) 72.5, 74.1) of all deaths were due to NCDs in 195 countries and territories; a 22.7% (95% CI 21.5, 23.9) increase over 10 years.[4] Although the surge of the global NCD mortality rate has slowed in recent years, forecasts warn that the percentage of death due to NCDs will reach 80% by 2040.[4] If current trends in NCDs and mental health conditions are not ameliorated, global economic losses associated with NCD-related morbidity and mortality could amount to $47 trillion in the period from 2010 to 2030.[5]

Cardiovascular diseases are now the biggest killer world-wide. The global age-standardised CVD death rate is estimated to 233.1 deaths per 100,000 people (95% CI 229.7, 236.4), which added up to 17.8 million (95% CI 17.5, 18.0) deaths in 2017 alone. Projections warn of continued increases in number of CVD deaths world-wide, however at a slower rate than previously observed, especially in high-income countries (HICs). The burdens of atherosclerotic CVDs, mainly ischemic heart disease (IHD) and cerebrovascular disease, have increased dramatically over the past decade and accounted for 84.9% (95% CI 84,3, 86.3) of all CVD-related deaths in 2017 (IHD: 8,930,400 million (95% CI 8,790,700, 9,138,700); stroke: 6,167,300 million (95% CI 6,044,300, 6,327,600)).[3] Four other conditions; hypertensive heart disease and resulting heart failure, cardiomyopathy, rheumatic heart disease, and atrial fibrillation, account for another 15% of CVD mortality.[6] The number of life-years lost due to IHD is particularly high in the large populations of South and East Asia[6] where CVDs manifest earlier than in other populations.[6, 7]

1.2 India's epidemiological transition

Life expectancy has improved markedly in India during the past two to three decades among both men (8.6 years (95% CI 7.8, 9.5) between 1990 and 2016) and women (10.6 years (95% CI 9.7, 11.6) between 1990 and 2016).[8] This improvement was mainly attributed to considerable reductions in communicable diseases (e.g. measles and diarrheal diseases) and undernutrition, particularly among women. During the same period, NCD mortality increased to account for 61.8% (95% CI 58.2, 64.0) of all deaths, and the number of people suffering from CVDs more than doubled from 25.7 million (95% CI 25.1, 26.0) in 1990 to 54.5 million (95% CI 53.7, 55.3) in 2016.[8, 9] The contribution of CVDs to all-cause mortality has also risen rapidly from an estimated 6.9% (95% CI 6.3, 7.4) of all deaths in 1990 to 28.1% (95% CI 26.5, 29.1) 2016[i].[9]
The steepest age-standardised increases in disease prevalence in India overall between 1990 and 2016 were for IHD (India's number one killer), stroke and diabetes.[8] It is maybe not surprising that levels of major modifiable CVD risk factors rose concurrently, including systolic blood pressure (SBP), fasting plasma glucose (FPG), total cholesterol, body mass index (BMI) and dietary risks (e.g. low fruits and vegetable consumption).[8, 9] The burden of CVDs and related risk factors vary considerably between India's 29 states, with the South being most affected (Figure 1.1).[9, 10] Disproportionate rises in CVD prevalence and risk factors during the past three decades were further reported by rural versus urban residence in India. The Lancet series on hypertension (2012) reported a 30-fold rise in the prevalence of hypertension in urban India over 25 years, and a 10-fold rise in rural India over 36 years.[11] The prevalence of hypertension and CVDs were reported to range from <10% in rural areas to >30% in urban areas across India.[10, 11] With an expected doubling of the adult prevalence of hypertension between 2014 and 2025 (to approximately 60%),[12] it is therefore becoming increasingly important to understand the epidemiology of hypertension in relation to place of residence.
The high case fatality (at relatively low levels of risk factors) and early manifestation of CVDs in India (five to 10 years earlier than in Western populations) pose a substantial threat to the health and productivity of India's working age populations.[12] In 2005, the median age of a first myocardial infarction was estimated to 53 years in India, with 5 to 10% of first events occurring before the age of 40.[13] The CVD trends have high opportunity costs for individuals and households (from spending on health care and drug therapies) as well as society overall.[5, 14] In

[i] Cardiovascular diseases accounted for 14.1% (95% CI 12.9, 15.3) of all DALYs in India in 2016, of which 61.4% were due to IHD and 24.9% were due to stroke.[27]

the absence of interventions to curb current trends, the expected continued high burden of CVDs and related premature death and disability are estimated to cost India $2.17 trillion before 2030.[5]

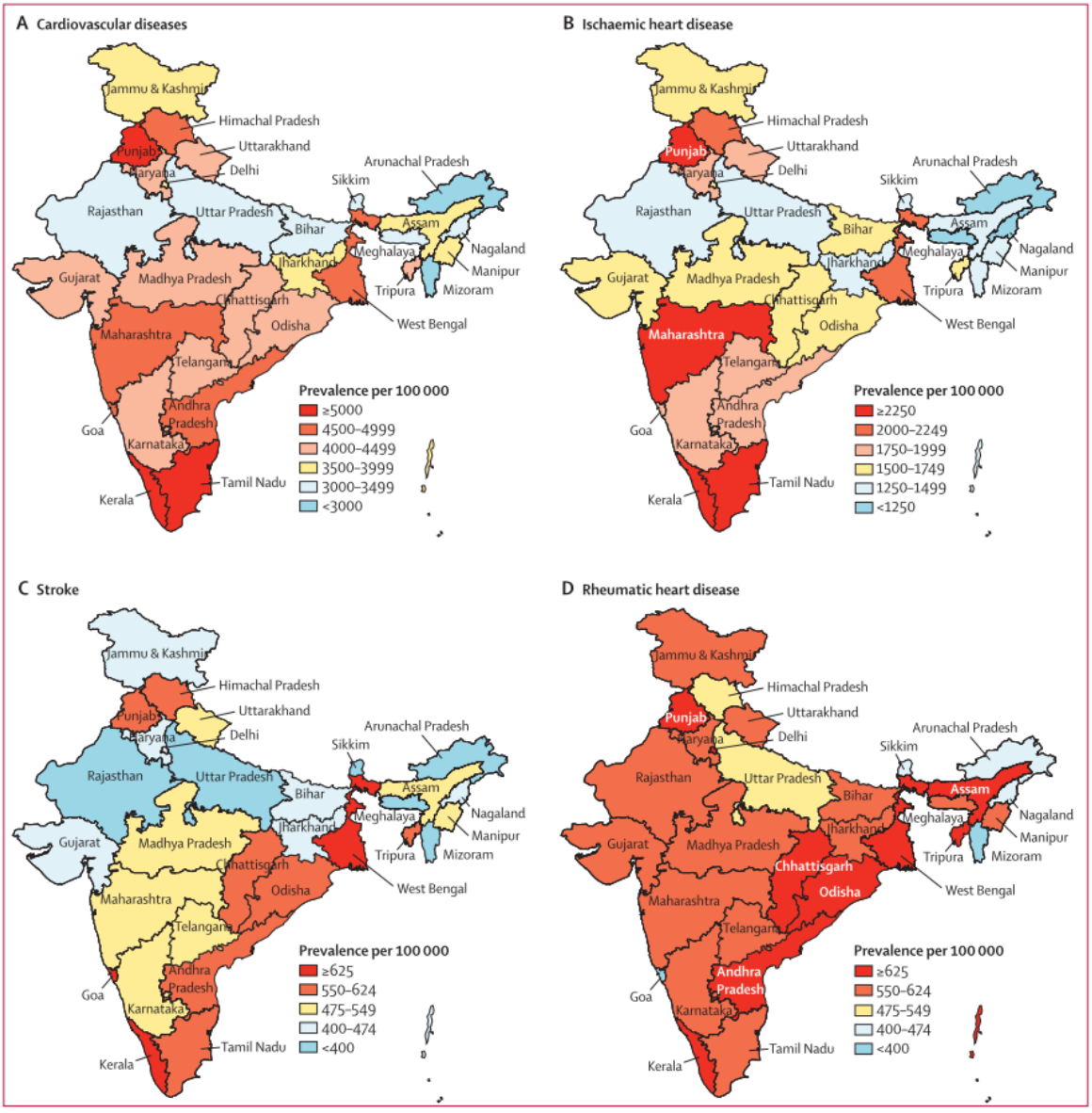


Figure 1.1 Crude prevalence of major cardiovascular diseases by states in India, 2016

Figure reproduced from the India State-Level Disease Burden Initiative Collaborators 2018[9]

1.3 Cardiovascular diseases and leading immediate risk factors

The term CVD covers a group of disorders of the heart and blood vessels.[15] Two major conditions, IHD (e.g. myocardial infarction) and cerebrovascular disease (e.g. stroke) are responsible for almost 80% of the deaths from these diseases.[16] Most CVDs develop through a complex pathological process known as atherosclerosis. Atherosclerosis involves the build-up

of fatty and fibrous deposits, calcium and blood products (mostly macrophages) in the intima of arteries[ii] (known as plaque), which causes them to narrow and harden. The plaque can rupture and form a blood clot (thrombus) locally, or the debris can travel with the blood stream to form a clot elsewhere. The clot may reduce or completely obstruct the flow of blood in a vessel and result in e.g. myocardial infarction (if forming in a coronary artery), or a stroke (if forming in the brain).[15, 16] Table 1.1 describes major atherosclerotic CVDs. Some forms of CVDs, not described in Table 1.1, commonly develop from infection or genetic predisposition (such as rheumatic heart disease, myocarditis, heart valve disease and congenital heart disease), but may also be influenced by atherosclerosis, smoking and hypertension.[17, 18]

Large international representative studies attribute most of the global atherosclerotic CVD burden, with some variation by country, ethnicity, age and gender, to a set of individual-level behavioural and metabolic risk factor: diet, physical activity, smoking, alcohol consumption, hypertension/systolic blood pressure (SBP), obesity/abdominal obesity/body mass index (BMI), diabetes/fasting plasma glucose (FPG), abnormal blood lipids and psychosocial factors (individual or composite scores of depression, locus of control, perceived stress, and life events).[6, 7, 19-21] The relative importance of individual risk factors vary by type of CVD.[7, 20] For example, The INTERHEART study[iii] identifies dyslipidaemia (ApoB/ApoA1 ratio) as the main contributor to acute myocardial infarction globally (contributing 54.1% of population attributable risk (PAR) (95% CI 49.6, 58.6)),[7] whereas the INTERSTROKE study[iv] identified hypertension as the largest contributor to both haemorrhagic (56.4% of PAR (95% CI 52.0, 60.6)) and ischemic stroke (45.7 % of PAR (95% CI 42.4 to 49.0)).[20] The relative importance of metabolic (and behavioural) CVD risk factors have some regional variations, which have been linked to variations in the prevalence of conditions and the magnitude of association between risk factors and conditions.[7, 20] For example, abdominal obesity has a stronger association with myocardial infarction than smoking in South Asia contrary to other regions of the world.[7]

[ii] The innermost of the three layers of an artery

[iii] INTERHEART: International standardised case-control study of first presentation of acute myocardial infarction among 12461 cases 14637 controls (men and women) from 52 countries in all inhabited continents of the world

[iv]INTERSTROKE: International standardised case-control study of first presentation of acute stroke among 13447 cases (ischaemic stroke (n=10388), intracerebral haemorrhage (n=3059)) and 13472 controls, from 32 countries in Africa, America, Asia, Australia, Europe and the Middle East

1.3.1 Immediate risk factors for cardiovascular diseases in India

A recent publication from the India Sate-level Disease Burden Initiative identified six behavioural and metabolic risk factors as accounting for most of CVD related DALYs in India: diet, high SBP, high total cholesterol, tobacco use, high FPG, and high BMI (Table 1.2). It should be noted that the estimated contribution of the six risk factor overlap and should not be interpreted as individual complimentary contributions.[9]

Table 1.1 Description of atherosclerotic cardiovascular diseases and their main risk factors

Atherosclerotic CVDs	**Main clinical manifestation**	**Main risk factors**	**Rank of contribution to global all-cause DALYs***	**Percent (%) deaths from all CVDs***
Coronary heart disease (Ischaemic heart disease)	Partial or complete obstruction of blood supply to the heart muscle Thrombosis, ischemia (oxygen deprivation) and cell necrosis; myocardial infarction (MI); stable angina pectoris; unstable angina pectoris; heart failure or sudden death	Abnormal lipids (leading risk factors for MI), smoking, hypertension, diabetes, abdominal obesity, psychosocial stress, decreased consumption of fruits and vegetables, moderate consumption of alcohol, and physical inactivity	1	49.8
Cerebrovascular disease	Partial or complete obstruction of blood supply to the brain Thrombosis, ischemia (oxygen deprivation) and cell necrosis; stroke	Hypertension (leading risk factors for both haemorrhagic and ischemic stroke), abnormal lipids, smoking, diabetes, abdominal obesity, psychosocial stress, decreased consumption of fruits and vegetables, moderate consumption of alcohol, and physical inactivity, and cardiac causes (atrial fibrillation/flutter, previous myocardial infarction, rheumatic valve disease, prosthetic heart valve)	2 (haemorrhagic and other non-ischemic stroke) 3 (ischemic stroke)	35.3
Hypertensive heart disease	Structural changes to the myocardium (heart muscle), coronary vasculature (blood vessels), and electrical conduction system of the heart left ventricular hypertrophy (thickening of the heart muscle), ventricular arrhythmias, coronary heart disease, cardiac arrhythmias, heart failure	Uncontrolled and prolonged hypertension	4	5.3

Atherosclerotic CVDs	Main clinical manifestation	Main risk factors	Rank of contribution to global all-cause DALYs*	Percent (%) deaths from all CVDs*
Atrial fibrillation (most common cardiac arrhythmia)	Disorders of the electrical conduction system of the heart, resulting in random contractions of the upper heart chambers Reduced efficiency and performance of the heart leading to irregular and uncoordinated atrial impulse, stroke and heart failure	Diabetes mellitus, arterial hypertension, obesity, metabolic syndrome, smoking and cardiac diseases. Other risk factors include, alcohol (particularly binge drinking), bypass surgery, medicines, hyperthyroidism, sick sinus syndrome	8	1.1
Aortic aneurysm	Bulging, swelling and weakening of the aortic wall, can result in shoulder, neck, lower back, hip and abdominal pain and rupture of the blood vessel	Atherosclerosis, break-down of the middle (muscular) layer of the aortic wall, hypertension and chest injury	9	0.9
Peripheral vascular disease	Partial or complete obstruction of the blood vessels supplying the arms, legs and organs aside from the heart and brain Thrombosis, ischemia (oxygen deprivation) and cell necrosis	Abnormal cholesterol, diabetes, CHD, hypertension, kidney disease, smoking	11	0.3

CHD – Coronary heart disease; CVD – cardiovascular disease; DALYs – disability adjusted life years; IHD – ischemic heart disease; MI – myocardial infarction

Table adapted from multiple sources[6, 15, 16 17, 18, 22, 23]; * reproduced from Joseph et al 2017[6]

Table 1.2 Contribution of leading CVD risk factors to CVD related DALYS in India in 2016

CVD risk factor	Percent contribution to CVD DALYs	95% CI
Diet	56.4	48.5, 63.9
High systolic blood pressure	54.6	49.0, 59.8
High total cholesterol	29.4	24.3, 34.8
Tobacco use	18.9	16.6, 21.3
High fasting plasma glucose	16.7	11.4, 23·5
High body mass index	14.7	8.3, 22.0

CVD – cardiovascular disease, DALYs – disability adjusted life years

Source[9]

1.4 India's public health context

Successive governments in India have committed to strengthen the national health system, albeit with various success. Major long-term shortcomings relating to public investment in health-care, particularly by the central government; poor health care infrastructure and coordination of health care delivery sectors; and insufficient health-care governance and accountability have eroded the availability, access, affordability, quality and trust in the Indian public and private health-care system.[24-26] In 2014 public health spending in India was reported amongst the lowest in the world (at approximately 1% of GDP) while out-of-pocket expenditure on health-care ranked amongst the highest, sending 50 to 60 million Indians into property every year.[25, 26] Despite the introduction of more than 20 central and state government sponsored financial protection schemes (such as Yeshaswini in Karnataka and Rajiv Aarogyasri in Andhra Pradesh), out-of-pocket health-care expenditure accounted for 65% of the county's total health expenditure in 2016, with more than half of the cost incurred on drugs.[24, 26-28] A study that quantified disease-specific out-of-pocket expenditure on health-care in India suggests that the treatment of cancer, CVDs and injuries are responsible for the highest out-of-pocket costs in India.[28] Vast disparities have further been reported in availability, access and quality of health services across Indian states, place of residence (rural vs urban), and social and economic strata.[24, 29] It is common that states with poor performance on health determinants also perform worst on public health services. Rapid growth in an unregulated private sector has drained the public health sector from specialists and both utilisation and quality of public health-care has steadily declined at a particular disadvantage for the poorest and vulnerable Indians.[24]

However, several important steps have been taken to ameliorate the long-standing shortcomings of the Indian health system. Namely, health-care gaps between rural and urban areas was sought bridged by supplementing the Rural Health Missions with an Urban Health Mission, joined under one National Health Mission in 2013 that aims to achieve "equitable, affordable & quality health care services"[30] through strengthening public health infrastructure and service delivery and produce more and better qualified human resource in the health sector.[29] The most recent

National Health Policy (2015) clearly articulates the governments intentions to achieve universal health care, including free essential drugs and diagnostics (under the National Health Mission). In 2009 the voluntary Aadhaar Card (linked to information on age, sex, residence and biometrics) was launched to provide proof of identity (12-digit identification number) and address as well as to shorten bureaucratic processes of accessing government benefits.[31] In March 2018 the government launched what has been called 'one of the most ambitious health missions in the world', the Ayushman Bharat Pradhan Mantri Jan Arogya Yojana (AB-PMJAY) that increased benefit coverage to 40% of India's population (the poor and vulnerable). However, impact evaluation of the AB-PMJAY is pending to assess whether the ambitious scheme will succeed in achieving what numerous Indian insurance schemes has failed to deliver in the past.[29]

1.5 Why consider urbanisation in low-and middle-income countries?

Evidence based risk factor modification and therapeutic intervention targeted at individual level immediate risk factors for CVD have been instrumental in slowing CVD mortality rates in HICs.[32] Therapeutic interventions and life-long treatment regimens are, however, expensive and may have several barriers to adherence in LMICs.[6] It is well recognised that, in order to efficiently target and modify immediate risk factors for CVDs at the individual- and household levels, the broader context needs to be considered, including the underlying economic and social determinants.[16, 33, 34] Population growth and aging have received much attention as important upstream determinants of increasing CVD burdens in LMICs.[3, 4, 6] Approximately 7.6 billion people inhabited the globe in 2017, with the highest concentration of people living in Asia (60%) and Africa (17%).[35] Although growth rates appear to be slowing in most countries, the global total population is projected to increase to 9.8 billion people by 2050[v]. However, economic growth, market integration, foreign direct investment and urbanisation together have been suggested to influence long-term changes in CVD mortality almost three times more than population growth, with the strongest influences observed in LMICs.[37] Low- and middle-income countries have undergone rapid urbanisation during the past decades at the same time that morbidity and mortality from CVDs have risen to unprecedented levels.[3, 36] As a result, there has been much interest in exploring the extent to which and the pathways through which, urbanisation influences burdens of CVDs in LMICs, to inform effective opportunities for prevention.

[v] India, with its 1.3 billion inhabitants, is expected to be the largest contributor to global population growth between 2017-2050, and overtake China to become the most populous country in the world with a projected population of 1.7 billion people.[35, 36]

1.6 Defining urbanisation

Before synthesising existing evidence on associations of urbanisation or urbanisation levels with CVDs and CVD risk factors, I will address a key challenge of this topic: how to measure the exposure? There is no international consensus on the definition of urbanisation or how to best measure 'level of urbanisation'.[38] It is commonly accepted that urbanisation reflects the process and dynamics of permanent concentration of a population in urban settlements.[39, 40] It also describes the transition from rural towards urban ways of life.[41, 42] As such, urbanisation may represent varying rates and intensities of environmental, demographic and social changes that influence both urban and rural landscapes and livelihoods, e.g. expansion of urban land cover; increased size, density, and heterogeneity of settlements; development (or deterioration) of infrastructure and migration.[38, 41, 42] Urbanisation is commonly represented by the growth rate of a country's urban population.[38] There is, however, no standard metric for measuring what is 'urban' and separating inhabited territories into rural and urban is not straightforward. As a result, cross-country comparisons and global urbanisation estimates and projections usually reflect various country-specific definitions of 'urban'.[38, 40-42] There are large variation in how national authorities, international organisations and researchers define 'urban'. Occasionally, definitions are reclassified and may thus further vary over time within countries.[38, 40] This makes it difficult to compare directly estimates of urbanisation across different countries and sometimes even within countries over time.

Most urban definitions draw on population criteria; predominantly population size and density and less commonly, economic activities and function, e.g. proportion of population engaging in agriculture. Threshold population sizes typically fall between 2000 and 5000 people, however large variations exist between some countries. For example, an urban settlement can be as small as 200 inhabitants in Sweden, while in Mali the definition of "urban" is a minimum of 40,000 people.[38, 43] Additionally, many countries include in their definitions administrative criteria, political boundaries or physical characteristics, e.g. the presence or degree of electric lighting, paved roads or sewerage infrastructure.[38, 43] The criteria used to separate inhabited territories into 'urban' and 'rural' rely heavily on national censuses, which may be infrequent in LMIC, particularly in times of conflict or economic difficulty.[40] As a result, periodic (annual or similar) estimates of rural and urban populations are often produced by extrapolation of available data (e.g. from neighbouring countries) and previous trends in LMICs settings.[40]

Rural urban dichotomies and migrant studies are frequently used to explore health outcomes by 'level of urbanisation'. However, it is difficult to compare findings between studies due to the aforementioned heterogeneity in what is considered 'urban' as well as differences in selected comparator groups. While broad rural urban dichotomies are useful for examining differences

between either extreme of the urbanisation continuum, they likely mask variations in urbanisation-related characteristics across the continuum, and thus ignore important information on underlying mechanisms.[41, 44] Understanding the mechanisms involves exploring the individual factors of urbanisation level, including differences in environmental factors and social determinants of health that change with urbanisation, e.g. levels of air pollution, affluence, education and type of occupation.[41] Although these factors may change at different rates during urbanisation, they tend to be more similar within than between urban and rural areas.[44-46] For example, urban residents may enjoy higher levels of affluence, education and occupation (if informal and illegal urban settlements are not considered[47]), whereas the opposite may be the case in rural areas.[44, 48] Such limited discordance at either extreme of the urbanisation continuum prevents identification of the relative importance of individual factors when they are measured as part of broad dichotomies.

The level of urbanisation (or urbanicity, as it is sometimes referred to in the literature) is static, as opposed to the process of urbanisation. That is, it refers to the degree of urbanisation of a given settlement at a given point in time.[49] It represents the extent to which urbanisation-related population, physical and social characteristics are present in a given place and time, including e.g. population size, infrastructure and health services.[42, 49] During the past decade, researchers have explored the utility of multi-component urbanisation scales to construct more detailed representations of the urbanisation continuum from information on e.g. population density, physical environment, infrastructure, services, housing, living conditions, economic activities and education.[45, 46, 49-55] Although multi-component scores may be more informative for studying the impact of urbanisation level on health and implicated pathways,[44, 46] it remains unclear which scale components best capture levels of urbanisation.[41, 49] The multi-component urbanisation scales further require large amounts of data that are often not available from LMICs and can be expensive to collect. A recent systematic review concluded that the reliability and validity of existing urbanisation scales are insufficiently established and called for urgent work to identify a reliable and valid standardised measure of urbanisation levels.[49]

More recent efforts have gone into exploring the utility of remote sensing data to characterise urban landscapes and urbanisation dynamics[56, 57] as well as study related health effects.[56-58] Much interest has been taken in the night-time light intensity (NTLI) data obtained by The United States' Defence Meteorological Satellite Programme's Operational Linescan System (DMSP-OLS).[56, 57] The DMSP-OLS was established in 1972 to monitor global distribution of cloud cover and top temperature using visible and infrared imagery from satellite sensors. Each day and night a satellite orbits the globe repeatedly to record near global imagery (longitude 180° W - 180° E; latitude 65°S - 75° N).[56, 59] On cloud free nights the sensors measure light emission from the earth's surface, e.g. from human settlements, industrial sites, natural gas

flares, and illuminated marine vessels.[20, 59] Additionally, the sensors capture transient light, such as lightning, fire, aurora, solar and lunar information.[60] The raw data is processed to estimate annual averages of light intensity from 'persistent sources' associated with human settlement, i.e. excluding transient light. The processed data are aggregated into 30 arc seconds grids, equivalent to a resolution of approximately 1km x 1km pixels.[61] Each pixel in the published data contains the annual average NTLI represented by a unitless digital number ranging from 0 (no light) to 63 (light saturation).[56, 62]

The NTLI data is suggested to be a valid proxy for level of urbanisation due to its strong association with built-up area,[56-58] population density, economic activity, and energy use at global, regional and local levels.[57, 62] With its near global coverage the NTLI may offer a novel, globally comparable, single standard metric for levels of urbanisation. Night-time light intensity data are published annually and are available from the National Oceanic and Atmospheric Administration, National Geophysical Data Center.[63] The NTLI data has been widely used in research on cancer[64-72] and more recently, in studies of blood pressure, overweight and obesity.[73, 74] The calibrated NTLI time series, consisting annual averages dating back to 1992,[63] has additionally shown promise as a metric for the urbanisation process over time,[56, 62] especially in rapidly developing countries, such as China, India, and Brazil.[62]

1.7 Associations of urbanisation levels with atherosclerotic CVDs and risk factors

A large body of evidence identifies urbanisation and changing levels of urbanisation as broad systemic drivers of the CVD epidemic in LMICs.[6, 10, 75-77] Due to a lack of reliable data on CVD morbidity and mortality in LMICs, most studies to understand links between urbanisation or urbanisation levels and CVDs in these settings, come from studies of major, well-established immediate risk factors for CVDs, e.g. blood pressure or LDL.[6, 78]

1.7.1 Rural-urban comparisons and migrant studies

A number of rural urban comparisons and migrant studies consistently observe that migrants and urban residents have higher levels of BMI, overweight, obesity,[79-89] SBP, hypertension, [79-81, 83-86, 88-94] FPG, diabetes,[79-81, 83, 85, 89] blood lipids and dyslipidaemia[80, 82, 84, 85, 88, 89, 94, 95] than rural residents in various LMIC. Results vary by age[90, 91, 93] and gender[82, 85, 86, 91] in some studies from India, China and Guatemala. Some studies further link longer time spent living in urban areas with greater mean BMI[80] and likelihood of overweight,[81] dyslipidaemia[94] and hypertension[96] among both genders, as well as greater mean SBP and FPG among men.[80] These links were however not supported by other studies.[85] It is difficult directly to compare the results from urban-rural comparisons due to differences in the classification of rural and urban. This

approach is further challenged by comparing outcomes between populations that may be inherently different in other, often unmeasured, aspects than exposure status. The overall consistent directions of associations, however, do suggest that high levels of urbanisation compared to low levels have adverse associations with cardiovascular health. Issues of residual confounding due to differences in unmeasured characteristics in rural-urban comparisons are discussed in further detail in the second research paper and the general discussion of this thesis (see Chapters 4 and 6).

1.7.2 Studies using multi-component scores to measure level of urbanisation

In order to identify the body of evidence investigating associations of urbanisation levels measured by urbanisation scales with CVDs and risk factors in LMICs, a rapid PubMed title/abstract search was conducted. Keywords relating to 'urbanisation/urbanicity score/scale' and 'cardiovascular diseases' were used. Bibliographies of relevant articles were manually search for additional publications. Studies were included if (i) the exposure was derived from a data-based urbanisation scale with any number of components; (ii) outcomes included CVDs, stroke, ischemic (or coronary) heart disease, SBP/hypertension, BMI/overweight and obesity, FPG/diabetes or blood lipids/dyslipidaemia; (iii) the study was conducted in or included at least one LMIC.

A total of eight studies were included (Table 1.3). Six of the studies were set in China (five used different waves of the China Health and Nutrition Survey (CHNS)), one was set in India and one in Sri Lanka. Sample sizes ranged from 3705 to 31333 men and women. One study restricted analysis to women. Most studies included participants from 18 year of age (n=4), two studies included participants as young as 15 or 16 years, one study included participants aged 35 years or older and one study reported including 'adults'. Three studies reported the age of the oldest included participants (64, 70 and 90 years respectively). Seven of the studies used the term urbanicity to describe level of urbanisation. In the following section I use the terms urbanicity and urbanisation level according to their use in the original publications. All eight studies categorised urbanisation/urbanicity levels from the scores, with large variations in applied cut-points. Three studies used terciles as cut-points for defining low, medium and high levels of urbanicity and one study used quintiles to determine five urbanicity levels. One study distributed urbanicity score points equally between three groups (low urbanicity 0, <24 points; medium 24, <46 points; high 46, 70 points). Methods for categorising urbanisation level (into 8 categories) were not reported in one study. Two studies using CHNS data waves between 1991 and 2004 included change in urbanicity level over time in their categorisations: One cohort study categorised level of urbanicity according to four levels of net urbanicity change across the five waves of the CHNS data, separated by 10 point intervals. The remaining cross-sectional

study calculated urbanicity scores for multiple CHNS data waves and dichotomised the change in urbanicity score over time. Low urbanicity level was defined as communities that remained below the median level of urbanisation, estimated from data pooled across all waves. The high urbanicity level included communities that moved from below to above the median. One study reported on stroke prevalence, five reported on SBP and/or hypertension; six on BMI and/or overweight and obesity; and two on FPG and/diabetes.

One study from Taiwan, China, (n=20,855) compared the prevalence of stroke between eight levels of urbanisation based on a 14 component score.[54] The likelihood of stroke was highest at the greatest level of urbanisation (level 1) and generally declined with decreasing level of urbanisation in analyses adjusting for age, gender, SES, BMI, hypertension, diabetes, smoking, age at stroke and history heart disease. Results in the mid-range of the urbanisation categories were however not statistically significant at the 95% level. Five studies from China, India and Sri Lanka, which used three or five categories of multi-component urbanicity scores, observed a positive association of urbanicity levels with BMI, overweight, obesity[50, 51, 55], SBP, hypertension,[53, 55] FPG and diabetes[51, 52] with increasing level of urbanicity among men and women after adjusting for different sets of potential confounders such as age, gender and income. Three studies from India and Sri Lanka reported associations of increasing urbanicity level with hypertension (n=2)[50, 51] and overweight and obesity (n=1)[52] among men but not women after adjusting for a limited number of confounders. A cross-sectional and a cohort study used data from the CHNS to construct 12[46] and 26[45] component urbanicity scales respectively. The studies, which adjusted for different sets of confounders, observed no difference in the likelihood of overweight and obesity or hypertension when dichotomising the scale into high vs. low urbanicity. One of the studies additionally performed a subgroup analysis among women, in which urbanicity level was further categorised into four groups and stratified by five baseline urbanicity levels.[46] The likelihood of being overweight or obese was lowest in the group with the lowest baseline urbanicity level and no apparent change over time. The likelihood of overweight and obesity further rose with increasing level of urbanicity over time in the stratum with the lowest baseline urbanicity level, whereas 95% CIs overlapped for ORs in the remaining four strata. Two other studies using CHNS data waves collected over 13 and 19 years respectively, observed a decrease in the magnitude of association over time for SBP[53] and obesity.[55] One age-specific analysis suggested that associations of urbanicity level with SBP is greater in older men and women.[53]

Table 1.3 Studies exploring the association of urbanisation levles measured by multi-component scales with metabolic CVD risk factors

Author (year), country	Study design/ data source	Population	Scale / categories (categorisation method)	covariates	CVD/ metabolic CVD risk factor	Findings
Allender (2010), Tamil Nadu, India,	Cross-sectional, 2003 Indian NCD risk factor surveillance study (health outcomes) and Census of India 2001 (urbanicity scale) Urban areas (Chennai): cluster-random sample); rural areas (six villages and one small rural town in Kancheepuram district): purpose sample	3705 men and women 15, 64 years of age (mean 39.2 (sd 14.4))	Three categories (low= 0, <24, medium= 24, <46), high =46, 70) of 7 component urbanicity scale: population size, population density, access to markets, communications, transport, education and health services (low (0 to < 24), medium (24 to < 46) and high (46–70))	Age, gender	SBP, Hypertension (BP ≥140/90 mmHg) BMI, overweight/obesity (≥ 25 kg/m2)	Men: Increasing level of urbanisation is associated with rising mean and odds of overweight/obesity (OR medium 2.4 (95% CI 1.1, 5.0); high 7.0 (95% CI 3.7, 12.9)) as well as rising mean SBP and odds of hypertension (OR medium 1.2 (95% CI 0.8,1.9); high 1.90 (95% CI 1.3, 2.7)) after adjusting for age Women: Increasing level of urbanisation is associated with rising mean and odds of overweight/obesity OR medium 2.16 (95% CI 1.3, 3.5); high 7.34, 95% CI [5.1, 10.5)) after adjusting for age, but not levels of SBP or hypertension
Allender (2011), 7 provinces, Sri Lanka	Cross-sectional, Sri Lankan Diabetes and Cardiovascular Study	4,485 men and non-pregnant women (response rate 89.7%) 18+ years of age	Three categories (terciles) of multi-item urbanicity scale: population size, population density, and access to markets, transportation, communications/media, economic factors, environment/sanitation, health, education, and housing quality	Age and income	Overweight/obesity (>23 kg m2) Diabetes (fasting plasma glucose ≥7.0 mmol/l or plasma glucose ≥11.1 mmol/l 2 h post-OGTT, Hypertension (BP ≥120/80 mmHg diastolic	Men: Rising level of urbanisation is associated with higher age and income-adjusted ORs (reference=low) of overweight and obesity (medium 1.42 (95% CI 1.08, 1.86); high 1.80 (95% CI 1.41, 2.48)); diabetes (medium 1.30 (95% CI 0.86, 1.98); high 2.05 (95% CI 1.35, 3.11)); and hypertension (medium 1.30 (95% CI 1.02, 1.66); high 1.05 (95% CI 0.80, 1.36)) women: Rising level of urbanisation is associated with higher age and income-adjusted ORs (reference=low) of overweight/obesity OR (medium NA; high 2.47 (95% CI 2.02, 3.01)); diabetes (medium 1.27 (95% CI 0.93, 1.74); high 2.14 (95% CI 1.58, 2.91)); but not hypertension (medium 1.06 (95% CI 0.88, 1.29); high 0.99 (95% CI 0.81, 1.22))

Author (year), country	Study design/ data source	Population	Scale / categories (categorisation method)	covariates	CVD/ metabolic CVD risk factor	Findings
Attard (2012), 9 provinces in China	Cross-sectional of 2009 cohort wave of the China Health and Nutrition Survey (CHNS) Multistage, random cluster sampling stratified by income	7742 men and women, 217 communities 18, 90 years of age	Three categories (terciles) of 12 component urbanicity scale: population density, economic activity, traditional markets, modern markets, transport infrastructure, sanitation, communication, housing, education, diversity (education and income), health infrastructure and social services	Region (north, south), clustering at community and province levels, BMI	Diabetes prevalence (fasting blood glucose ≥7.0 mmol/l or doctor diagnosis) overweight (BMI 24, 28 kg/m2), obesity prevalence (BMI ≥28 kg/m2)	Twofold higher prevalence of diabetes in urban vs rural areas after adjusting for residential community, province, age and household income (men OR 2.02, 95% CI 1.47, 2.78; women, OR 1.94, 95% CI 1.35, 2.79). Results not reported for middle urbanicity group The prevalence of overweight and obesity using international and Chinese criteria (overweight BMI 24–28 kg/m2; obese BMI ≥28 kg/m2) increased with increasing level of urbanisation among men but not women Including BMI in models for diabetes did not change estimates (by the pre-specified 10% cut off for confounding)
Attard (2015), 9 provinces in China	Cohort (min one follow-up between 1991–2009) of the CHNS Multistage, random cluster sampling stratified by income	18754 men and women, 228 communities 18, 70 years of age	Three categories (terciles) of 12 component urbanicity scale: Population density, economic activity, traditional markets, modern markets, transport infrastructure, sanitation, communication, housing, education, diversity (education and income), health infrastructure and social services	Time, gender, age, clustering at province, community, and individual levels	SBP, hypertension (BP 140/90 mmHg)	Prevalence of hypertension and level of urbanisation increased simultaneously over 18 years (hypertension: 13% in 1991, ~26% in 2009) Difference in SBP (2009 minus 1991 values) are larger at lower than higher levels of urbanisation Differences between high (75th percentile) versus low (25th percentile) urbanisation level is greatest among men in 1991 and narrow over time. Differences were greater in the older cohort
Jones-Smith (2010), 9 provinces in China	cross-sectional of 1991 and 2004 waves of the CHNS Multistage, random cluster sampling stratified by income	1433 women, 218 neighbourhoods Adults	Four categories (net urbanicity change (0, 10, 20, 30 points) across 5 baseline urbanicity levels) of 12 component urbanicity scale: population density, economic activity, traditional markets, modern markets, transport infrastructure, sanitation, communication, housing, education, diversity (education and income), health infrastructure and social services	Time, baseline level of urbanicity and outcome	Overweight/obesity (BMI ≥25) incidence	Increased level of urbanisation score (over 13 years) is associated with incident overweight/obesity The odds of overweight/obesity increased with increasing level of urbanicity (greater change over time) among women with the lowest baseline urbanicity score, but not among four groups of women with higher baseline urbanicity score. Increasing level of urbanicity over time may be associated with higher odds of overweight/obesity only up to a certain point of urbanicity

Author (year), country	Study design/ data source	Population	Scale / categories (categorisation method)	covariates	CVD/ metabolic CVD risk factor	Findings
						When using traditional rural and urban categorisation, there is no difference in the odds of overweight/obesity
Lin (2007), Taiwan, China	Cross-sectional, 2001 National Health Interview Survey (NHIS) Multistage stratified systematic sampling	9794 men and women 35+ years of age (mean 50.6 (sd 12.2))	8 categories (na) of 14 component urbanisation scale: population density, age, employment rate, density of manufacturing industry, male immigration rate, female immigration rate, economic activities, annual income, annual expenditure per person, daily amount of garbage per 1,000 population, number of telephones per family, education, number of physicians per 1,000 population, availability of health care facilities in each city/county	Age, gender, SES, BMI, hypertension, diabetes, smoking, age at stroke, history of heart disease	self-reported stroke prevalence	The odds of stroke were highest at the greatest level of urbanisation (level 1). The odds of stroke declined, when comparing the highest level of urbanisation to each of the remaining 7 levels and adjusting for age, gender, SES, BMI, hypertension, diabetes, smoking, age at stroke and history heart disease (although statistical significance at the 95% level was limited to the second highest and the two lowest levels)
Poel et al 2012, China	Cohort, five waves of the CHNS (1991-2004) Multistage, random cluster sampling stratified by income	31333 person–wave observations among men and women 18+ years	High vs. low urbanicity (remaining at the bottom half (median when pooling all waves) of the distribution vs. moving from the bottom to the top half) measured as above and below the median score from multi-component urbanicity index: factors pertaining to population size, land use, transportation facilities, economic activity, and public services	Age, sex, marital status, and household size), socioeconomic status (education and income), and household living conditions*	Obesity (BMI>30) hypertension (≥140/ 90mmHg and/or respondent was taking medication to lower blood pressure	No association of high vs low urbanicity with hypertension or obesity

Author (year), country	Study design/ data source	Population	Scale / categories (categorisation method)	covariates	CVD/ metabolic CVD risk factor	Findings
Poel et al 2009, China	Cohort, five waves of the CHNS (1991, 1993, 1997, 2000, 2004) Multistage, random cluster sampling stratified by income	6484 men and women 16+ years	Five categories (quintiles) of 26 component urbanicity scale: farmland, agricultural workers, bus station, train station, dirt roads, gravel roads, tarmac roads, any tarmac road, distance to tarmac, services, distance to market, telephone, post office, newspaper, primary school, secondary school, vocational school, distance to health care, power cut, childcare <3 years, childcare <6 years, socioeconomic context, restaurants, enterprises, % workers in large firms, % workers in small firms, open trade area, population	Income, education, occupation, physical activity, fat intake, smoking, alcohol, gender, marital status, province, error term	Overweight BMI> 25 kg/m2 Hypertension (≥140/90 mmHg and/or medication to lower blood pressure	Overweight and hypertension are more prevalent in urban than rural areas (lowest versus highest third of urbanicity score, no difference observed between lowest vs middle third). The trend weakens over time due to a more even distribution of risk factors for overweight and hypertension across urbanicity levels (e.g. smoking, younger people move away from rural areas) Middle-aged women (45-65 years) were at highest risk of being overweight. Men were at lower risk than women. Hypertension increases with age and men are at greater risk than women More than half of overweight prevalence and almost a third of the rise in hypertension prevalence in urban areas is explained by a decline in physical activity and farming

BMI – body mass index; CHNS - China Health and Nutrition Survey; na – not available; sd – standard deviation; SES – socio-economic status

* Availability of a flush toilet, use of solid fuels within the dwelling, water from a water plant, and the presence of excreta around the household dwelling

Overall, results from eight cross-sectional and cohort studies suggest that increasing levels of urbanisation was associated with likelihood of stroke, BMI, overweight, obesity, SBP, hypertension, FPG and diabetes in Asia. It is however important to note that that scale components and methods for categorising scores into urbanisation levels varied considerably and that five of the eight studies used different data waves of the CHNS cohort and their estimates were thus not independent. Seven studies categorised urbanisation/urbanicity scores into between three and eight levels of urbanisation/urbanicity. Several of the studies observed differences in stroke or CVD risk factors between the lowest and highest urbanisation/urbanicity levels, whereas results for the middle category or mid-range were often inconclusive. Using urbanicity scores to derive three or more levels of urbanisation thus appeared more robust to capturing differences in CVD risk factors that rural-urban comparisons cannot detect. This point was illustrated in one of the studies that compared the use of binary (rural vs. urban) and four-levels categorisation of the urbanicity score, and found no difference in the odds of BMI ≥25 kg/m2 between rural and urban groups, whereas the odds were greater at the highest compared to the lowest level when using four categories. Some evidence from the CHNS further indicated that adverse impacts of increasing level of urbanisation on CVD risk factors might be limited to certain degrees of urbanisation, i.e. communities at early stages of urbanisation.

1.7.3 Studies using remote sensing data of night-time light to measure urbanisation level

There are still only relatively few studies that have used NTLI to assess associations of urbanisation levels with CVDs and risk factors. Two recent studies from Asia explored associations of NTLI with likelihood of obesity and mean SBP. A cross-sectional study from Korea (n=8526) suggested higher odds of obesity among participants, who were exposed to a high level of NTLI (compared to a low level) after adjusting for age, gender, level of education, type of residential building, monthly household income, alcohol consumption, smoking, consuming caffeine or alcohol before sleep, delayed sleep pattern, short sleep duration and habitual snoring (odds ratio (OR) 1.20, 95% CI 1.06, 1.36).[73] A cross-sectional study from South India (n=3150) reported a 2.8 mmHg increase in SBP per interquartile range (IQR) increase in NTLI (measured by NASA's Visible Infrared Imaging Radiometer Suite) after adjusting for age, gender, anxiety, BMI, energy intake, type of household cooking fuel, SES, sedentary score and smoking.[74]

These two studies used NTLI as a proxy for artificial light at night and the availability of green spaces and were not designed to understand potential effects of urbanisation or urbanicity *per se*. However, the included sets of covariates were relevant for studying impacts of urbanisation and as such, the findings support associations of urbanisation level (as measured using NTLI) with SBP and obesity after adjusting for major confounders (e.g. age, gender). They further

suggest that at least some of the association of urbanicity on CVD risk factors are transmitted (mediated) via lifestyle and metabolic factors such as tobacco, alcohol and BMI, which were additionally included by both studies. However, potential mediation by these factors needs to be confirmed by studies applying causal methods.

1.8 Pathways via which urbanisation might act on CVDs and CVD risk factors

1.8.1 Conceptualising the links between urbanisation and CVDs and CVD risk factors

Grounded in a growing body of evidence, several broad conceptual frameworks propose potential pathways between different social, economic and environmental determinants of health with CVD risk or risk factors.[34, 48, 77, 97-100] Urbanisation or level of urbanisation is typically included in these frameworks as one of several major global and national trends that shape cardiovascular health. That is, few frameworks illustrate the impact of urbanisation on CVD or related risk factors in isolation from concepts such as globalisation and population growth. Due to the complexity of implicated pathways, the insights into how urbanisation may act on CVD is generally inferred from linking existing evidence along the plausible pathways. It is important to note that such exercises do not allow for estimating the magnitude of effects via specific pathways. Another approach is to estimate the effect via hypothesised mediators 'one at a time'. The INTERSTROKE study[20] for example, reported the estimated proportional contribution of individual established immediate risk factors for stroke mortality, e.g. diet and physical activity. There are however limitations to using this causal method in situations of complex interrelations between mediators. Mainly, the sum of the estimated effects via individual risk factors will usually add up to more than 100% of the total contribution of all risk factors considered together due to the shared pathways.[101] Nevertheless, previously proposed frameworks offer informative and straightforward illustrations of a very complex model that can be helpful for further conceptualisation and investigation of the implicated pathways. Several comprehensive reviews have elaborated on the complexity of links between urbanisation and CVDs, and the need for formally testing the hypothesised pathways is regularly emphasised. Formally testing the causal effects via the hypothesised pathways is however an ambitious task due to the large amount of information required and the substantial computational challenges.[101, 102] This is perhaps why to-date, no study has established causality of individual pathways in the context of the full causal web.

One way to conceptualise the hypothesised pathways between urbanisation and CVD is to develop a Directed Acyclic Graph (DAG) that includes all possible connections between exposure, outcomes, mediators, confounders and effect modifiers. Drawing on evidence from a wide range of sources, including systematic reviews, high quality global epidemiological

studies, international working papers and action plans[vi] (cited throughout this section), I developed a DAG for each of four leading CVD risk factors (BMI, SBP, LDL and FPG). The details of the DAG for SBP are discussed in the third research paper of this thesis (see Chapter 5). From the DAG, I created a conceptual framework (Figure 1.2) to guide the explorative and causal analyses that address the objectives of my PhD. The conceptual framework illustrates the complex web of risk factors that are hypothesised to effect cardiovascular health in the context of urbanisation at multiple levels, from global and national to household and individual levels. It is important to note that the pathways linking urbanisation level and CVD risk factors sometimes act in opposite directions. That is, urbanisation may have both positive and negative impacts on cardiovascular health through different and sometimes shared pathways. In the following section, I will illustrate key examples of hypothesised pathways in both directions.

1.8.2 Hypothesised pathways linking increasing levels of urbanisation with CVDs and CVD risk factors

The provision of most forms of infrastructure and services generally increase with rising level of urbanisation.[103] Better access to health and social services at greater levels of urbanisation[98, 104, 105] may improve the prevention, detection and treatment of CVDs and risk factors, such as high cholesterol or BMI. Increased level of urbanisation is generally reflected in better opportunities for higher education and the generation of additional, new or more desirable jobs, with potential benefits to productivity and income as well as knock-on effects e.g. on mental health[vii].[37, 103, 106] Greater affluence at greater levels of urbanisation is associated with better living conditions (e.g. access to safe water and sanitation) and improved indoor air quality due to less reliance on solid fuels (wood, coal and agricultural residues) for cooking, heating and lighting.[107, 108] Reduction in the exposure to fine particulate matter from combustion of solid fuels may in turn reduce vascular inflammation, hypertension, insulin sensitivity, diabetes, IHD risk and CVD mortality.[6, 108-111] Improved living conditions further reduces the spread of infectious diseases that may deteriorate cardiovascular health directly or indirectly, e.g. via poor nutritional outcomes or elevated blood pressure.[37, 112] It is important to keep in mind that these

[vi] Connections between intermediate risk factors and CVD were largely informed by causal webs published by the GBD study and major high-quality epidemiological studies such as the INTERHEART and INTERSTROKE. Pathways via more up-stream determinants were mainly based on systematic reviews, working papers, global reports and NCD action plans by international organisation, particularly the UN.

[vii] Impacts may extent to non-urban residents such as dependants and commuters[47, 103]

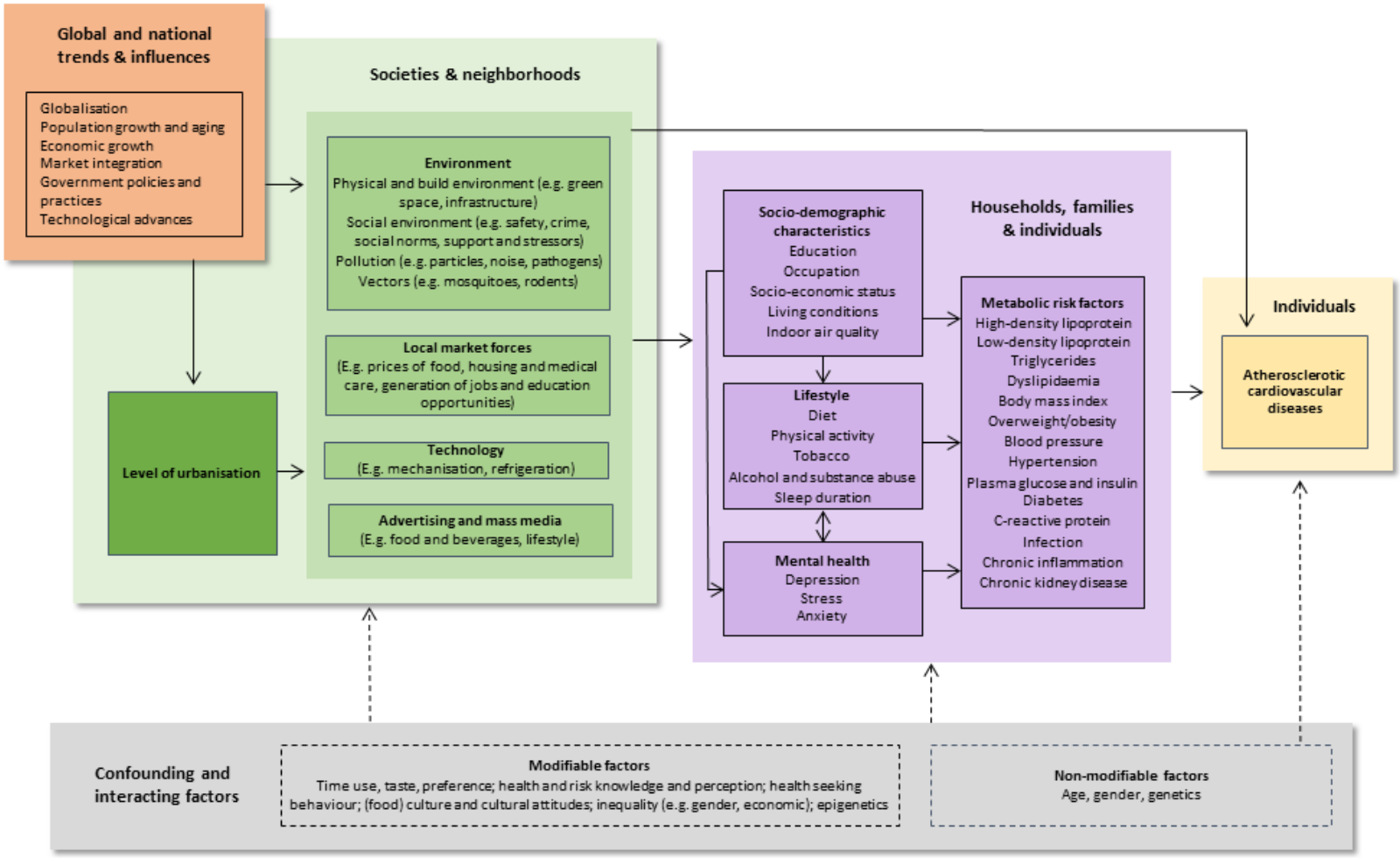


Figure 1.2 Conceptual framework of links between urbanisation and atherosclerotic cardiovascular diseases

Sources[8, 11, 21, 37, 41, 48, 49, 74, 97, 98, 103-107, 110-121]

benefits are not enjoyed equally among all urban residents, particularly at high levels of urbanisation.[48, 98] Poor residents of informal settlements or urban slums are for example, more likely to suffer from dual burdens of communicable and chronic diseases.[48, 98, 118] This is referred to, by some researchers, as the 'urban penalty'.[48, 98] Analysing and ameliorating health challenges in highly urbanised settings may therefore need a separate framework that is sensitive to high levels of intra-urban inequality.[98]

Agriculture (farming) employs approximately 70% of people in low-income countries,[122] and shifts away from agriculture and into the service and manufacturing sectors are typically among the first manifestations of urbanisation.[46] Farming households may for example, take advantage of new employment opportunities as they become available in order to diversify livelihood strategies or entirely transition out of agriculture in pursuit of improved income or work regarded as more desirable and resilient to environmental or market shocks.[47, 51, 116, 122-124] Our livelihoods or occupation greatly influence where and how we live,[118] and as such are important social determinants of health and disease.[118] A number of frameworks outline potential pathways through which agriculture may influence nutrition and chronic diseases,[125-130] however, few studies have investigated associations of transitions away from agriculture and into more urbanised employment with cardiovascular health. Assessments of recent urbanisation in China have highlighted a number of unforeseen economic and social challenges affecting agriculturalists who transitioned into non-agricultural employment in urban areas. Poor integration of transitioning agricultural workers into urban areas and the inability to secure employment rights have resulted in poor living conditions, social exclusion and restricted access to health and other services for this population group.[131, 132] Such changes in economic and social health determinants may well adversely affect CVD incidence, prevalence and/or risk factors in this population group. However, to the best of my knowledge, these associations are yet to be assessed in relation to China's recent urbanisation.

Among the better understood harmful impacts of greater urbanisation level on cardiovascular health is ambient (outdoor) air pollution that is typically worse in urban areas for several reasons, including a greater concentrations of motor vehicles, industry, power plants and households combusting biomass and fossil fuels.[49, 113, 114] Air pollution is consistently associated with considerable rises in blood pressure and risk of IHD and stroke,[8, 110, 113, 115] A growing evidence base suggests that the main mechanisms include abnormal CRP[110, 111, 114] possibly linked to chronic and systemic inflammation;[114, 115] triglycerides,[111] (oxidised) LDL,[110] lipid metabolism[111] and obesity.[115] In addition, higher levels of other forms of pollution including noise, water and solid waste pollution resulting from urbanisation may affect blood pressure via metabolic and mental health mechanism.[41, 104, 113, 115] Greater level of urbanisation is associated with features of infrastructure and built environment that promote sedentary travel and leisure

activities while limiting opportunities for active alternatives, for example due to limited access to green spaces and safe pedestrian or cycle paths.[74, 116] Sedentary recreational activities, such as TV watching and video games, have been reported to be more common at greater levels of urbanisation across LMICs.[112, 117] Snacking and exposure to food and beverage advertising during sedentary recreational activities may further promote (empty) calorie consumption and weight gain with adverse implications for cardiovascular health.[116]

A number of urbanisation-related physical and social changes give rise to rapid 'nutrition transitions'[8] and changing consumption patterns of alcohol and tobacco, which have well established deleterious impacts on cardiovascular health.[112, 113, 119, 121] For example, developments in infrastructure; improved processing, distribution and storage technologies; and the introduction of convenience stores and supermarkets improve availability and access to new and often highly processed foods and beverages, while the number of traditional fresh markets decline.[41, 112] Simultaneously, (aggressive) marketing of commodities and greater purchasing power enable consumption behaviours favouring a growing need or desire for convenience as well as changes is taste and social norms.[97, 98, 112] The association of level of urbanisation with tobacco use remains controversial to some extent in LMIC. Studies from India,[133] China,[92] and Guatemala[86] suggest that rural men and women are more likely to smoke or chew tobacco than their urban counterparts.[133] Other studies suggest that tobacco use is more prevalent in urban areas,[49] or particular population subgroups, e.g. youth, urban poor and less educated individuals in India.[12, 78] Finally, alcohol consumption was higher among some urban than rural populations in India[85] and in a systematic review of multiple countries,[49] whereas no difference was observed in analysis of data from six Indian states (Assam, Karnataka, Maharashtra, Rajasthan, Uttar Pradesh and West Bengal).[133]

[8] Transition away from traditional, predominantly grain based diets towards diets high in fat, sugar and salt, and convenience foods eaten away from the home.[97, 112, 119]

References

1. United Nations. Sustainable Development Goals. Goal 3: Ensure healthy lives and promote well-being for all at all ages. Available from: https://www.un.org/sustainabledevelopment/health/.
2. GBD 2017 Mortality Collaborators. Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018 Nov 10;392(10159):1684-735
3. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018 Nov 10;392(10159):1736-88
4. Foreman KJ, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. Lancet. 2018 Nov 10;392(10159):2052-90
5. Bloom DE, Cafiero-Fonseca ET, Candeias V, Adashi E, Bloom LR, Gurfein L, et al. Economics of Non-Communicable Diseases in India: The Costs and Returns on Investment of Interventions to Promote Healthy Living and Prevent, Treat, and Manage NCDs. World Economic Forum, Harvard School of Public Health, 2014.
6. Joseph P, Leong D, McKee M, Anand SS, Schwalm JD, Teo K, et al. Reducing the Global Burden of Cardiovascular Disease, Part 1: The Epidemiology and Risk Factors. Circulation research. 2017 Sep 1;121(6):677-94
7. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004 Sep 11-17;364(9438):937-52
8. India State-level Disease Burden Initiative Collaborators. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet. 2017 Nov 13
9. India State-Level Disease Burden Initiative CVD Collaborators. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990-2016. The Lancet Global health. 2018 Dec;6(12):e1339-e51
10. Gupta R, Mohan I, Narula J. Trends in Coronary Heart Disease Epidemiology in India. Annals of global health. 2016 Mar-Apr;82(2):307-15
11. Ibrahim MM, Damasceno A. Hypertension in developing countries. Lancet. 2012 Aug 11;380(9841):611-9
12. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India: Current Epidemiology and Future Directions. Circulation. 2016 Apr 19;133(16):1605-20
13. Sharma M, Ganguly NK. Premature coronary artery disease in Indians and its associated risk factors. Vascular health and risk management. 2005;1(3):217-25
14. Narayan KM, Ali MK. Commentary: Shielding against a future inferno: the not-so-problematic discourse on socioeconomic status and cardiovascular health in India. International journal of epidemiology. 2013 Oct;42(5):1426-9
15. World Health Organization. Cardiovascular diseases (CVDs) fact sheet 2017 [updated May 2017; cited 2017 19 April]. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/.
16. World Health Organization. Global atlas on cardiovascular disease prevention and control. 2011.
17. World health Organization. International Classification of Diseases 11th Revision ICD-11. Available from: https://icd.who.int/browse11/l-m/en.
18. National Library of Medicine. Medline Plus, Medical Encyclopedia. Bethesda, USA: National Library of Medicine.
19. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct 08;388(10053):1659-724
20. O'Donnell MJ, Chin SL, Rangarajan S, Xavier D, Liu L, Zhang H, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet. 2016 Aug 20;388(10046):761-75
21. Prabhakaran D, Singh K, Roth GA, Banerjee A, Pagidipati NJ, Huffman MD. Cardiovascular Diseases in India Compared With the United States. Journal of the American College of Cardiology. 2018 Jul 3;72(1):79-95
22. Kallistratos MS, Poulimenos LE, Manolis AJ. Atrial fibrillation and arterial hypertension. Pharmacological Research. 2018 2018/02/01/;128:322-6

23. The Free Medical Dictionary by Farlex. Online: Farlex; 2019. The Free Medical Dictionary.
24. Patel V, Parikh R, Nandraj S, Balasubramaniam P, Narayan K, Paul VK, et al. Assuring health coverage for all in India. The Lancet. 2015 2015/12/12/;386(10011):2422-35
25. Chatterjee P. The health system in India: the underserved majority. The Lancet. 2017 2017/12/02/;390(10111):2426-7
26. Angell BJ, Prinja S, Gupt A, Jha V, Jan S. The Ayushman Bharat Pradhan Mantri Jan Arogya Yojana and the path to universal health coverage in India: Overcoming the challenges of stewardship and governance. PLoS medicine. 2019;16(3):e1002759-e
27. World Bank. Out-of-pocket expenditure (% of current health expenditure) 2019 [cited 2020 26 January]. Available from: https://data.worldbank.org/indicator/SH.XPD.OOPC.CH.ZS.
28. Selvaraj S, Farooqui HH, Karan A. Quantifying the financial burden of households' out-of-pocket payments on medicines in India: a repeated cross-sectional analysis of National Sample Survey data, 1994–2014. BMJ Open. 2018;8(5):e018020
29. Zodpey S, Farooqui HH. Universal health coverage in India: Progress achieved & the way forward. The Indian journal of medical research. 2018;147(4):327-9
30. Ministry of Health and Family Welfare Government of India. National Health Mission 2020 [cited 2020 26 February]. Available from: https://nhm.gov.in/.
31. Unique Identification Authority of India Government of India. AADHAAR SEVA KENDRA 2019 [cited 2020 26 February]. Available from: https://uidai.gov.in/.
32. Leong DP, Joseph PG, McKee M, Anand SS, Teo KK, Schwalm JD, et al. Reducing the Global Burden of Cardiovascular Disease, Part 2: Prevention and Treatment of Cardiovascular Disease. Circulation research. 2017 Sep 1;121(6):695-710
33. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Switzerland: WHO, 2013.
34. World Health Organization. Equity, social determinants and public health programmes. Blas E, Sivasankara AK, editors. WHO press2010.
35. Division UNDoEaSAP. World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. United Nations Department of Economic and Social Affairs Population Division 2017.
36. United Nations. World Urbanization Prospects: The 2018 Revision 2018. Available from: https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf.
37. Eckert S, Kohler S. Urbanization and health in developing countries: a systematic review. World health & population. 2014;15(1):7-20
38. McGranahan G, Satterthwaite D. Urbanisation concepts and trends. IIED Working Paper. London: International Institute for Environment and Development's Human Settlements Group, 2014.
39. Organisation for Economic Co-operation and Development. Glossary of statistical terms: Urbanization 2003. Available from: https://stats.oecd.org/glossary/detail.asp?ID=2819.
40. United Nations DoEaSA, Population Division,. World Urbanization Prospects: The 2018 Revision, Methodology. New York: United Nation, 2018 Contract No.: Working Paper No. ESA/P/WP.252.
41. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008 Nov;85(6):938-51
42. Vlahov D, Galea S. Urbanization, urbanicity, and health. Journal of urban health : bulletin of the New York Academy of Medicine. 2002 Dec;79(4 Suppl 1):S1-s12
43. United Nations Children's Fund. Children in an increasingly urban world. New York: UN; 2012.
44. Dahly DL, Adair LS. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. Social science & medicine (1982). 2007 12/29;64(7):1407-19
45. Poel EV, O'Donnell O, Doorslaer EV. Is there a health penalty of China's rapid urbanization? Health economics. 2012 Apr;21(4):367-85
46. Jones-Smith JC, Popkin BM. Understanding community context and adult health changes in China: development of an urbanicity scale. Soc Sci Med. 2010 Oct;71(8):1436-46
47. Tacoli C. Urbainzation, gender and poverty. New York, USA: United Nations Population Fund, International Institute for Environment and Development, 2012 March. Report No.
48. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.
49. Cyril S, Oldroyd JC, Renzaho A. Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales. BMC public health. 2013 May 28;13:513
50. Allender S, Lacey B, Webster P, Rayner M, Deepa M, Scarborough P, et al. Level of urbanization and noncommunicable disease risk factors in Tamil Nadu, India. Bulletin of the World Health Organization. 2010 Apr;88(4):297-304

51. Allender S, Wickramasinghe K, Goldacre M, Matthews D, Katulanda P. Quantifying urbanization as a risk factor for noncommunicable disease. Journal of urban health : bulletin of the New York Academy of Medicine. 2011 Oct;88(5):906-18
52. Attard SM, Herring AH, Mayer-Davis EJ, Popkin BM, Meigs JB, Gordon-Larsen P. Multilevel examination of diabetes in modernising China: what elements of urbanisation are most associated with diabetes? Diabetologia. 2012 Dec;55(12):3182-92
53. Attard SM, Herring AH, Zhang B, Du S, Popkin BM, Gordon-Larsen P. Associations between age, cohort, and urbanization with SBP and DBP in China: a population-based study across 18 years. Journal of hypertension. 2015 May;33(5):948-56
54. Lin HC, Lin YJ, Liu TC, Chen CS, Chiu WT. Urbanization and stroke prevalence in Taiwan: analysis of a nationwide survey. Journal of urban health : bulletin of the New York Academy of Medicine. 2007 Jul;84(4):604-14
55. Van de Poel E, O'Donnell O, Van Doorslaer E. Urbanization and the spread of diseases of affluence in China. Economics and human biology. 2009 Jul;7(2):200-16
56. Yi K, Tani H, Li Q, Zhang J, Guo M, Bao Y, et al. Mapping and evaluating the urbanization process in Northeast China using DMSP/OLS nighttime light data. Sensors (Basel, Switzerland). 2014;14(2):3207-26
57. Ma T, Zhou C, Pei T, Haynie S, Fan J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. Remote Sensing of Environment. 2012 2012/09/01/;124(Supplement C):99-107
58. Imhoff M, Lawrence W, Stutzer D, Elvidge C. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. Remote Sensing of Environment. 1997 1997/09/01/;61(3):361-70
59. Li X, Zhou Y. Urban mapping using DMSP/OLS stable night-time light: a review. International Journal of Remote Sensing. 2017 2017/11/02;38(21):6030-46
60. National oceanic and atmospheric administration National centers for environmental information. OLS - Operational Linescan System. Available from: https://ngdc.noaa.gov/eog/sensors/ols.html.
61. Elvidge CD, Imhoff ML, Baugh KE, Hobson VR, Nelson I, Safran J, et al. Nighttime lights of the world: 1994-95 ISPRS Journal of Photogrammetry and Remote Sensing. 2001;56:81-99
62. Zhang Q, Seto KC. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sensing of Environment. 2011 2011/09/15/;115(9):2320-9
63. Zhang Q, Pandey B, Seto KC. A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data. IEEE Transactions on geoscience and remote sensisng. 2016
64. Kloog I, Haim A, Stevens RG, Portnov BA. Global co-distribution of light at night (LAN) and cancers of prostate, colon, and lung in men. Chronobiology international. 2009 Jan;26(1):108-25
65. Cho Y, Ryu SH, Lee BR, Kim KH, Lee E, Choi J. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. Chronobiology international. 2015;32(9):1294-310
66. Rybnikova NA, Haim A, Portnov BA. Is prostate cancer incidence worldwide linked to artificial light at night exposures? Review of earlier findings and analysis of current trends. Archives of environmental & occupational health. 2017 Mar 4;72(2):111-22
67. Kloog I, Haim A, Stevens RG, Barchana M, Portnov BA. Light at night co-distributes with incident breast but not lung cancer in the female population of Israel. Chronobiology international. 2008 Feb;25(1):65-81
68. Hurley S, Goldberg D, Nelson D, Hertz A, Horn-Ross PL, Bernstein L, et al. Light at night and breast cancer risk among California teachers. Epidemiology (Cambridge, Mass). 2014 Sep;25(5):697-706
69. James P, Bertrand KA, Hart JE, Schernhammer ES, Tamimi RM, Laden F. Outdoor Light at Night and Breast Cancer Incidence in the Nurses' Health Study II. Environmental health perspectives. 2017 Aug 17;125(8):087010
70. Kloog I, Stevens RG, Haim A, Portnov BA. Nighttime light level co-distributes with breast cancer incidence worldwide. Cancer causes & control : CCC. 2010 Dec;21(12):2059-68
71. Bauer SE, Wagner SE, Burch J, Bayakly R, Vena JE. A case-referent study: light at night and breast cancer risk in Georgia. International Journal of Health Geographics. 2013 Apr 17;12:23
72. Rybnikova N, Haim A, Portnov BA. Artificial Light at Night (ALAN) and breast cancer incidence worldwide: A revisit of earlier findings with analysis of current trends. Chronobiology international. 2015;32(6):757-73

73. Koo YS, Song JY, Joo EY, Lee HJ, Lee E, Lee SK, et al. Outdoor artificial light at night, obesity, and sleep health: Cross-sectional analysis in the KoGES study. Chronobiology international. 2016;33(3):301-14
74. Lane KJ, Stokes EC, Seto KC, Thanikachalam S, Thanikachalam M, Bell ML. Associations between Greenness, Impervious Surface Area, and Nighttime Lights on Biomarkers of Vascular Aging in Chennai, India. Environmental health perspectives. 2017 Aug 2;125(8):087003
75. Institute of Medicine Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing C. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. In: Fuster V, Kelly BB, editors. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. Washington (DC): National Academies Press (US), National Academy of Sciences; 2010.
76. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation. 2001 Nov 27;104(22):2746-53
77. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: Part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. Circulation. 2001 Dec 4;104(23):2855-64
78. Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart (British Cardiac Society). 2008 Jan;94(1):16-26
79. Subramanian SV, Smith GD. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. The American journal of clinical nutrition. 2006 Sep;84(3):633-40
80. Kinra S, Andersen E, Ben-Shlomo Y, Bowen L, Lyngdoh T, Prabhakaran D, et al. Association between urban life-years and cardiometabolic risk: the Indian migration study. American journal of epidemiology. 2011 Jul 15;174(2):154-64
81. Sobngwi E, Mbanya JC, Unwin NC, Porcher R, Kengne AP, Fezeu L, et al. Exposure over the life course to an urban environment and its relation with obesity, diabetes, and hypertension in rural and urban Cameroon. International journal of epidemiology. 2004 Aug;33(4):769-76
82. Torun B, Stein AD, Schroeder D, Grajeda R, Conlisk A, Rodriguez M, et al. Rural-to-urban migration and cardiovascular disease risk factors in young Guatemalan adults. International journal of epidemiology. 2002 Feb;31(1):218-26
83. Al-Moosa S, Allin S, Jemiai N, Al-Lawati J, Mossialos E. Diabetes and urbanization in the Omani population: an analysis of national survey data. Population health metrics. 2006 Apr 24;4:5
84. He J, Klag MJ, Wu Z, Qian MC, Chen JY, Mo PS, et al. Effect of migration and related environmental changes on serum lipid levels in southwestern Chinese men. American journal of epidemiology. 1996 Nov 1;144(9):839-48
85. Ebrahim S, Kinra S, Bowen L, Andersen E, Ben-Shlomo Y, Lyngdoh T, et al. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. PLoS medicine. 2010 Apr;7(4):e1000268
86. Gregory CO, Dai J, Ramirez-Zea M, Stein AD. Occupation is more important than rural or urban residence in explaining the prevalence of metabolic and cardiovascular disease risk in Guatemalan adults. The Journal of nutrition. 2007 May;137(5):1314-9
87. Mendez MA, Monteiro CA, Popkin BM. Overweight exceeds underweight among women in most developing countries. The American journal of clinical nutrition. 2005 Mar;81(3):714-21
88. Patel JV, Vyas A, Cruickshank JK, Prabhakaran D, Hughes E, Reddy KS, et al. Impact of migration on coronary heart disease risk factors: comparison of Gujaratis in Britain and their contemporaries in villages of origin in India. Atherosclerosis. 2006 Apr;185(2):297-306
89. Ramachandran A, Mary S, Yamuna A, Murugesan N, Snehalatha C. High prevalence of diabetes and cardiovascular risk factors associated with urbanization in India. Diabetes care. 2008 May;31(5):893-8
90. Poulter NR, Khaw KT, Hopwood BE, Mugambi M, Peart WS, Rose G, et al. The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. BMJ (Clinical research ed). 1990 Apr 14;300(6730):967-72
91. He J, Klag MJ, Whelton PK, Chen JY, Mo JP, Qian MC, et al. Migration, blood pressure pattern, and hypertension: the Yi Migrant Study. American journal of epidemiology. 1991 Nov 15;134(10):1085-101
92. Wang C, Wei D, Wang B, Zhang J, Zhang K, Ma M, et al. Effect of lifestyle on the prevalence of the metabolic syndrome among farmers, migrants with Yi ethnicity and Han population in Sichuan province of China. Asia Pacific journal of clinical nutrition. 2010;19(2):266-73

93. Norboo T, Stobdan T, Tsering N, Angchuk N, Tsering P, Ahmed I, et al. Prevalence of hypertension at high altitude: Cross-sectional survey in Ladakh, Northern India 2007-2011. BMJ Open. 2015;5 (4) (no pagination)(e007026)
94. Zhao J, Seubsman SA, Sleigh A, Thai Cohort Study Team T. Timing of urbanisation and cardiovascular risks in Thailand: evidence from 51 936 members of the thai cohort study, 2005-2009. Journal of epidemiology. 2014;24(6):484-93
95. Vorster HH, Kruger A, Venter CS, Margetts BM, Macintyre UE. Cardiovascular disease risk factors and socio-economic position of Africans in transition: the THUSA study. Cardiovascular journal of Africa. 2007 Jul;18(5):282-9
96. Sodjinou R, Agueh V, Fayomi B, Delisle H. Obesity and cardio-metabolic risk factors in urban adults of Benin: relationship with socio-economic status, urbanisation, and lifestyle patterns. BMC public health. 2008 Mar 4;8:84
97. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. The American journal of clinical nutrition. 2006;84(2):289-98
98. Vlahov D, Freudenberg N, Proietti F, Ompad D, Galea S. A conceptual framework for organizing determinants of urban health. Thematic paper 1 for Knowledge Network on Urban Settings WHO Centre for Health Development2008.
99. Vlahov D, Freudenberg N, Proietti F, Ompad D, Quinn A, Nandi V, et al. Urban as a determinant of health. Journal of urban health : bulletin of the New York Academy of Medicine. 2007 May;84(3 Suppl):i16-26
100. Misra A, Khurana L. Obesity and the metabolic syndrome in developing countries. The Journal of clinical endocrinology and metabolism. 2008 Nov;93(11 Suppl 1):S9-30
101. VanderWeele T, Vansteelandt S. Mediation Analysis with Multiple Mediators. Epidemiologic methods. 2014 Jan;2(1):95-115
102. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015 Mar;71(1):1-14
103. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
104. Ezzati M, Webster CJ, Doyle YG, Rashid S, Owusu G, Leung GM. Cities for global health. 2018;363:k3794
105. World Health Organization. Urbanization and health 2019 [cited 2019 16 July]. Available from: https://www.who.int/bulletin/volumes/88/4/10-010410/en/.
106. Estache A, Garsous G. The scope for an impact of infrastructure investments on jobs in developing countries International Finance Corporation, 2012.
107. Zhao B, Zheng H, Wang S, Smith KR, Lu X, Aunan K, et al. Change in household fuels dominates the decrease in PM2.5 exposure and premature mortality in China in 2005-2015. Proceedings of the National Academy of Sciences of the United States of America. 2018 Dec 4;115(49):12401-6
108. GBD 2017 Mortality Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018 Nov 10;392(10159):1923-94
109. Kwan GF, Mayosi BM, Mocumbi AO, Miranda JJ, Ezzati M, Jain Y, et al. Endemic Cardiovascular Diseases of the Poorest Billion. Circulation. 2016 Jun 14;133(24):2561-75
110. Yamamoto SS, Phalkey R, Malik AA. A systematic review of air pollution as a risk factor for cardiovascular disease in South Asia: limited evidence from India and Pakistan. International journal of hygiene and environmental health. 2014 Mar;217(2-3):133-44
111. Uzoigwe JC, Prum T, Bresnahan E, Garelnabi M. The emerging role of outdoor and indoor air pollution in cardiovascular disease. N Am J Med Sci. 2013 Aug;5(8):445-53
112. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012 Jan;70(1):3-21
113. Moore M, Gould P, Keary BS. Global urbanization and impact on health. International journal of hygiene and environmental health. 2003 2003/01/01/;206(4):269-78
114. Vogel CF, Sciullo E, Wong P, Kuzmicky P, Kado N, Matsumura F. Induction of proinflammatory cytokines and C-reactive protein in human macrophage cell line U937 exposed to air pollution particulates. Environmental health perspectives. 2005 Nov;113(11):1536-41
115. Meo SA, Suraya F. Effect of environmental air pollution on cardiovascular diseases. European review for medical and pharmacological sciences. 2015 Dec;19(24):4890-7

116. Goryakin Y, Suhrcke M. Economic development, urbanization, technological change and overweight: what do we learn from 244 Demographic and Health Surveys? Econ Hum Biol. 2014 Jul;14:109-27
117. Sullivan R, Kinra S, Ekelund U, Bharathi AV, Vaz M, Kurpad A, et al. Socio-demographic patterning of physical activity across migrant groups in India: results from the Indian Migration Study. PloS one. 2011;6(10):e24898
118. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March. Report No.
119. UNDP/UNFPA, UNICEF, WFP. Agenda Item 1: Population Growth and Rapid Urbanization: Food Insecurity on the Rise in Urban Settings. UNDP/UNFPA, UNICEF, WFP, 2009.
120. Popkin BM, Horton S, Kim S, Mahal A, Shuigao J. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutrition reviews. 2001 Dec;59(12):379-90
121. Godfrey R, Julien M. Urbanisation and health. Clin Med (Lond). 2005 Mar-Apr;5(2):137-41
122. World Bank. Employment in agriculture (% of total employment) (modelled ILO estimate): World Bank; 2018 [updated 2018; cited 2018 28 September]. Available from: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?contextual=employment-by-sector&locations=XO.
123. Bhandari PB. Rural livelihood change? Household capital, community resources and livelihood transition. Journal of rural studies. 2013 Oct 1;32:126-36
124. Cervantes-Godoy D, Brooks J. Smallholder Adjustment in Middle-Income Countries: Issues and Policy Responses. OECD publishing, 2008.
125. Masset E, Haddad L, Cornelius A, Isaza-Castro J. Effectiveness of agricultural interventions that aim to improve nutritional status of children: systematic review. Bmj. 2012;344:d8222
126. Berti PR, Krasevec J, FitzGerald S. A review of the effectiveness of agriculture interventions in improving nutrition outcomes. Public Health Nutr. 2004 Aug;7(5):599-609
127. Girard AW, Self JL, McAuliffe C, Olude O. The effects of household food production strategies on the health and nutrition outcomes of women and young children: a systematic review. Paediatr Perinat Epidemiol. 2012 Jul;26 Suppl 1:205-22
128. Turner R, Hawkes C, Jeff W, Ferguson E, Haseen F, Homans H, et al. Agriculture for improved nutrition: the current research landscape. Food Nutr Bull. 2013 Dec;34(4):369-77
129. Kadiyala S, Harris J, Headey D, Yosef S, Gillespie S. Agriculture and nutrition in India: mapping evidence to pathways. Annals of the New York Academy of Sciences. 2014 Dec;1331:43-56
130. Hawkes C, Ruel M. The links between agriculture and health: an intersectoral opportunity to improve the health and livelihoods of the poor. Bulletin of the World Health Organization. 2006 Dec;84(12):984-90
131. Chen M, Gong Y, Lu D, Ye C. Build a people-oriented urbanization: China's new-type urbanization dream and Anhui model. Land Use Policy. 2019 2019/01/01/;80:1-9
132. Chen M, Liu W, Lu D, Chen H, Ye C. Progress of China's new-type urbanization construction since 2014: A preliminary assessment. Cities. 2018 2018/08/01/;78:180-93
133. Patel V, Chatterji S, Chisholm D, Ebrahim S, Gopalakrishna G, Mathers C, et al. Chronic diseases and injuries in India. Lancet. 2011 Jan 29;377(9763):413-28

CHAPTER 2 THESIS AIMS, OBJECTIVES AND STUDY SETTING

2.1 Summary of identified research gaps

Greater urbanisation levels are consistently associated with greater levels of risk factors for atherosclerotic CVDs, such as SBP, hypertension, BMI, obesity, dyslipidaemia, FPG and diabetes. However, results between studies are rarely directly comparable due to substantial differences in exposure measures and categorisations. There are no standard definitions of urbanisation or urbanisation level. Most studies compare rural to urban or migrant populations who may be inherently different in aspects other than level of urbanisation (exposure). The frequent use of binary rural-urban comparison may mask variations in urbanisation-related characteristics, e.g. education or occupation, across the urbanisation level continuum. Several recent studies use physical and social data-based scales to categorise levels of urbanisation (usually low, medium and high levels). There is currently no consensus on the components of a scale that best capture levels of urbanisation. There is a need to identify a globally comparable standardised measure of urbanisation level that can be used in health research.

Labour-force transitions out of agriculture and into types of non-agricultural employment (often in more urbanised locations) are a first manifestation of urbanisation in LMICs. Associated changes in CVD risk factors at multiple levels, such as physical activity, income and health care access, are hypothesised to affect CVD incidence and prevalence, sometimes in opposite directions. Progress reports from China's recent urbanisation suggest that transitioning agrarians face a number of adverse economic and social pressures that may deteriorate their cardiovascular health. Studies that formally assess the associations of transitions out of agriculture and into 'more urbanised' types of employment with CVDs and risk factors are rare. There has to-date been no systematic review assessing the association of current engagement in agriculture compared to other types of employment (including types of employment that typically employ transitioning agrarians, for example in services and industry) with CVD incidence, prevalence or risk factors.

Several reviews have proposed different conceptual frameworks of potential pathways between urbanisation level and CVDs or related risk factors. The pathways are largely hypothesised by linking existing evidence along plausible pathways. Pathways between urbanisation level and CVD risk factors are sometimes explored by adjusting analyses for selected potential mediators in addition to confounders. While results are informative, strong assumptions underlie this approach to causal inference and the risk of bias is high. Appropriate mediation analyses are needed to quantify hypothesised pathways through which urbanisation levels might act on CVDs and CVD risk factors.

2.2 My PhD journey

Before outlining the final aims and objectives of my PhD research in Section 2.3, this section provides the background for their development. My PhD research initially set out to address the evidence gap pertaining to the association of transitioning out of agricultural and into more urbanised employment on CVD incidence, prevalence and risk factors in urbanising South India. Following the Food and Agriculture Organization's (FAO) definition, I defined 'engaging in agriculture' as being involved in (a) horticulture or agro-forestry, e.g. preparing the soil, planting, fertilising, weeding, watering or harvesting field or tree crops (pulses, roots and tubers, fruits, vegetables, nuts, herbs/spices, tree or palm sap, flowers etc.) on own or others' land and/or (b) animal husbandry (of domesticated animals such as poultry, cattle, swine, sheep, goat etc.), e.g. rearing, feeding, breading and caring for animals used for food, wool/fur or economic purposes, beekeeping, aquaculture, fishing and hunting

The APCAPS cohort provided data on current primary occupation according to pre-specified categories and supplementary free-text descriptions of participants' occupations (Table 2.1). Data were also provided on time spent in selected agricultural activities outside work (e.g. caring for animals and gardening) and ownership of agricultural assets (e.g. a plough or tractor). Several of the pre-specified occupation groups included agriculture and I used different strategies (including cross-referencing different agriculture-related variables and Finite Fixture Modelling on broader characteristics) to separate out participants who engaged in agriculture, either as a primary occupation, subsistence farming, or as a 'livelihood strategy' (see section 2.2.1). The final occupation categories were defined by cross-referencing the pre-specified current occupation and the supplementary free-text variables and as such reflected participants' current primary occupations.

Initial cross-sectional exploration among participants aged 30 years or older from the third survey wave suggested that participants who engaged in agriculture had lower levels of major CVD risk factors, such as tobacco use, physical inactivity and hypertension, than participants in non-agricultural work after adjusting for age and gender (data not shown). Unfortunately, on closer analysis there were not sufficient data overlap between the survey waves to perform longitudinal analysis of the association of shifts out of agriculture (and into other types of employment) with CVD incidence, prevalence or risk factors (Table 2.2). It was, for example, not possible to identify participants who engaged in agriculture in the first survey wave, as this data did not contain the free-text descriptions of occupations or other occupation-related variables that enabled identification of agricultural activities in the second and third survey waves. The time period between the second (2009-10) and third survey waves (2010-12) was

too short to perform meaningful longitudinal analysis on CVDs and risk factors that typically develop over longer periods of time.

Table 2.1 Occupation data collected in the third survey wave of the APCAPS

Survey code	Pre-specified occupation categories	Types of occupations or tasks included in the pre-specified occupation groups	Common participant description (recorded in free-text variable)	Includes agriculture
1	At home doing house work	Housework	Housewife, house work	N/A
2	Unemployed (retire/disabled)	Retired, disabled	Unemployed	N/A
3	Unemployed (seeking work)	Unemployed, seeking work	Unemployed	N/A
4	Student/training	Student/training	Student, vocational training	No
5	Unskilled manual labourer	Car cleaner, coolie, delivery boys (paper, milk etc.), garbage collectors, hawkers/vendors, landless labourers, packers, labellers, servant, sweeper/dhobi (washing and iron), other unskilled manual labourer, watchman/chowkidar/gate-keepers	Agricultural labour, agriculture animal rearing, farmer, labour work /daily wage earner, (less common: shopkeeper/vendor, hotel business, housekeeping, sweeper)	Yes
6	Semi-skilled manual	Barber, butcher, cobbler, farmer/gardener, fisherman, marginal landowner, peon, petty shopkeeper, rickshaw driver, other semi-skilled manual labour, sweet maker (halwai), welder/fitter	Agriculture, (small) businesses, driver, farming, hotel, own land agriculture, shop keeper/vendor, tailor	Yes
7	Skilled manual	Army jawan, blacksmith/goldsmith/engravers, carpenters/furnishers, driver, firefighters, hunters/trappers, machine and plant operators, mason, mechanic, painters/plumber, poultry farmers/animal rearer, sculptors/potters, spinners/weavers/carpet makers, street artist and performers/circus people	Farmer(small) business, barber, carpenter, centering work, electrician, tailor, agricultural labour (own land), fisherman, landlord, tractor driver	Yes
8	Skilled non-manual	Small business owner(<15 employees), alternative healers, big store keeper/shopkeeper, clerk/typist/stenographer/librarian, electrical repair works/electrician/watch, farm owner/landlord, makers, midwives/health visitors/field workers/vaccinators, musicians/dancers/artists (village level), home teachers, postmasters/telegraph masters, postman, receptionist (small organization), station masters and superintendents, telephone/telegraph operators, ticket collectors/sellers and examiners/bus conductors, x-ray technician/lab technicians/OT assistants	(Small) business, agricultural worker, landlord, shop keeper/vendor (less common: contractor, electrician, police)	Yes
9	Semi-professional	Accountants, administrators, diploma engineers, inspectors (police, school etc.)/agents (customs etc.), maintenance (in-charge), medium business owner (15-49 employees), music/dance/art teachers, teachers/college lecturers, nurse/pharmacist/dietician, personnel managers/junior, secretary/ receptionist (large organisation)	Teacher, business(man) (less common: bank, doctor, executive)	No

Survey code	Pre-specified occupation categories	Types of occupations or tasks included in the pre-specified occupation groups	Common participant description (recorded in free-text variable)	Includes agriculture
10	Professional	Bank managers/auditors, big business (>50 employees), class I IAS/IFS/IPS officers, doctors (allopathy, Ayurveda, homeopathy)/veterinarians, engineers/architects/designers, lawyers/judges/magistrate, musicians/dancers/artists (national/international level), newspaper editors, pilots/navigators, senior administrative officers/managing directors, university lectures/readers/ professors/principals	Government employment	No

Table 2.2 Data overlaps across survey waves of the Andhra Pradesh Children and Parents Study

Type of data	Available data from the first survey wave (2003-05)			Available data from the second (2009-10) and third (2010-12) survey waves
	Index child	Index child's mother	Index child's father	Index children, parents and siblings (completed the same individual survey)
Pre-specified occupation group	✓	x	✓*	✓
Free-text occupation description	x	x	x	✓
Tobacco use	✓	x	✓*	✓
Passive smoking	✓	x	x	x
Alcohol consumption	✓	x	x	✓
Diet	✓	x	x	✓
Physical activity	✓	x	x	✓
Blood samples	✓	x	x	✓
Anthropometry	✓	✓	✓ (mostly missing)	✓
Known cardiovascular disease	✓	x	x	✓
Blood pressure	✓	x	x	✓

✓ - recorded in survey; x – not recorded in survey

* Data provided by index child's mother

I decided to continue my work exploring links between employment in agriculture and CVD in the third data wave of the APCAPS and planned to conduct a supplementary qualitative study in India to identify underlying risk factors. I set out systematically to review the published evidence on my initial research question relating to shifts out of agriculture compared to remaining in agricultural and other types of employment over time with CVD incidence, prevalence and risk factors. However, my search returned only one relevant longitudinal analysis (summarised in Chapter 3) and I therefore amended the focus of my systematic review to the available published evidence on the associations of engaging in agriculture compared with types of non-agricultural employment with CVD incidence, prevalence and risk factors in

LMICs. During the review process and concurrent further exploration of the APCAPS data, I realised a number of limitations to conceptualising engaging in agriculture from data not collected for this purpose that I briefly summarise below.

2.2.1 Process of categorising occupation in the third APCAPS survey wave

In the APCAPS survey, a participant's current primary occupation was recorded according to 10 pre-specified categories supplemented with their own descriptions, captured in a free-text variable (Table 2.1). As agricultural workers were included in several of the pre-specified occupation groups, I cross-referenced these with the free-text variable to identify as many agricultural workers as possible. A large number of participants described their occupation as 'coolie', 'daily wage earner' or 'labour worker', without detailing the type of work typically performed. We followed-up with a group of APCAPS participants (n=20) to clarify the typical tasks performed by 'coolies', 'daily wage earners' and 'labour workers'. Coolies were generally reported to perform agricultural tasks, such as ploughing, planting and harvesting, but also non-agricultural tasks such as carrying, loading and unloading small or heavy loads. Daily wage earners were perceived to performed a more even mix of agricultural and other activities, whereas labour workers appeared to almost exclusively do non-agricultural work, such as building and construction; carrying, loading and unloading. Unless participants described themselves as 'agricultural labour workers' or 'agricultural coolies', 'coolies', 'daily wage earners' and 'labour workers' were combined in one occupation category.

Some participants, who did not report their primary occupation to be agriculture, reported doing agriculture-related physical activities (ploughing, harvesting, watering or weeding fields) during the two weeks preceding the survey[9], or owning agricultural assets, such as agricultural land, a tractor or a thresher (Table 2.3). I expected some seasonal diversification of livelihood strategies in the study population[1] and did not attempt to separate these individuals out. Before finalising the occupation categories, I explored an alternative data driven approach to categorising occupations in collaboration with Dr Rosemary Green. Dr Green used Finite Fixture Modelling to predict clusters (or groupings) of participants from a board range of livelihood-related variables, e.g. pre-specified occupation group, working hours, occupation and leisure activities and assets ownership. Gender-specific and overall Finite Fixture Modelling models were run with full specification, including all available livelihood-related variables; and with reduced specification, excluding variables with large proportions of missing data and collapsing occupation groups to five categories (i) agriculture, (ii) daily wage and labour workers, (iii) manual workers, (iv) non-manual workers and professionals, (v) unemployed & students. The

[9] In the physical activity questionnaire that had a two-week recall period

results did not indicate any clear occupation-related groupings of participants and I thus returned to refining the aforementioned occupation variable. Table 2.4 lays out the final occupation variable used in the research of this thesis, together with the rationale behind each category and potential limitations.

Table 2.3 Agriculture related household activities and ownership of agricultural assets by occupation groups in the third data wave of the Andhra Pradesh Children and Parents Study, 2010-12

	Household activities		Household asset ownership				Any agricultural Household activity or asset
	Animal care	Gardening	Land	Bullock cart	Tractor	Thresher	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Agricultural labourer	18 (10.7)	5 (3.0)	5 (3.0)	0 (0)	0 (0)	0 (0)	26 (15.5)
Agriculture own land	171 (19.4)	56 (6.4)	880 (99.8)	27 (3.1)	38 (4.3)	5 (0.6)	881 (99.9)
Unemployed	12 (7.6)	16 (10.1)	85 (52.2)	1 (0.6)	5 (3.1)	1 (0.6)	97 (59.5)
Student/training	64 (9.7)	104 (15.7)	375 (57.1)	13 (2.0)	20 (3.0)	5 (0.8)	436 (65.9)
Housework	38 (4.3)	115 (13.0)	408 (46.1)	7 (0.8)	18 (2.0)	4 (0.5)	478 (54.0)
Daily wage earner	103 (12.3)	23 (2.8)	390 (46.7)	9 (1.1)	8 (1.0)	0 (0)	429 (51.2)
Coolie	99 (8.7)	68 (5.9)	515 (45.0)	4 (0.4)	5 (0.4)	1 (0.1)	572 (50.0)
Unskilled manual	14 (5.9)	11 (4.7)	102 (43.2)	2 (0.9)	3 (1.3)	2 (0.9)	109 (46.2)
Semi-skilled manual	20 (6.1)	32 (9.8)	170 (52.2)	0 (0)	9 (2.8)	1 (0.3)	196 (59.9)
Skilled manual	33 (5.1)	46 (7.1)	318 (49.1)	6 (0.9)	11 (1.7)	1 (0.2)	355 (54.7)
Skilled non-manual	10 (5.2)	10 (5.2)	99 (51.8)	1 (0.5)	0 (0)	0 (0)	102 (53.1)
Semi-professional	5 (3.6)	14 (9.9)	77 (55.0)	0 (0)	1 (0.7)	0 (0)	84 (59.6)
Government	0 (0)	3 (16.7)	8 (47.1)	0 (0)	0 (0)	0 (0)	10 (55.6)

n – sample size

Occupation groups were derived from a combination of data collected by pre-specified occupation codes and free-text descriptions

Table 2.4 Description and rationale for the categorisation of occupation for use in the PhD research

Occupation category	Includes	Excludes	Rationale
People engaging/employed in agriculture as their self-reported main occupation, including for subsistence	Agricultural workers working on own, leased, or someone else's land, identified by the free-text occupation description. This group likely includes seasonal workers currently working in agriculture but may miss those interviewed out of the agricultural season	Participants describing themselves as non-agricultural workers but reported doing agricultural activities e.g. caring for animals and gardening in the previous two weeks	Landowners may benefit from access to the produce for own consumption and income generation from selling excess or cash crops. Landless agricultural contract workers and agricultural day labourers, who are unlikely to share these benefits, may be more vulnerable and perform less desirable tasks at low wages and higher health risk. However, it is common for landless agricultural workers in the current context to lease land or sharecrop with other farmers and thus live off the land with its associated benefits. We do not have information on land leasing and share cropping and thus cannot separate agricultural workers who live off the land from those who do not. In addition, land is

Occupation category	Includes	Excludes	Rationale
			commonly suggested a stronger driver of wealth than agriculture in the current context
Daily wage employment	Coolies, daily wage earners and day labourers identified by the free-text occupation description. This group likely includes some seasonal agricultural workers interviewed out of the agricultural season	Participants describing themselves as agricultural day labourers or coolies	Daily wage earners and coolies perform various strenuous and less desirable tasks at low wages and higher health risk on a day-to-day basis. The irregular and insecure livelihood strategies of this group of people may be associated with greater vulnerability to stresses and shocks compared to most other categories
Unskilled manual work	Unskilled manual workers and housework identified by the pre-specified occupation code. Seasonal agricultural workers interviewed out of the agricultural season are likely also included	Participants describing themselves as agricultural or daily wage workers in the free-text variable	This group will be a mix of daily wage workers, self-employed and some (lower) salary workers. Unskilled-manual workers include individuals performing less-desirable manual tasks with low skills requirements, e.g. street sweepers, garbage collectors and servants. These workers are likely employed in low-grade work in the informal sector, and are unlikely to enjoy the employment-related benefits of formal or higher-grade work. This group includes 'housewives', 'homemakers' and 'housekeepers'
Self-employed	Semi-skilled and skilled manual workers identified by the pre-specified occupation code		Semi-skilled and skilled workers include individuals who use acquired occupational skills to earn a living from providing a service or selling a product, e.g. electricians, shopkeepers and drivers. This group is likely dominated by self-employed workers (informal sector), who do not enjoy benefits of formal (contract) employment. This group has a high proportion of manual workers
Salaried workers	Skilled non-manual and professional workers identified by the pre-specified occupation code		Salaried workers include skilled manual workers and professionals who are more likely to be employed in contract work and therefore may enjoy higher livelihood security and more occupation-related benefits, e.g. pension or health insurance, than other occupation groups
Students and unemployed	Students and unemployed identified by the pre-specified occupation code. This group includes students at all levels and unemployed individuals either seeking work or unable to seek work (e.g. due to disability)		Students and the unemployed are highly reliant on other HH members. This group may be highly heterogeneous in their characteristics

Following the categorisation process, I reconsidered whether cross-sectional analysis using the derived occupation variable would contribute new knowledge to the existing evidence base (appraised in the systematic review), and considered if other types of analyses might be more informative for CVD prevention in South India. The uneven urbanisation experienced by the 28 APCAPS villages provided me with a unique opportunity to contribute to existing knowledge on the impacts of urbanisation levels on CVDs and risk factors and identify important pathways.

I therefore continued my PhD research with the aim to understand associations of increasing levels of urbanisation with CVD, incidence, prevalence and risk in India. I considered using the second and third APCAPS data waves in longitudinal analysis of the association of urbanisation level change (measured by remote sensing data described in Chapters 4 and 5) with CVDs and risk factors[10]. Technical issues with the satellite data from 2010 and 2012 however prevented this approach. The third data wave of the APCAPS cohort provided an opportunity to assess associations of urbanisation levels with self-reported clinician-diagnosed CVDs. The prevalence of CVD was low in the study population and would have underpowered statistical analyses using NTLI in the multi-level models that are needed to account for cluster sampling of participants. Therefore, for my outcome variables I used the clinician-measured information on individual-level risk factors with established strong associations with CVDs, such as blood pressure and fasting plasma glucose. My revised final aims and objectives for this PhD research are outlined in the following section.

2.3 Thesis aims and objectives

2.3.1 Aims

The aim of the thesis was to assess the association of place of residence on CVD incidence, prevalence and risk factors, and quantify potential pathways during early stages of urbanisation in South India. A research paper-style PhD thesis was prepared to address research gaps with the overarching aim to contribute to curbing expected rises in premature death and disability from CVDs with continued, rapid urbanisation in South India. The thesis aim was addressed in three connected research papers addressing three objectives.

2.3.2 Objective 1

Systematically review the available published evidence on the associations of engaging in agriculture compared with types of non-agricultural employment (across the urbanisation continuum) with CVD incidence, prevalence and risk factors in LMICs

2.3.3 Objective 2

Explore associations of NTLI, a novel, globally comparable, continuous proxy for level of urbanisation, with mean levels of SBP, BMI, fasting serum LDL, and FPG among adults at early stages of urbanisation in Telangana, South India

[10] This would be done by matching survey years during the second and third data waves (2009-12), with remote sensing derived urbanisation levels during those years

2.3.4 Objective 3

Decompose the total effect of level of urbanisation on SBP into direct and indirect effects via hypothesised pathways, among adults at early stages of urbanisation in Telangana, South India.

2.4 Study setting and data

The following section provides supplementary information on the study setting and datasets used in the second and third research papers. Data were combined from two sources (i) demographic and health data from the third wave of the APCAPS conducted in 2010-12 and (ii) remote sensing satellite data on NTLI from the United States' Defence Meteorological Satellite Programme's Operational Linescan System (DMSP-OLS) measured in 2011.[2] See the submitted research papers described in Chapters 4 and 5 for details on the NTLI data.

2.4.1 The Andhra Pradesh Children and Parents Study (APCAPS)

The APCAPS is a prospective, trans-generational cohort study, set in 29 villages 30-80 km from Hyderabad in the Ranga Reddy district of Telangana (formerly Andhra Pradesh), South India (Figure 2.2).[3] The first survey wave of the cohort was between 2003 and 2005 with the first follow-up of the Hyderabad Nutrition Trial (HNT) (1987-90) (Box 2.1).[4] The first wave of the APCAPS followed up all available mothers and children who participated in the HNT, with the aim to explore 'the developmental origins of health and disease hypothesis' concerned with the effects of early life nutrition on chronic disease risk throughout the life course. The second wave of the APCAPS followed up the birth cohort in early adulthood in 2008-10 to explore 'the effect of nutritional shortage in early life on the amount and distribution of body fat, and the development of type 2 diabetes and early markers of coronary heart disease'. The third and most recent wave of the APCAPS was completed in 2010-12. The sample additionally included all available parents and siblings of the original HNT birth cohort (Figure 2.3), to enable analyses of 'how nutritional supplementation, environment, lifestyle, physical activity and dietary habits are associated with health and chronic disease, particularly cardiovascular disease'.[4, 5] Trained field staff and medical doctors collected the APCAPS data in accordance with the study protocol. Standardised semi-structured questionnaires in the local language (Telugu) were used to obtain information on socio-demographic characteristics, health, diet, and physical activity. Venous blood samples, blood pressure and anthropometric measurements were taken in local clinics and at the National Institute of Nutrition in Hyderabad. Details on data collection, biochemical analysis and data management are provided in the publications described in the research paper presented in Chapters 4 and 5.

Box 2.1 The Integrated Child Development Service Scheme and the Hyderabad Nutrition Trial

In 1976, the Integrated Child Development Service (ICDS) initiated a national community outreach program aiming to improve the nutrition, health and development of Indian children. The ICDS scheme offered corn-soya blend and soybean oil based (energy-protein) food supplements to pregnant women and young children (<6 years), together with early childhood, nutrition, health and hygiene education. Implementation of the intervention was combined with the delivery of existing national ICDS programs involving immunisation, anaemia control and provision of basic health care services. During 1987-90, The National Institute of Nutrition (NIN) in Hyderabad conducted the Hyderabad Nutrition Trial (HNT) to evaluate the nutritional outcomes of children born by mothers who had received the ICDS intervention (n=15 villages) and children born by mother who were awaiting the ICDS intervention (14 villages) (see Figure 2.2).

Source Kinra et al 2013

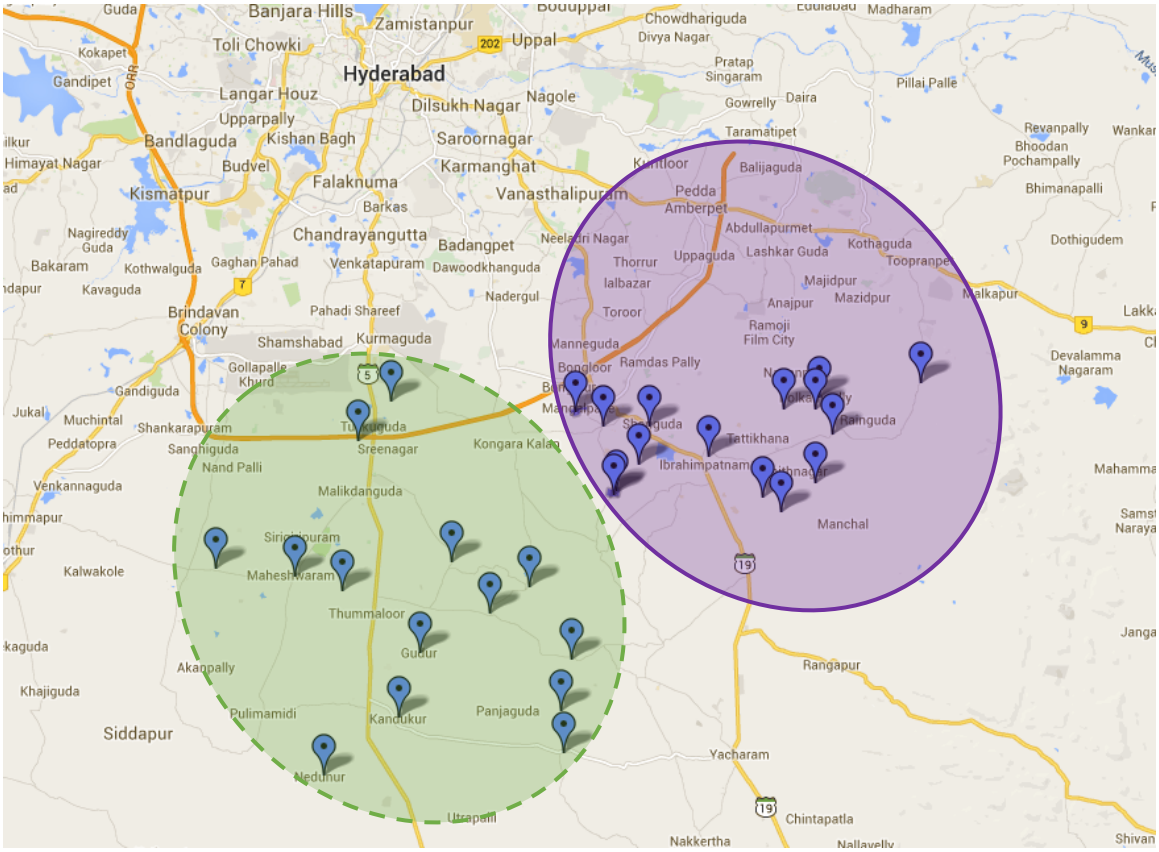


Figure 2.1 APCAPS villages (n=29) in the Ranga Reddy district of Andhra Pradesh (now Telangana), South India

Source: Google Maps by Santhi Bhogadi, APCAPS project manager

Green oval (– –) = Intervention villages (n=15); Purple oval (——) = Control villages (n=14)

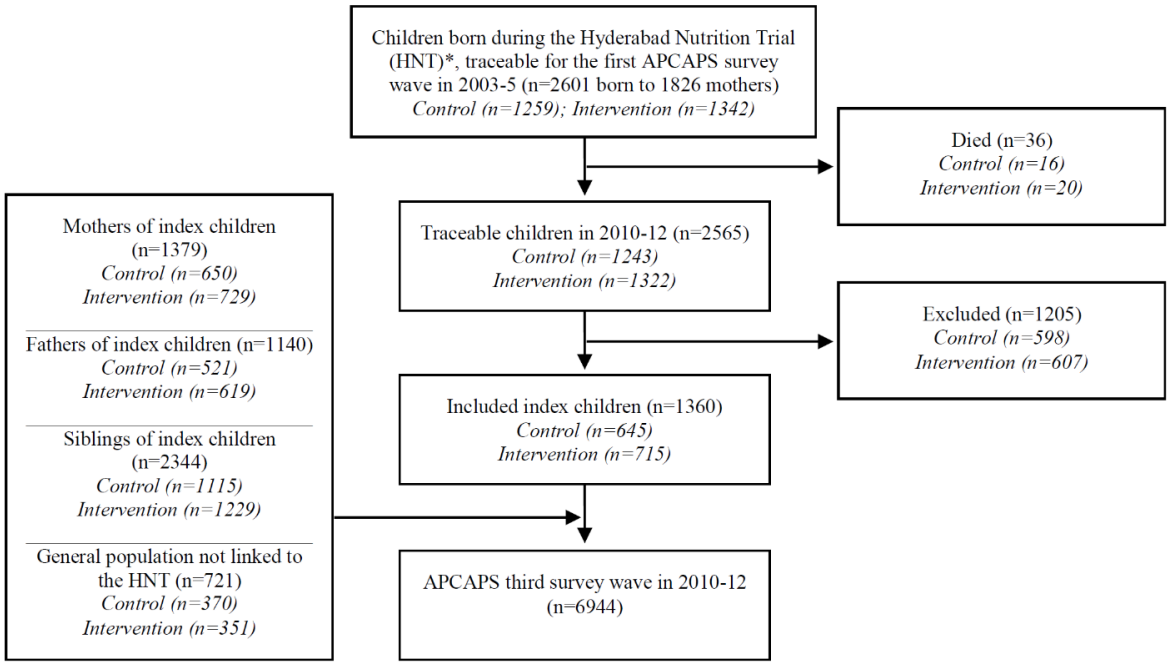


Figure 2.2 Selection of participants for the third wave of the APCAPS

APCAPS – Andhra Pradesh Children and Parents Study; HNT – Hyderabad Nutrition Trial

2.4.2 Representativeness of the third survey wave of the APCAPS

The HNT was not designed for long-term follow up. As a result 2601 of an estimated 4338 children born by mothers who participated in the HNT (60%) were traceable for the inclusion in the APCAPS cohort in 2003-2005. The HNT employed an ecological design and a lack of a reliable identifier prevented the linkage of all traceable mothers and children to their collected socio-demographic and health information[11]. It was therefore not possible to ascertain if any major differences in characteristics existed between traceable and untraceable HNT participants. A total of 1360 HNT children (an estimated 31% of the children born during the HNT between 1987 and 1990), their available parents and siblings participated in the third follow-up. The proportion of attrition across the three APCAPS follow-ups did not appear to differ by trial status.[4] It was however not possible to assess differences in attrition over time by sample characteristics from the third survey wave data used in the current PhD analyses.

[11] Socio-demographics; medical and obstetric history for women; feeding and immunisation history for offspring; nutritional supplementation and anthropometrics for both

2.4.3 Estimating urbanisation levels of APCAPS villages

The APCAPS village boundaries were defined by aerial tracing of Open Street Map satellite imagery and GPS-based ground surveying by the APCAPS field team. The village boundaries were then overlaid with remote sensing data on NTLI, extracted and managed by Dr Robin Wilson (University of Southampton). Several villages were intersected by two or more NTLI pixels, which in some cases had different NTLI values. To ensure the appropriate weighting of each pixel present over a village (i.e. minimise the inclusion of light intensity from areas beyond a village boundary also present in some NTLI pixels), the super resolution method was applied to resample the data to a higher resolution (≈100m^2).[6] Validation studies of NTLI as a proxy for urbanisation levels, report strong correlations of the sum of NTLI values over given areas with urbanisation dimensions such as population density, economic development and activities, extent of build-up area and energy consumption.[4-27] The NTLI values were therefore summed over each APCAPS village to obtain one village level estimate for each village. To this end, the final village-level NTLI values (urbanisation level) used in my analyses also included an element of village size. The derived urbanisation levels (NTLI sums) was strongly correlated with a multi-component urbanisation scale constructed by researchers of the APCAPS team (study in peer-review) as well as the proportion of village residents working in non-agricultural employment or owning motorised vehicle derived from a household census of APCAPS villages in 2013 (results available from the author on request).

References

1. Dodd W, Humphries S, Patel K, Majowicz S, Little M, Dewey C. Determinants of internal migrant health and the healthy migrant effect in South India: a mixed methods study. BMC Int Health Hum Rights. 2017 Sep 12;17(1):23
2. Google Earth Engine. Earth Engine Data Catalog. DMSP OLS: Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System na [cited 2019 22 October]. Available from: https://developers.google.com/earth-engine/datasets/catalog/NOAA_DMSP-OLS_NIGHTTIME_LIGHTS.
3. Andhra Pradesh Children and Parents Study (APCAPS). Andhra Pradesh Children and Parents Study: London School of Hygiene and Tropical Medicine; 2018 [cited 2018 12 July]. Available from: http://apcaps.lshtm.ac.uk/.
4. Kinra S, Radha Krishna K, Kuper H, Rameshwar Sarma K, Prabhakaran P, Gupta V, et al. Cohort Profile: Andhra Pradesh Children and Parents Study (APCAPS). International journal of epidemiology. 2013 Sep 9
5. Kinra S, Rameshwar Sarma KV, Ghafoorunissa, Mendu VV, Ravikumar R, Mohan V, et al. Effect of integration of supplemental nutrition with public health programmes in pregnancy and early childhood on cardiovascular risk in rural Indian adolescents: long term follow-up of Hyderabad nutrition trial. BMJ (Clinical research ed). 2008;337:a605
6. Zhang Q, Pandey B, Seto KC. A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data. IEEE Transactions on geoscience and remote sensisng. 2016

CHAPTER 3 IS AGRICULTURAL ENGAGEMENT ASSOCIATED WITH LOWER INCIDENCE OR PREVALENCE OF CARDIOVASCULAR DISEASES AND CARDIOVASCULAR DISEASE RISK FACTORS? A SYSTEMATIC REVIEW OF OBSERVATIONAL STUDIES FROM LOW- AND MIDDLE-INCOME COUNTRIES

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Tina Bonde Sorensen
Principal Supervisor	Professor Alan Dangour
Thesis Title	Influences of place of residence on risk factors for atherosclerotic cardiovascular diseases in South India

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLOS ONE		
When was the work published?	1 April 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Tina Bonde Sorensen developed the inclusion/exclusion criteria, the risk of bias assessment, the data extraction form, drafted the protocol, developed and performed the searches, performed the analysis and drafted the manuscript with input from all contributing authors. Tina Bonde Sorensen and Dr Mika Matsuzaki independently

	screened titles and abstracts against the inclusion/exclusion criteria. Tina Bonde Sorensen extracted data from original publications for analysis and Dr Mika Matsuzaki double-checked the extracted data against the original publications. Professor Alan D Dangour provided expertise on the links between agriculture, nutrition and health and Professor Sanjay Kinra provided expertise on cardiovascular diseases and epidemiology. Dr John Gregson advised on the analysis strategy. All authors contributed substantially to revising the manuscript critically, approved the final manuscript for publication and agreed to be accountable for the work.

Student Signature: [redacted] **Date:** 25 May 2020

Supervisor Signatu [redacted] **Date:** 25 May 2020

Full title: Is agricultural engagement associated with lower incidence or prevalence of cardiovascular diseases and cardiovascular disease risk factors? A systematic review of observational studies from low- and middle-income countries

Short title: Agricultural engagement, cardiovascular diseases and cardiovascular risk factors: A systematic review

Tina B Sørensen,[1*] Mika Matsuzaki,[1] John Gregson,[1] Sanjay Kinra,[2] Suneetha Kadiyala,[1,3] Bhavani Shankar[3,4] and Alan D Dangour[1,3]

[1]Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

[3]London Centre for Integrative Research in Agriculture and Health (LCIRAH), London School of Hygiene & Tropical Medicine, London, United Kingdom

[4]Centre for Development, Environment and Policy, School of Oriental and African Studies, London, United Kingdom

*Corresponding author, E-mail: tina.sorensen@lshtm.ac.uk.

Author contributions

Dr Tina Bonde Sorensen developed the inclusion/exclusion criteria, the risk of bias assessment, the data extraction form, drafted the protocol, developed and performed the searches, performed the analysis and drafted the manuscript with input from all contributing authors. Dr Tina Bonde Sorensen and Dr Mika Matsuzaki independently screened titles and abstracts against the inclusion/exclusion criteria. Dr Tina Bonde Sorensen extracted data from original publications for analysis and Dr Mika Matsuzaki double-checked the extracted data against the original publications. Professor Alan D Dangour provided expertise on epidemiology and the links between agriculture, nutrition and health, and Professor Sanjay Kinra provided expertise on epidemiology and cardiovascular diseases. Dr John Gregson advised on the analysis strategy. All authors contributed substantially to revising the manuscript critically, approved the final manuscript for publication and agreed to be accountable for the work.

Conflict of interest: The authors declare that they have no conflict of interest.

Funding

This study was funded by the Bloomsbury Colleges PhD Studentship (http://www.bloomsbury.ac.uk/studentships) to TBS via LSHTM. ADD and BS acknowledge the Sustainable and Healthy Food Systems (SHEFS) programme supported by the Wellcome Trust's 'Our Planet, Our Health' programme (grant no. 205200/Z/16/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Non-communicable diseases, such as cardiovascular diseases (CVDs), diabetes and cancer account for more than half of the global disease burden, and 75% of related deaths occur in low- and middle-income countries (LMICs). Despite large regional variations in CVD incidence and prevalence, CVDs remain the leading causes of death worldwide. With urbanisation, developing nations are undergoing unprecedented labour-force transitions out of agriculture and into types of non-agricultural employment, mainly in the industry and service sectors. There are few studies on the effect of these transitions on CVDs and CVD risk factors in LMICs. We systematically searched MEDLINE, PubMed, EMBASE and the Cochrane Library from January 1950 to January 2017 to assess the association of engaging in agriculture compared to types of non-agricultural employment (e.g. services and manufacturing) with CVD incidence, prevalence and risk factors. Studies were included if they: included participants who engaged in agriculture and participants who did not engage in agriculture; measured atherosclerotic CVDs

or their modifiable risk factors; and involved adults from LMICs. We assessed the quality of evidence in seven domains of each study. Prevalence ratios with 95% confidence intervals were calculated and compared in forest plots across studies. Study heterogeneity did not permit formal meta-analyses with pooled results. There was a lack of publications on the primary outcomes, atherosclerotic CVDs (n=2). Limited evidence of varying consistency from 13 studies in five countries reported that compared with non-agricultural workers, mainly living in urban areas, rural agriculture workers had a lower prevalence of hypertension, overweight and obesity; and a higher prevalence of underweight and smoking. High quality evidence is lacking on the associations of engaging in and transitioning out of agriculture with atherosclerotic CVDs and their modifiable risk factors in LMICs. There is a need for interdisciplinary longitudinal studies to understand associations of types of employment and labour-force transitions with CVD burdens in LMICs.

Key words: Systematic review; epidemiology; agriculture; employment; occupation; cardiovascular disease (CVD); CVD risk factor; hypertension; overweight; obesity; underweight; tobacco; smoking; low- and middle-income countries (LMICs)

List of abbreviations

Agri	agriculture
BMI	body mass index
CI	confidence interval
CVD	cardiovascular disease
Govt	government
HH	household
LMIC	Low- and middle-income country
LSHTM	London School of Hygiene and Tropical Medicine
n	sample size
NCD	non-communicable disease
OR	odds ratio
PICOS	participants, interventions, comparisons, outcomes, and study design
PR	prevalence ratio

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
R	rural
Stud	student
UK	United Kingdom

List of abbreviations, online supporting information

% BF	percent body fat
BMI	body mass index
CHD	coronary heart disease
cm	centimetre(s)
CVD	cardiovascular disease
d	day
DBP	diastolic blood pressure
EE	energy expenditure
g	gram
HDL	high-density lipoprotein cholesterol
HH	household(s)
INCAP	Nutrition of Central America and Panama longitudinal study
Kcal	kilocalories
kg	kilograms
KJ	kilojoule
l	litre
LDL	low-density lipoprotein cholesterol
LMIC	low- and middle-income country
m	metre(s)

mg	milligram
min	minute
mmHg	millimetre mercury
mmol	millimoles
n	sample size
N/A	not available
PA	physical activity
SBP	systolic blood pressure
TC	total cholesterol
TG	triglycerides
WC	waist circumference
WHO	World Health Organization
WHR	waist-to-hip ratio
YMS	Yi Migrant Study
µg	microgram

Introduction

Non-communicable diseases (NCDs), such as cardiovascular diseases (CVDs), diabetes and cancers are the leading causes of death and disability worldwide, accounting for more than half of the global disease burden.[1] Almost 75% of NCD-related deaths occur in low- and middle-income countries (LMIC),[2] often among working-age adults as young as 40 years.[3] Although disease patterns vary across world regions, CVDs remain the leading causes of death throughout.[4] Most CVDs develop from atherosclerosis (the hardening and narrowing of major blood vessels).[5] As such, atherosclerotic CVDs are largely preventable by addressing risk factors including unhealthy diets, physical inactivity, harmful use of alcohol, tobacco use, hypertension, overweight and obesity, diabetes and dyslipidaemia.[5-7] Urbanisation is demanding still more non-agricultural labour, and LMICs have undergone unprecedented labour-force transitions out of agriculture and into the industry and service sectors (Table 1).[8, 9] Labour-force transitions out of agriculture have been particularly steep in middle-income countries such as China, however, large low-income countries, such as India, are quickly catching up.

Table 3.1 Percent employment in agriculture, services and industry of total employment in low-and middle-income countries

	% of total employment		
	1991	**2004**	**2018**
Low-income countries			
Agriculture	71	70	63
Services	20	21	26
Industry	9	9	11
Low- and middle-income countries			
Agriculture	53	46	34
Services	27	34	43
Industry	19	20	23
Upper middle-income countries			
Agriculture	48	37	22
Services	29	38	52
Industry	24	25	27
India			
Agriculture	63	57	44
Services	22	25	31
Industry	15	18	25

Source World Bank Group[10]

Type of employment is an important social determinant of health.[11] Types of employment contribute significantly to shaping the conditions of daily life that strongly associate with major immediate CVD risk factors such as hypertension, obesity, diabetes and dyslipidaemia.[11-13] For

example, the type, amount and stability of labour and income influence people's ability to acquire diverse and nutritious foods (through access to own-produce or purchase), assets and health-services.[13, 14] The type and duration of labour influence physical activity, nutritional needs, time available for food preparation and levels of exposure to biological and chemical hazards.[13, 15, 16] The environment in which people live and work influence the availability, access and affordability of commodities that may have beneficial or harmful effects on cardiovascular health, such as fruits and vegetables, highly processed energy-dense foods, tobacco and alcohol.[13, 16-18]

Agriculture has the potential to benefit nutrition and cardiovascular health, through the increased production and availability of nutritious foods and higher levels of physical activity that is associated with agricultural labour.[14, 15] Engaging in agriculture is also associated with cardiovascular health risks such as prolonged exposure to disease vectors, food borne diseases and toxic pesticides.[13] Most systematic reviews on the links between agriculture, nutrition and cardiovascular health synthesise evidence from studies that introduce, improve or intensify agriculture. In light of the expected continued labour-force transitions away from agriculture in LMICs, this paper aims to address two additional questions (i) is engaging in agriculture compared to types of non-agricultural employment (e.g. services and manufacturing) associated with lower levels of CVD incidence, prevalence and risk factors? (ii) Is the process of transitioning out of agriculture and into types of non-agricultural employment associated with higher levels of CVD incidence, prevalence and risk factors? Our initial systematic search returned only one eligible study pertaining to question (ii) and we therefore set out systematically to review the published evidence of the associations of engaging in agriculture compared to (any) types of non-agricultural employment with CVD incidence, prevalence or CVD risk factors in LMICs. We hypothesised that people who engage in agriculture have lower levels of CVDs and associated risk factors (not considering chemical, ambient and noise pollutants) than individuals who engage in other types of labour, particularly types of sedentary work. A review of the evidence might identify types of higher-risk employment, and with that, provide guidance to categorising employment in future longitudinal studies of employment transitions out of agriculture.

Methods

This systematic review asked the following questions (i) is engaging in agriculture compared to types of non-agricultural employment (e.g. services and manufacturing) associated with lower levels of CVD incidence, prevalence and risk factors?; and (ii) is the process of transitioning out of agriculture and into types of non-agricultural employment associated with higher levels of CVD incidence, prevalence and risk factors? The review is reported according to the Preferred

Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [19] (see S1 Table for PRISMA checklist) and the Guidelines for Meta-Analyses and Systematic Reviews of Observational Studies.[20] The protocol was published in advance (ID=CRD42015025488).[21] Eligibility criteria were defined in relation to PICOS (participants, interventions, comparisons, outcomes, and study design) as recommended by PRISMA.

Eligibility criteria

Population

We included studies that reported on individuals i) from at least one LMIC as defined by the World Bank at the time of the study[22] and ii) aged 15 years and above or described as 'adults', 'men' or 'women'.

Interventions, exposures and comparators

Following the Food and Agriculture Organization's (FAO) definition, we defined 'agriculture' as horticulture and agro-forestry (e.g. preparing the soil, planting, fertilising, weeding, watering or harvesting food or other crops) as well as animal husbandry (e.g. rearing, feeding, breading and caring for animals used for food, wool/fur or economic purposes), beekeeping, aquaculture, fishing and hunting.

Studies were eligible for inclusion if one group of participants reported to engage in agriculture on their own or someone else's land, either as a primary occupation or by predominantly depending on agriculture for their livelihood. Studies were included if they had at least one comparator group of participants who reported to not engage in agriculture as defined above. To reduce contamination of comparator groups, we excluded studies that sourced comparator groups from 'agricultural communities' or similar without specifying if participants engaged in agriculture.

Outcomes

Studies were included if they measured one or more atherosclerotic CVDs (primary outcomes) or related modifiable risk factors (secondary outcomes) (Box 1).[6, 23-25] When multiple publications analysed data from one study, we included results for all unique outcomes. When multiple publications presented overlapping analyses from the same study, we included results from the most comprehensive analysis based on methods and sample size.

Box 2 Atherosclerotic cardiovascular diseases (primary outcomes) and their modifiable risk factors (secondary outcomes)

Atherosclerotic cardiovascular diseases and events

Ischaemic heart disease or coronary artery/heart disease, for example heart attack
Cerebrovascular disease, for example stroke
Peripheral vascular disease
Deep vein thrombosis
Pulmonary embolism
Unspecified cardiovascular disease (CVD)
Specified or unspecified CVD mortality

Diet

Saturated fat, trans fat, cholesterol, carbohydrate (including sugar), fibre, antioxidants: vitamin C, E, Ubiquinone (coenzyme Q10), bioflavonoids, selenium, folate, vitamin B6, vitamin B12, potassium, fruits and vegetables, whole grain cereals, unsalted nuts, fish, salt/sodium

Physical activity

(Low) physical activity

Tobacco and alcohol

Any or harmful alcohol consumption
Any tobacco use

Metabolic cardiovascular disease risk factors

Body mass index
Overweight and obesity
Underweight
Systolic and diastolic blood pressure
Hypertension
(High) total cholesterol
(Low) high-density lipoprotein
(High) low-density lipoprotein
Total cholesterol:high-density lipoprotein ratio
(High) triglycerides
Fasting glucose
Impaired fasting glucose
Diabetes Mellitus Type II
Augmentation index
Carotid intima-media thickness
Homocysteine
High sensitivity C - reactive protein
(High) serum apolipoprotein B
(Low) serum apolipoprotein A-I

Composite measures

CVD risk score, e.g. the Framingham 10-year coronary heart disease risk score
Homeostatic model assessment
Dyslipidaemia
Metabolic syndrome

CVD – cardiovascular disease

Sources[6, 23-25]

Study design

We included comparative studies of any design and duration. No restriction was placed on sample size in the initial review phase. However, because of their limited generalisability and power, case studies and studies with small sample sizes (typically between 30 and 50 participants) were excluded. The characteristics of excluded studies can be found in S3 Table in the online information.

Data

We included studies that, as a minimum, reported crude estimates of associations, e.g. prevalence ratios (PRs) and 95% confidence intervals (CIs), or enough information to calculate them. Studies that in addition to crude estimates provided adjusted estimates (of any type) were eligible if they did not also adjust for mediators (factors on the causal pathway between employment and CVDs or risk factors) in non-mediation analyses. We restricted analyses to risk factors that were reported in four or more studies and for which clear descriptions of measurement methods and outcome categorisations were provided. In addition to the main analysis, we summarised studies on primary outcomes, atherosclerotic CVDs, if they met eligibility criteria other than that relating to 'four or more available studies'.

Information sources and search strategy

Our search was conducted in January 2017 and searched databases dating back to January 1950. We systematically searched the electronic databases MEDLINE, PubMed, EMBASE and the Cochrane Library by using key words and Medical Subject Headings delimiting 'agriculture' and 'CVD' or 'CVD risk factors' (S2 Table). The search was limited to human subjects and texts in English, Danish, Norwegian, Swedish, German and Spanish. We manually searched bibliographies of included primary publications, relevant reviews and supplementary grey literature (the latter identified from Google and Google Scholar searches). We included only primary peer-reviewed literature in the final review.

Selection process and data extraction

We imported and managed citations in Endnote X7. Two investigators (TBS and MM) individually screened all titles and abstracts against the eligibility criteria and resolved any disagreement in study selection by discussion. One reviewer (TBS) used data extraction forms that were designed for the review to extract data on PICOS and data were imported into Excel 2016. A second reviewer (MM) double-checked the extracted data against the original publications.

Quality of the evidence

We adapted 'A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI)[26] to assess the quality of evidence. The assessment covered seven risk of bias domains: confounding, selection of study participants, measurement of exposures/interventions, departures from intended interventions, missing data, measurement of outcomes and selection of the reported results. We rated the quality of evidence within each domain and overall at the study level as 'well covered', 'adequately addressed', 'poorly addressed', 'not applicable', 'not described adequately to classify' or 'not described'.

Compliance with ethical standards

Ethical approval for the current systematic review was obtained from the London School of Hygiene and Tropical Medicine, London, United Kingdom. Informed consent was obtained from all participants in the original publications, from which the data for the current review were extracted.

Analysis

Where outcome units differed between studies, we converted results to a common unit using biomedical research conversion tables.[27] Some studies described one risk factor with multiple estimates, e.g. mean body mass index (BMI) and prevalence of overweight and obesity. The type of information reported by most studies were included in the analysis, resulting in the exclusion of continuous data. Characteristics of studies that were excluded from the analysis as a result of these restrictions are available in S3 Table in the online information.

We used formulas suggested by Sterne (Equation (1))[28] to calculate PRs with 95% CIs and produce forest plots that graphically describe patterns of outcomes by livelihood or occupation groups across studies. Adjusted PRs (95% CIs) could not be calculated because we did not have access to raw datasets. Narrative analyses of adjusted estimates were presented separately. It was not appropriate to perform formal meta-analyses or generate funnel plots because of the substantial heterogeneity of measurement methods, categorisation of exposures and outcomes, study settings and populations of the included studies. All analyses were performed in Stata 14.

$$\text{Log prevalence ratio} = \log\left(\left[\frac{\text{exposed cases}}{\text{total exposed}}\right] \Big/ \left[\frac{\text{unexposed cases}}{\text{total unexposed}}\right]\right)$$
$$\text{Standard error of log prevalence ratio} = \sqrt{\left(\frac{1}{\text{exposed cases}} + \frac{1}{\text{unexposed cases}} - \frac{1}{\text{total exposed}} - \frac{1}{\text{total unexposed}}\right)} \quad (1)$$

Results

Our search yielded 3159 records, including 166 studies identified through other sources, such as manually searching bibliographies and contacting authors (Fig 1). After removing duplicates (n=995) we screened 2164 titles and abstracts and reviewed 189 full-text publications against inclusion/exclusion criteria. We included 13 publications that reported data from 12 unique studies. Only one eligible study provided appropriately adjusted estimates (i.e. not including mediators in non-mediation analysis). Adjusted results on hypertension from this study is presented separately. A summary of identified studies on primary outcomes, atherosclerotic CVDs (n=2), which did not meet inclusion/exclusion criteria for the main analysis, is further provided.

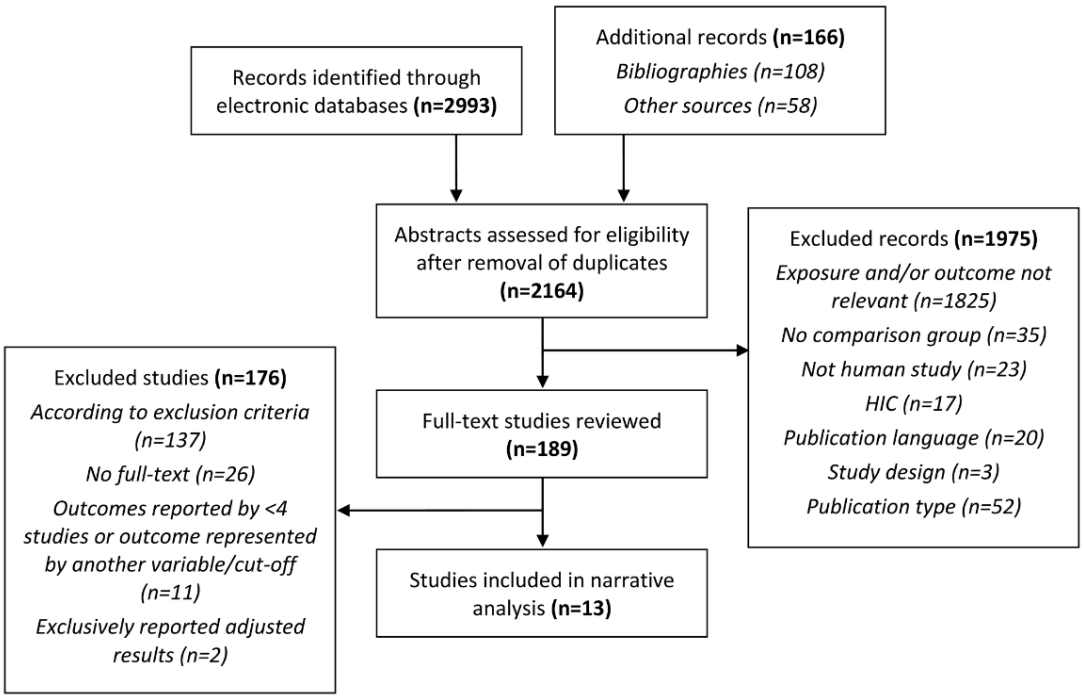


Fig 1 Flow diagram of the review process

CVD – Cardiovascular disease, HIC – high-income country, n – number of studies

Study characteristics

Characteristics of included studies are provided in Table 2 (see S4 Table for further details). We included one longitudinal[29] and 12 cross-sectional studies. The longitudinal study followed up participants 10 years from baseline. The cross-sectional studies comprised eight analyses of primary data[30-36]; two secondary analyses of existing cross-sectional data[37, 38]; and three cross-sectional analyses of longitudinal data,[39-41] of which two used data from different time points of one study.[40, 41] Study settings and characteristics differed substantially between studies. Most

studies were from Asia (India[31, 33, 35-39] and China[29, 40, 41]); one was from Latin America (Guatemala[32]) and two were from Sub-Saharan Africa (Ghana[30] and Nigeria[34]). Eight studies analysed data from rural populations and five studies additionally included urban participants (including migrants). Sample sizes ranged from 195 to 77,220 participants. Twelve studies reported the age range of participants; eight included younger adults aged from 15 years,[38, 42] 18 years,[30, 36, 39] and 20 years.[34, 35, 41] The maximum age of participants varied from 49 to 99 years. One study did not report the age range, however, described participants as adults.[32]

Table 1 Characteristics of included studies from five low- and middle-income countries (n=13)

Author and year	Population and site	Study design	Outcomes	Case definition	Exposure and comparators	n	Age, mean (SD)	Age range in years
Addo et al. 2006	Ghana, four rural farming communities	Cross-sectional	Hypertension	≥140/90 mmHg	Farmer	107	42.4 (18.6)	18, 99
					Trader	152		
					Other	103		
Arlappa et al. 2009	India, rural areas in nine states	Cross-sectional	Underweight	BMI <18.5 kg/m^2	Agriculture	399		60, 70+
					Non-agriculture	1,170		
Asgary et al. 2013	Jamkhed, India, six rural villages	Cross-sectional	Hypertension	≥140/90 mmHg	Farmer	112		40, 85
					Housekeeper	100		
Balagopal et al. 2012	Gujarat, India, rural community	Cross-sectional	Hypertension; underweight, overweight, obese; tobacco	SBP ≥140 mmHg; BMI <18.5, 23-24.99, ≥25 kg/m^2	Agrarian (low socio-economic status)	764	43.4 (15.9)	18+
					Business (high socio-economic status)	874	40.2 (15.7)	
Gregory et al. 2007	Guatemala, people born in four rural villages	Cross-sectional	Hypertension; overweight, obese; smoking	≥130/85 mmHg; BMI ≥25, ≥30 kg/m^2	Rural agriculture	88	31.7 (4.4)	
					Rural non-agriculture	153	31.4 (4.2)	
					Urban	119	33.6 (4.3)	
Hazarika et al. 2004	Assam, India, 25 rural villages	Cross-sectional	Hypertension	≥140/90 mmHg	Service			≥30
					Business			
					Cultivator			
					Daily wager			
					Unemployed			
					Others			
					Total	3,180		
He et al. 1991	Sichuan province, China, mountains, city and county seats	Cross-sectional	Age standardised hypertension I and II; smoking	140-159/90-94, ≥160/95 mmHg	Farmer	8,241	31.4	15, 89
					Migrant	2,575	33.1	
					Urban	3,689	33.9	
Norboo et al. 2015	Jammu and Kashmir, India, rural and urban areas	Cross-sectional	Hypertension; overweight	≥140/90 mmHg, BMI ≥25 kg/m^2	Farmer	1,247		20, 94
					Nomad	220		
					Sedentary worker	549		
					Other, including:	784		
					Housewife	*325*		
					Manual labourer	*63*		

Author and year	Population and site	Study design	Outcomes	Case definition	Exposure and comparators	n	Age, mean (SD)	Age range in years
					Monk	*157*		
					No job	*138*		
					Retired sedentary	*101*		
					Total	*2,800*	53.8 (15.0)	
Olugbile & Oyemade 1982	Nigeria, two rural areas in different states	Cross-sectional	Hypertension	≥140/90 mmHg	Agriculture company	112		20, 59
					Factory worker	136		
Subasinghe et al. 2014	Andhra Pradesh, India, 12 rural villages	Cross-sectional	Underweight	BMI <18 kg/m^2	Non-government, government	376		18, 55+
					Self-employed	165		
					Farming and livestock	326		
					Homemaker	209		
					Unemployed, student, retired	93		
Subramanian & Davey Smith 2006	India, rural and urban areas in 26 states	Cross-sectional	Underweight, overweight, obese	BMI <16; 16-16.9, 17-18.49, <18.5; 23-24.9, 25-29.9; ≥30 kg/m^2	Not working	48,160		15, 49
					Non-manual	4,433		
					Agricultural	17,758		
					Manual	6,869		
Wang et al. 2010	South-western China, mountains, city and county seats	Cross-sectional	Hypertension; overweight/obesity; smoking	≥130/85 mmHg; BMI ≥24 kg/m^2	Farmer	1,535	39.6	≥20
					Migrant	1,306	38.8	
					Urban	2,130	44.3	
Zhou et al. 2003	Beijing, Northern China, and Guangzhou, Southern China, rural areas near big cities	Cohort	Smoking		Agriculture 1983–84	326		35, 54 (at baseline)
					Remained in agriculture 1993–94			
					Agriculture 1983–84	102		
					Shifted out of agriculture 1993-94			
					Factory work 1983–84	135		
					Remained in factory work 1993-94			
					Office work 1983–84	70		
					Remained in office work 1993-94			

BMI - body mass index; HH - household(s); kg – kilograms; m - metre(s); mmHg - millimetre mercury; n – sample size; SBP - systolic blood pressure

Quality of the evidence

We identified substantial methodological shortcomings in all included studies (S5 Table). A particular concern was that none of the included studies described the measurements of exposure and comparators or outcomes in adequate detail to rate the quality of evidence relating to these domains. Studies rarely defined agriculture, but described participants in broad terms, such as farmers or agriculturalists. Two studies described participants' agricultural practices in more detail, such as whether they farmed their own, leased or someone else's land. The included 13 studies explored associations of engaging in agriculture with secondary outcomes, modifiable risk factors for atherosclerotic CVDs. Study outcomes were measured and categorised in various ways across the studies. None of the studies addressed blinding of outcome assessors although all assessed outcomes might be subjective and therefore open to bias from either assessors or participants.

Primary outcomes

Two studies reported on primary outcomes (S3 Table). One study from India classified participants as having coronary heart disease (CHD) if angina or infarction and CHD had previously been diagnosed; affirmative response was given to the Rose questionnaire; or changes were observed in electrocardiograms according to the Minnesota code classification system. The prevalence of CHD did not differ between agricultural, business, professional, government and household workers (n=3,148).[43] The second study, from Vietnam, combined mortality, identified by verbal autopsy, from undifferentiated CVDs, pulmonary heart disease, stroke and CHD during a two year period (n=49,543 person-years).[44] The rate of CVD was almost six times higher among non-pension retired individuals than among farmers in a subsample of older participants (≥ 50 years, n=15,193 person-years). The association more than halved when adjusting for age and gender.[44] Farmers did not differ from government employees and 'others'.

Secondary outcomes

Hypertension

Nine studies reported sufficient data for calculating PRs (95% CIs) of hypertension in agricultural and non-agricultural groups. Seven studies were set in rural areas and two included rural and urban residents. The prevalence of hypertension varied considerably between compared groups and studies, from 0.3% among farmers in one study[40] to 48.5% among retired sedentary workers in another.[35] However, different cut-offs were used (range: 130/85mmHg, ≥ 160/95mmHg), which hampered direct comparisons of results (Fig 2). In three publications (from two studies), hypertension was less prevalent among farmers than migrants[40, 41] and urban

workers,[32, 40, 41] with the exception of migrant women in one study.[40] The pattern was less clear in the studies that compared more specific occupation groups that combined men and women, who were predominantly living in rural areas. Farmers had a lower prevalence of hypertension than factory workers[34] and retired sedentary workers[35] in two studies. In two other studies, farmers had a higher prevalence of hypertension than manual labourers[35] and those with other employment than farming and trade.[30] Four studies did not observe differences in the prevalence of hypertension between agricultural and a range of non-agricultural groups comprising: non-farming rural workers,[32] housewives/housekeepers,[31, 35] daily wage earners,[33] service workers,[33] nomads,[35] sedentary workers,[35] monks,[35] the unemployed[33, 35] and a group of 'others' (who were not unemployed, business, service or daily wage workers).[33] Three comparisons of hypertension between agricultural workers and trade or business workers were contradictory: the prevalence was lower among agricultural workers than business workers in one study that analysed women and men separately[39] and similar between groups of agricultural and business/trade workers in analyses that combined women and men.[30, 33]

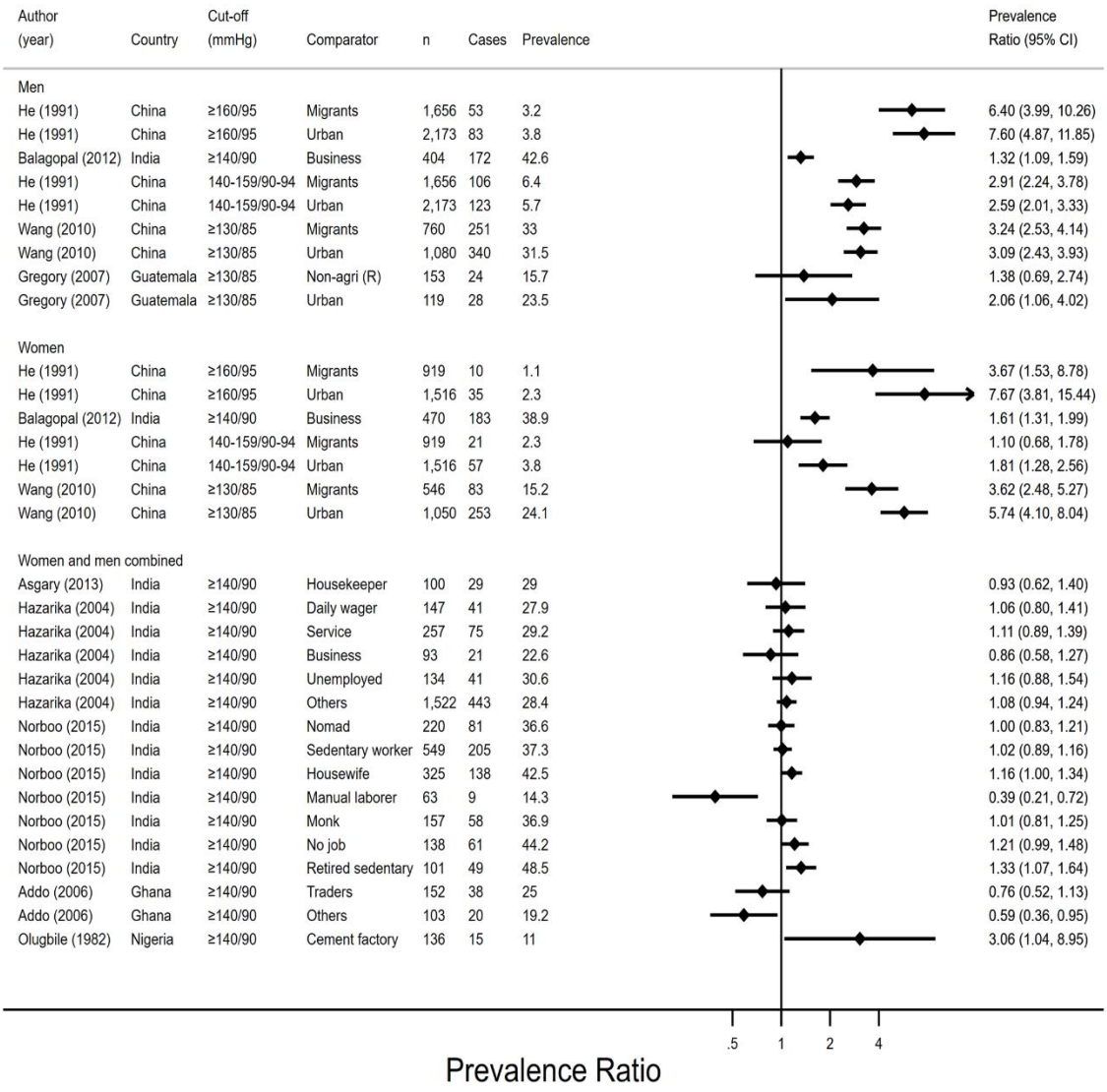


Fig 2 Prevalence ratios and 95% confidence intervals of hypertension by employment status (n=9)

Agri – agriculture, CI – confidence interval, n – sample size, PR – prevalence ratio, R – rural

Prevalence ratios were derived from comparing each non-agricultural group (coded 1) to the agricultural group (coded 0)

Hypertension remained considerably less common among farmers than migrants (urban people were excluded) after adjusting the odds ratios (ORs) for age in one study[40] (blood pressure ≥ 140-159/90-94 mmHg: OR^{men} 1.38 [95% CI 1.19, 1.59], OR^{women} 1.03 [95% CI 1.14, 2.93]; Blood pressure ≥ 160/95: OR^{men} 1.96 [95% CI 1.52, 2.52], OR^{women} 1.83 [95% CI 1.14, 2.93]). The adjusted OR was a slight attenuation among men and a slight increase among women compared the crude ORs.

Overweight and obesity

Five studies reported enough information to calculate PRs (95% CIs) for overweight and/or obesity. Studies predominantly combined overweight and obesity using international cut-offs (BMI ≥ 25 kg/m^2) (Fig 3). Data from all five studies suggested that overweight and obesity (analysed separately and combined) were less prevalent among farmers than migrant,[41] urban,[32, 41] business,[39] non-manual,[38] sedentary[35] and retired sedentary workers[35]; nomads,[35] housewives,[35] monks,[35] and the unemployed.[35, 38] Overweight (BMI 25-29.9 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) were less prevalent among women engaging in agriculture than manual labour in one study,[38] whereas overweight and obesity combined did not differ between agricultural and manual workers in another study combining men and women.[35] Overweight and obesity (BMI ≥ 25 kg/m^2), but not obesity separately (BMI ≥ 30 kg/m^2), was less prevalent among agricultural than non-agricultural rural men in one study.[32] However, large confidence intervals of the association suggested low precision of the estimates.

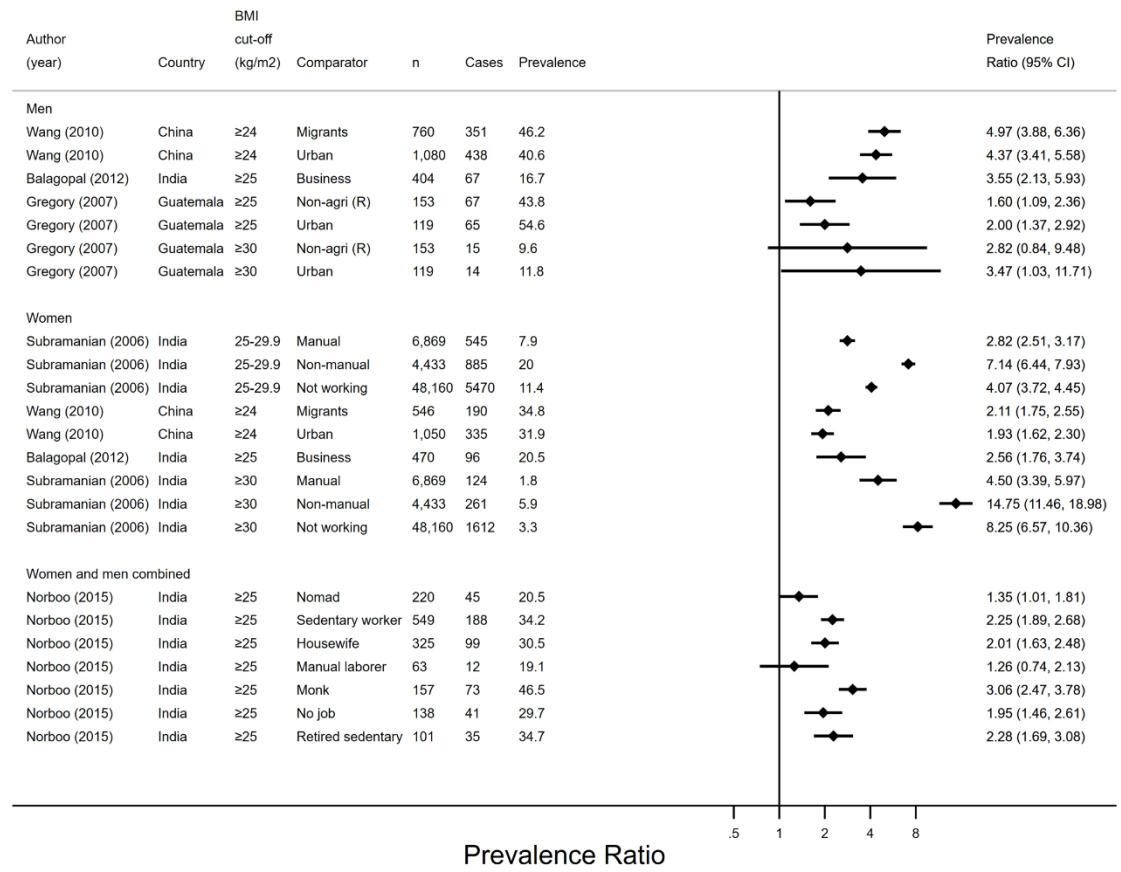


Fig 3 Prevalence ratios and 95% confidence intervals of overweight and obesity by employment status (n=5)

Agri – agriculture, BMI – body mass index, CI – confidence interval, n – sample size, PR – prevalence ratio; R – rural

Prevalence ratios were derived from comparing each non-agricultural group (coded 1) to the agricultural group (coded 0)

Underweight

Four studies presented sufficient data to calculate PRs (95% CIs) for underweight (BMI < 18.5kg/m^2) (Fig 4). Three studies were set in rural areas and one included rural and urban workers. All studies suggested that farmers had higher prevalence of underweight than homemakers,[36] manual and non-manual women[38]; business,[39] government and non-government[36] women and men; and 'other' (than non-agricultural) workers in analyses combining genders.[37] The prevalence of underweight was higher among farming than among self-employed men but not women.[36] Underweight was more prevalent among farmers than among unemployed women,[38] but similar to students, retired and unemployed people (the latter two were combined in gender-specific analyses).[36] There were no differences in the prevalence of underweight between farmers and other (non-specific) non-agricultural workers.[37]

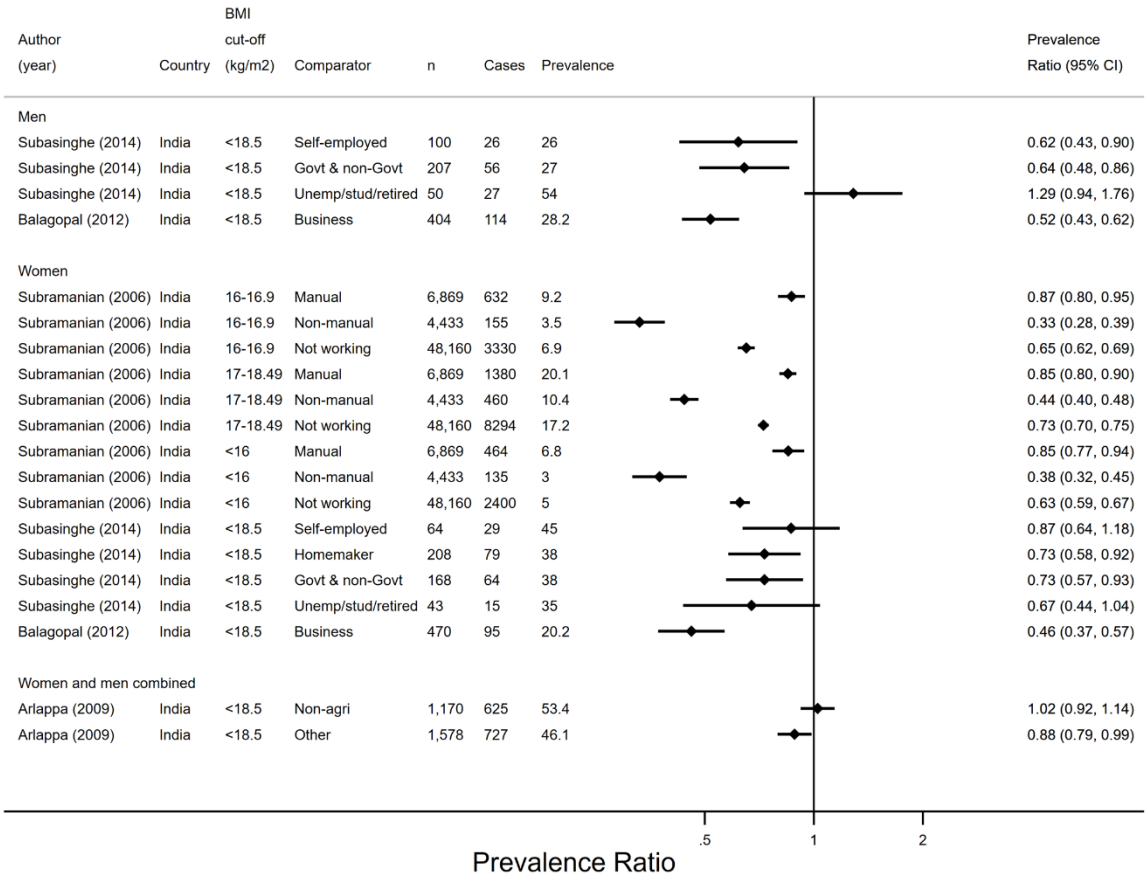


Fig 4 Prevalence ratios and 95% confidence intervals of underweight by employment status (n=4)

Agri – agriculture, BMI – body mass index, CI – confidence interval, Govt – government, n – sample size, PR – prevalence ratio, stud – student

Prevalence ratios were derived from comparing each non-agricultural group (coded 1) to the agricultural group (coded 0)

Tobacco

Five studies presented sufficient data to calculate PRs (95% CIs) for current smoking (n=2), ever smoked (n=1), unspecified period of smoking (n=1) and chewing, snuffing and smoking tobacco (n=1) (Fig 5). Two studies were set in rural areas and three were set in rural and urban areas. In one longitudinal study,[29] the prevalence of smoking decreased over 10 years among men who transitioned out of agriculture and into to unspecified occupations. The prevalence of smoking did not change among men who remained in agriculture, factory or office work over the 10 years.[29] Smoking prevalence did not differ between the four groups at the 10-year follow-up (Fig 5). Farmers were more likely to smoke than urban workers in three publications (from two studies).[32, 40, 41] Two publications used data from different time-points of the same study and found that migrant men were more likely to smoke than farming men in 1991[40] and less likely to smoke than farming men in 2010[41]. Although PRs for migrant versus farming women were similar to those of men in the two studies, they had wide 95% CIs that included the null-value of one. Two studies did not support differences between agricultural workers and non-agricultural rural workers[32] or business workers.[39]

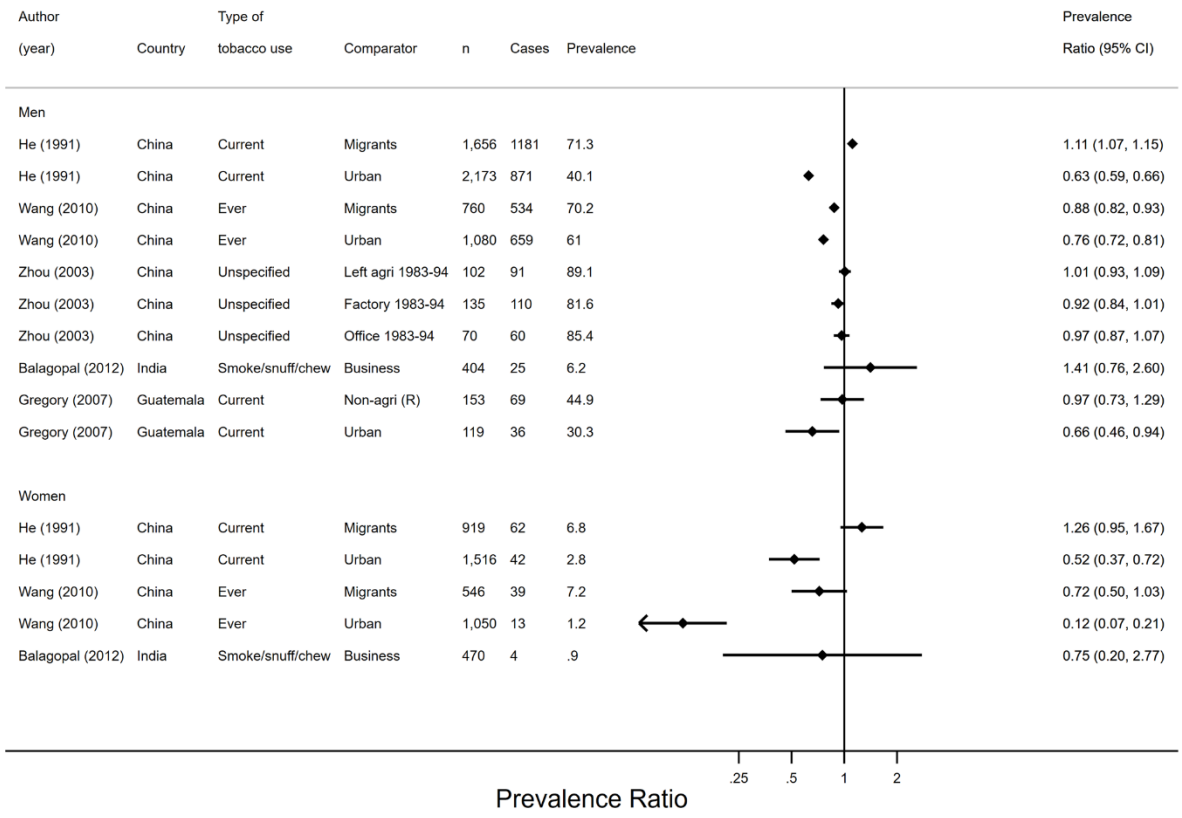


Fig 5 Prevalence ratios and 95% confidence intervals of tobacco use by employment status (n=5)

Agri– agriculture, CI – confidence interval, n – sample size, PR – prevalence ratio, R – rural

Prevalence ratios were derived from comparing each non-agricultural group (coded 1) to the agricultural group (coded 0)

Discussion

We aimed systematically to review the published evidence on the association of engaging in agriculture compared to other types of non-agricultural employment with atherosclerotic CVDs and over 50 associated modifiable risk factors. Following the protocol, we found two studies on the primary outcomes, atherosclerotic CVDs and 13 studies on four secondary outcomes, hypertension, overweight and obesity, underweight and tobacco use. Included studies were predominantly from India (n=7) and China (n=3). Heterogeneity in study settings, populations under study, measurement methods and categorisation of exposures and outcomes prevented formal meta-analyses with pooled results and generation of funnel plots.

Older agricultural workers had a lower CVD mortality rate than retired individuals in a cohort of rural residents from Vietnam, whereas the prevalence of CHD did not differ with employment in a study from rural India. Five out of nine studies suggested that people who engaged in agriculture had a lower prevalence of hypertension than migrant, urban, factory and retired sedentary workers. The evidence suggested no difference or contradictory results on the prevalence of hypertension for a number of other occupations with no clear pattern. One study reported that migrants had a higher likelihood of hypertension than farmers after adjusting for age and gender. Most evidence suggested that people who engaged in agriculture were less likely to be overweight and obese and more likely to be underweight than most non-agricultural workers they were compared to. Urban men and women appeared less likely to smoke than farmers in two studies, whereas results were contradictory for migrants. The prevalence of smoking declined among men transitioning out of agriculture and remained unchanged among men who did not change their occupation during 10 years. However, there were no differences between groups at the 10-year follow-up. The associations of type of labour could not be separated from those of location of residence for any of the outcomes because of the way different types of employment were sampled, for example, rural agricultural workers and urban government workers.

To the best of our knowledge, this is the first systematic review to assess associations of engaging in agriculture compared to types of non-agricultural employment with CVD incidence, prevalence and risk factors. A main contribution of this review was the calculation and comparison of PRs and 95% CIs across studies with available data, including studies that did not perform statistical tests. We attempted to reduce overestimation of associations from studies with high outcome prevalence by calculating PRs (95% CIs), as ORs are prone to overestimate the strength of associations in this context.[45] There are also a number of limitations to this review. A main limitation was the lack of eligible studies, particularly on the primary outcomes, atherosclerotic CVDs. The lack of studies on primary outcomes was likely due to a lack of this

type of data from LMIC, e.g. from national surveys and disease surveillance.[3] It is also possible that we missed some relevant studies. The lack of studies on secondary outcomes were more surprising and is likely rooted in a disconnect between sectors and sciences concerned with employment/labour and chronic diseases that are inherently interlinked.[11, 13, 46] A number of studies and risk factors were excluded from the review as a result of restricting analyses to risk factors that were reported in four or more studies and for which clear descriptions of measurement methods and outcome categorisations were provided. Our list of risk factors was comprehensive, but not exhaustive and we may in turn have neglected some important relationships of agriculture and CVD risk factors, e.g. mental health.

Substantial limitations of the included studies made it difficult to draw conclusions. The absence of definitions and descriptions of the measurement methods and categorisations of exposure and comparators complicated the interpretation of results. It may be challenging to measure livelihoods and types of employment in LMIC settings where the informal sector dominates and livelihood diversification and seasonal migration are common.[47] It is possible that the 'current status' in cross-sectional studies reflect the work done on the day or season of the survey, at least for some employment groups, e.g. daily wage earners. In turn, some misclassification of results is expected to dilute our results (towards the null). In all studies, participants may have belonged to their current exposure or comparator group for varying lengths of time. Survivor bias arising from ill-health or death from the reviewed risk factors or related CVD events in the time of employment prior to study start may explain some of this review's inconsistent and contradictory results.[48] To varying degree, healthy worker bias may have diluted and possibly reversed associations, particularly in studies of older adults or other high-risk groups.[48, 49] However, studies of younger adults in LMICs may also be at risk of this kind of bias, as premature deaths and disability from chronic diseases become more common.[3] Some employment groups, e.g. highly physical work, may attract healthier individuals. For example, it is common for young adults in LMIC to migrate from rural to urban areas for work in e.g. construction (men) or housekeeping (women),[47] leaving older, potentially less healthy, adults behind in agriculture.[50] None of the included studies adequately addressed differences between included and excluded participants. Large differences in time between included studies (1991 to 2015) further pose challenges to directly comparing results within and between studies, as well as inferring relevance of the observed associations in the present contexts of LMICs.

We could not determine the extent to which potential confounders might have accounted for observed associations (or absence of associations) because of the lack of appropriately adjusted analyses. An important example is socio-economic status, which particularly may mask results of studies that included or compared rural and urban workers. For example, wealthier urban residents may enjoy health benefits from availability of and access to goods and services, such

as food, water, sanitation and health-services. In contrast, the health of the urban poor may be worse than that of the rural poor,[18] for example as a result of unsafe living conditions and high living costs.[47] This may explain some of the inconsistent findings, e.g. among manual workers, business/trade and the unemployed, in studies sourcing participants from different settings. We appreciate that hypothesis-generating studies, that analysed multiple exposure-outcome relationships may have limited their number of analyses to reduce the risk of producing 'statistically significant' results by chance. Few studies reported that they based sample sizes on power calculation and wide confidence intervals in several studies suggested that the statistics power might not be high enough to detect existing differences. Missing data may have additionally favoured or diluted associations in several studies. Finally, there were some indications of selective reporting of results from most studies. To some extent, this could be explained by the large scope of several exploratory studies, which may not allow for reporting all analyses. It is also possible that some studies gave preference to 'statistically significant' results in order to improve chances of publication.[26]

It is common for LMIC governments, e.g. in Indian and China, to facilitate labour-force shifts out of agriculture and into industry and service sectors to promote economic growth during development.[51, 52] Concurrently, national age-standardised prevalence of death and disability from CVDs have risen and the Employment Conditions Knowledge Network expresses concern that LMICs will be unable to provide the growing urban labour-force with fair employment opportunities.[11, 53] Progress reports on China's recent urbanisation have shown unexpected challenges with integrating agricultural workers into new types of employment and urban settings. Chinese migrant workers, for example, suffered social exclusion and were denied equal rights to fair employment, housing and health services, which is associated with infectious and chronic diseases.[51, 54] As previously discussed, our search only returned one study that examined the association of employment transitions out of agriculture with cardiovascular health in LMICs. The unexpected developments in China warrant a deeper understanding of how employment transitions are associated with CVDs and risk factors to ensure current economic growth strategies do not add to already expected rises in CVD burdens with development in LMICs.

Evidence on associations of types of employment, and particularly employment transitions, with chronic diseases and related risk factors is fragmented. Where the health sciences typically assess an association of an intervention with an outcome within a single type of employment (e.g. agriculture [55, 56]) the employment sector tends to focus on shorter-term issues relating to health and safety regulation-outcomes (e.g. ergonomics, injuries or exposure to noise or hazardous agents).[11] The data collected across sectors vary widely, for example in relation to sources and collection methods, and are not easily combined for epidemiological studies that

can assess health risks or benefits that are associated with types of employment. Prospective longitudinal studies should take advantage of the current momentum for interdisciplinary initiatives across the 2030 agenda for sustainable development, such as 'decent work and economic development', 'sustainable cities and communities' and 'health and well-being for all' in LMICs.[57] New study models and data collection approaches are likely to be required to address these challenging needs. Studies should also address the methodological shortcomings identified in this review by complying with appropriate reporting guidelines,[58] namely ensuring appropriate selection of participants that accommodates conditional comparability of outcomes across exposure and comparator groups; use of appropriate measurement methods for collection of all data; and appropriate reporting of study protocol, methods and results (including the reporting of missing data and 'non-significant' findings).

Conclusion and implications

There was some, however limited, evidence of negative associations of engaging in (rural) agriculture compared to types (usually urban) employment with prevalence of hypertension, overweight and obesity, and positive associations with prevalence of underweight and smoking. There were no clear patterns indicating which types of employment were associated with higher or lower prevalence of outcomes. High quality evidence is lacking on how engaging in agriculture compared to types of non-agricultural employment may act on CVDs and risk factors in LMICs. Rigorous studies that address the methodological shortcomings identified by this review are needed. They should cultivate a consensus of how best to measure and categorise employment when investigating chronic disease outcomes as well as appropriately control for potential confounders. We further call for new models of interdisciplinary longitudinal studies, for example across health sciences, livelihoods, demography and policy, that assess the association of livelihood transitions and migration with core health outcomes in LMICs. These sorts of studies should seek to inform evidence-based chronic disease prevention measures in national economic growth strategies to safeguard health during development in line with the sustainable development goals.

Acknowledgements

We would like to thank Ms Katharine Ker (Faculty of Epidemiology Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine (LSHTM), United Kingdom (UK)) for guidance on the search strategy, data extraction form and risk of bias assessment strategy; Dr MinHae Park (Faculty of Epidemiology Faculty of Epidemiology and Population Health, LSHTM, UK) for contributing to the data extraction form and risk of bias assessment strategy; Ms Jane Falconer, User Support Services Librarian (Library and Archives

Service, LSHTM, UK) for developing the search on ‘low- and middle-income countries’; Dr Clara Calvert (Faculty of Epidemiology and Population Health, LSHTM, UK) for advising on analysis strategy and manuscript preparation; and MD Karin Østergaard (Department of Internal Medicine, Hospital of South West Jutland, Denmark) for manuscript revisions.

References

1. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease Study and the Preventable Burden of NCD. Global heart. 2016 Dec;11(4):393-7
2. World Health Organization. Noncommunicable diseases: Fact Sheet 2015. Available from: http://www.who.int/mediacentre/factsheets/fs355/en/.
3. Roth GA, Huffman MD, Moran AE, Feigin V, Mensah GA, Naghavi M, et al. Global and regional patterns in cardiovascular mortality from 1990 to 2013. Circulation. 2015 Oct 27;132(17):1667-78
4. Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, Brauer M, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015 Dec 05;386(10010):2287-323
5. World Health Organization. Global atlas on cardiovascular disease prevention and control. Geneva: World Health Organization, 2011.
6. National cholesterol education program, National heart lung and blood institute, National institutes of health. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). NIH Publication: 2002.
7. Reddy SK, Katan MB. Diet, nutrition and the prevention of hypertension and cardiovascular diseases. Public Health Nutr. 2004 Feb;7(1a):167-86
8. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
9. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March. Report No.
10. World Bank Group. Employment in agriculture (% of total employment) (modelled ILO estimate): World Bank; 2018 [updated 2018; cited 2018 18 June]. Available from: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=XO.
11. Employment Conditions Knowledge Network. Employment Conditions and Health Inequalities: Final Report to the WHO. 2007.
12. World Health Organization. Social determinants of health: Employment conditions 2017 [updated 2017; cited 2017 21 September]. Available from: http://www.who.int/social_determinants/themes/employmentconditions/en/.
13. Hawkes C, Ruel M. The links between agriculture and health: an intersectoral opportunity to improve the health and livelihoods of the poor. Bulletin of the World Health Organization. 2006 Dec;84(12):984-90
14. Girard AW, Self JL, McAuliffe C, Olude O. The effects of household food production strategies on the health and nutrition outcomes of women and young children: a systematic review. Paediatr Perinat Epidemiol. 2012 Jul;26 Suppl 1:205-22
15. Atkinson K, Lowe S, Moore S. Human development, occupational structure and physical inactivity among 47 low and middle income countries. Preventive Medicine Reports. 2016 June 01;3:40-5
16. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutr Rev. 2012 Jan;70(1):3-21
17. Reddy KS. Cardiovascular diseases in the developing countries: dimensions, determinants, dynamics and directions for public health action. Public Health Nutr. 2002 Feb;5(1a):231-7
18. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009 Aug 18;151(4):264-9, w64
20. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. MOOSE Guidelines for Meta-Analyses and Systematic Reviews of Observational Studies. 2008.
21. Sørensen TB. Is working in agriculture associated with cardiovascular disease risk? A systematic review.2015. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015025488.
22. World Bank. World Bank Country and Lending Groups 2017. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

23. Reddy KS, Katan MB. Diet, nutrition and the prevention of hypertension and cardiovascular diseases. Public Health Nutr. 2004 Feb;7(1a):167-86
24. World Health Organization. Global status report on non communicable diseases 2014. Geneva, Switzerland: WHO, 2014.
25. Bazzano LA. Dietary intake of fruit and vegetables and risk of diabetes mellitus and cardiovascular diseases. Background paper for the Joint FAO/WHO workshop on fruit and vegetables for health, 1-3 September 2004, Kobe, Japan. World Health Organization, 2004.
26. Sterne JAC, Higgins JPT, Reeves BC, on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI). 24 September 2014.
27. Society for Biomedical Diabetes Research. Conversion table for chemical compounds from conventional to SI units 2002-2016 [cited 2016 5 July]. Available from: http://www.soc-bdr.org/rds/authors/unit_tables_conversions_and_genetic_dictionaries/e5196/index_en.html#conversion.
28. Sterne JAC. Meta-Analysis in Stata: An updated Collection from the Stata Journal. 1 ed. Newton J, Cox J, editors. Texas, USA: Stata Press 2009 2009.
29. Zhou B, Li Y, Stamler J, Tao S, Davis CE, Wu Y, et al. Relation of occupational change to cardiovascular risk factor levels in rural Chinese men: the People's Republic of China-United States Collaborative Study on Cardiovascular and Cardiopulmonary Epidemiology. Am J Public Health. 2003 Dec;93(12):2049-51
30. Addo J, Amoah AG, Koram KA. The changing patterns of hypertension in Ghana: a study of four rural communities in the Ga District. Ethnicity & disease. 2006 Autumn;16(4):894-9
31. Asgary R, Galson S, Shankar H, O'Brien C, Arole S. Hypertension, pre-hypertension, and associated risk factors in a subsistent farmer community in remote rural central India. Journal of Public Health (Germany). 2013 June;21(3):251-8
32. Gregory CO, Dai J, Ramirez-Zea M, Stein AD. Occupation is more important than rural or urban residence in explaining the prevalence of metabolic and cardiovascular disease risk in Guatemalan adults. The Journal of nutrition. 2007 May;137(5):1314-9
33. Hazarika NC, Biswas D, Narain K, Kalita HC, Mahanta J. Hypertension and its risk factors in tea garden workers of Assam. National Medical Journal of India. 2004;15(2):63-8
34. Olugbile AO, Oyemade A. Health and the environment-a comparative study of agricultural and industrial workers in Nigeria. Afr J Med Med Sci. 1982 Sep-Dec;10(3-4):107-12
35. Norboo T, Stobdan T, Tsering N, Angchuk N, Tsering P, Ahmed I, et al. Prevalence of hypertension at high altitude: Cross-sectional survey in Ladakh, Northern India 2007-2011. BMJ Open. 2015;5 (4)(e007026)
36. Subasinghe AK, Walker KZ, Evans RG, Srikanth V, Arabshahi S, Kartik K, et al. Association between farming and chronic energy deficiency in rural South India. PLoS ONE [Electronic Resource]. 2014;9(1):e87423
37. Arlappa N, Venkaiah K, Rao KM, Reddy Ch G, Kumar SS, Ravindranath M, et al. Prevalence of chronic energy deficiency in rural-dwelling older Indian adults during a period of severe drought. Journal of Nutrition for the Elderly. 2009 Jul;28(3):301-12
38. Subramanian SV, Smith GD. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. Am J Clin Nutr. 2006 Sep;84(3):633-40
39. Balagopal P, Kamalamma N, Patel TG, Misra R. A community-based participatory diabetes prevention and management intervention in rural India using community health workers. Diabetes Educ. 2012 Nov-Dec;38(6):822-34
40. He J, Klag MJ, Whelton PK, Chen JY, Mo JP, Qian MC, et al. Migration, blood pressure pattern, and hypertension: the Yi Migrant Study. American journal of epidemiology. 1991 Nov 15;134(10):1085-101
41. Wang C, Wei D, Wang B, Zhang J, Zhang K, Ma M, et al. Effect of lifestyle on the prevalence of the metabolic syndrome among farmers, migrants with Yi ethnicity and Han population in Sichuan province of China. Asia Pacific journal of clinical nutrition. 2010;19(2):266-73
42. He J, Tell GS, Tang YC, Mo PS, He GQ. Effect of migration on blood pressure: the Yi People Study. Epidemiology. 1991 Mar;2(2):88-97
43. Gupta R, Prakash H, Gupta VP, Gupta KD. Prevalence and determinants of coronary heart disease in a rural population of India. Journal of clinical epidemiology. 1997 Feb;50(2):203-9
44. Van Minh H, Byass P, Wall S. Mortality from cardiovascular diseases in Bavi District, Vietnam. Scandinavian journal of public health Supplement. 2003;62:26-31
45. Tamhane AR, Westfall AO, Burkholder GA, Cutter GR. Prevalence odds ratio versus prevalence ratio: choice comes with consequences. Statistics in medicine. 2016 Dec 30;35(30):5730-5

46. Kadiyala S, Harris J, Headey D, Yosef S, Gillespie S. Agriculture and nutrition in India: mapping evidence to pathways. Annals of the New York Academy of Sciences. 2014 Dec;1331:43-56
47. Tacoli C. Urbainzation, gender and poverty. New York, USA: United Nations Population Fund, International Institute for Environment and Development, 2012 March. Report No.
48. Delgado-Rodríguez M, Llorca J. Bias. Journal of Epidemiology and Community Health. 2004;58(8):635-41
49. Buckley JP, Keil AP, McGrath LJ, Edwards JK. Evolving methods for inference in the presence of healthy worker survivor bias. Epidemiology. 2015 Mar;26(2):204-12
50. Bhandari PB. Rural livelihood change? Household capital, community resources and livelihood transition. Journal of rural studies. 2013 Oct 1;32:126-36
51. Chen M, Gong Y, Lu D, Ye C. Build a people-oriented urbanization: China's new-type urbanization dream and Anhui model. Land Use Policy. 2019 2019/01/01/;80:1-9
52. The National Institution for Transforming India. The Three Year Action Agenda 2017-18 to 2019-20. Delhi, India: Government of India, 2017.
53. Global Burden of Disease Study (GBD). GBD Foresight Viz Hub 2019 [cited 2019 30 September]. Available from: https://vizhub.healthdata.org/gbd-foresight/.
54. Chen M, Liu W, Lu D. Challenges and the way forward in China's new-type urbanization. Land Use Policy. 2016 2016/09/01/;55:334-9
55. Masset E, Haddad L, Cornelius A, Isaza-Castro J. Effectiveness of agricultural interventions that aim to improve nutritional status of children: systematic review. Bmj. 2012;344:d8222
56. Berti PR, Krasevec J, FitzGerald S. A review of the effectiveness of agriculture interventions in improving nutrition outcomes. Public Health Nutr. 2004 Aug;7(5):599-609
57. United Nations. Sustainable Development goals 2020. Available from: https://www.un.org/sustainabledevelopment/sustainable-development-goals/.
58. Equator Network. Enhancing the QUAlity and Transparency Of health Research NA. Available from: https://www.equator-network.org/reporting-guidelines.

Supporting information

S1 Table. PRISMA checklist.

S2 Table. Search strategy

S3 Table. Characteristics of studies excluded from results synthesis (n=11)

S4 Table. Detailed characteristics of included studies (n=13)

S5 Table. Quality of evidence rating of included studies from five countries (n=13)

S1 Table. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1, 2, 6
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6, 9
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow–up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6-8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supporting information: S2 Table Search strategy
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta–analysis).	9-10
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	9-10

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	10
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta–analysis.	9-10
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre–specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	10-11
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	10-13
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	13, S5 Table Quality of evidence rating
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	13-16
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	13, S5 Table Quality of evidence rating
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			

Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	16-17
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	17-19
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	20
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Financial Disclosure Statement

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

S2 Table. Search strategy

Search number (#) and description
#1 Search (animal husbandry/ OR aquaculture/ OR beekeeping/ OR dairying/ OR gardening/ OR organic agriculture/[MeSH Terms])
#2 Search (agri*[Title/Abstract] OR agrarian[Title/Abstract] OR agro*[Title/Abstract] OR farm*[Title/Abstract] OR fishing[Title/Abstract] OR aquaculture[Title/Abstract] OR fisherm#n[Title/Abstract] OR mariculture[Title/Abstract] OR pisciculture[Title/Abstract] OR pastoral[Title/Abstract] OR livestock[Title/Abstract] OR dairying[Title/Abstract] OR floriculture*[Title/Abstract] OR horticulture*[Title/Abstract] OR forestry[Title/Abstract])
#3 Search ("Animal husbandry"[Title/Abstract] OR "animal rearing"[Title/Abstract] OR "livestock rearing"[Title/Abstract] OR "crop production"[Title/Abstract] OR "food production"[Title/Abstract] OR "animal production"[Title/Abstract] OR "poultry production"[Title/Abstract] OR "swine production"[Title/Abstract] OR "livestock production[Title/Abstract] OR "poultry farm*"[Title/Abstract] OR "swine farm*"[Title/Abstract])
#4 Search (land-use[Title/Abstract] OR land-own*[Title/Abstract] OR pasture[Title/Abstract])
#5 Search ("Arable land"[Title/Abstract] OR "agricultur* land"[Title/Abstract] OR "farm land"[Title/Abstract] OR "own* land"[Title/Abstract] OR "land use*"[Title/Abstract])
#6 Search ((((#1) OR #2) OR #3) OR #4) OR #
#7 Search (cardiovascular diseases/ OR myocardial ischemia/ OR pulmonary heart disease/ OR brain ischemia/ OR carotid artery thrombosis/ OR stroke, lacunar/ OR cerebral arterial diseases/ OR intracranial arteriosclerosis/ OR intracranial embolism and thrombosis/ OR intracranial thrombosis/ OR stroke/ OR embolism and thrombosis/ OR hypertension/ OR hypotension/ OR peripheral vascular diseases/ OR prehypertension/ OR poisoning/ OR substance-related disorders/[MeSH Terms])
#8 Search (Stroke [Title/Abstract] OR "coronary heart disease" [Title/Abstract] OR CHD[Title/Abstract] OR "cardio-vascular"[Title/Abstract] OR CVD[Title/Abstract] OR "heart disease?"[Title/Abstract] OR "cardiometabolic"[Title/Abstract]
#9 Search (exp body fat distribution/[Title/Abstract] OR body mass index/[Title/Abstract] OR waist circumference/[Title/Abstract] OR waist-height ratio/[Title/Abstract] OR cholesterol/[Title/Abstract] OR apolipoproteins/[Title/Abstract] OR exp apolipoproteins a/[Title/Abstract] OR exp apolipoproteins b/OR Triglycerides/[Title/Abstract] OR exp Hypertriglyceridemia/[Title/Abstract] OR diabetes mellitus, type 2/[Title/Abstract] OR diabetes, gestational/[Title/Abstract] OR diabetic ketoacidosis/[Title/Abstract] OR prediabet* state/[Title/Abstract] OR hyperglycemia/[Title/Abstract] OR glucose intolerance/[Title/Abstract] OR insulin resistance/[Title/Abstract] OR metabolic syndrome x/[Title/Abstract] OR Hyperinsulinism/[Title/Abstract] OR Vascular Stiffness/[Title/Abstract] OR C-Reactive Protein/[Title/Abstract] OR Carotid Intima-Media Thickness/[Title/Abstract] OR Homocysteine/[Title/Abstract] OR exp *Dyslipidemias/[Title/Abstract] OR *Alcohols/[Title/Abstract] OR dietary fats/[Title/Abstract] OR cholesterol, dietary/[Title/Abstract] OR *fatty acids/[Title/Abstract] OR Trans Fatty Acids/[Title/Abstract] OR Dietary Carbohydrates/[Title/Abstract] OR Antioxidants/[Title/Abstract] OR Vitamin E/[Title/Abstract] OR Ascorbic Acid/[Title/Abstract] OR Ubiquinone/[Title/Abstract] OR exp Flavonoids/[Title/Abstract] OR exp carotenoids/[Title/Abstract] OR beta carotene/[Title/Abstract] OR Selenium/[Title/Abstract] OR Folic Acid/[Title/Abstract] OR Vitamin B Complex/[Title/Abstract] OR exp Fatty Acids, Omega-3/[Title/Abstract] OR sodium chloride, dietary/OR Dietary Fibre/[Title/Abstract] OR exp Fruit/[Title/Abstract] OR Vegetables/[Title/Abstract] OR Exercise/[Title/Abstract])
#10 Search ("Body mass index"[Title/Abstract] OR BMI[Title/Abstract] OR "waist circumference"[Title/Abstract] OR "hip circumference"[Title/Abstract] OR "waist-to-hip ratio"[Title/Abstract] OR WHR[Title/Abstract] OR "body fat"[Title/Abstract] OR "fat percentage"[Title/Abstract] OR "blood pressure"[Title/Abstract] OR hypertension[Title/Abstract] OR cholesterol[Title/Abstract] OR HDL[Title/Abstract] OR LDL[Title/Abstract] OR Apolipoprotein*[Title/Abstract] OR triglycerides[Title/Abstract] OR hypertriglycerid?emia[Title/Abstract] OR diabetes[Title/Abstract] OR DMTII[Title/Abstract] OR DMT2[Title/Abstract] OR TIIDM[Title/Abstract] OR T2DM[Title/Abstract] OR hyperglyc?emia[Title/Abstract] OR "impaired fasting glucose"[Title/Abstract] OR "insulin resistance"[Title/Abstract] OR hyperinsulin*[Title/Abstract] OR "Homeostatic Model Assessment"[Title/Abstract] OR HOMA[Title/Abstract] OR "augmentation index"[Title/Abstract] OR AIX[Title/Abstract] OR "arterial stiffens"[Title/Abstract] OR "C-reactive protein"[Title/Abstract] OR CRP[Title/Abstract] OR "hs-CRP"[Title/Abstract] OR "carotid intima-media thickness"[Title/Abstract] OR CIMT[Title/Abstract] OR IMT[Title/Abstract] OR homocysteine[Title/Abstract] OR Framingham[Title/Abstract] OR dyslipid?emia[Title/Abstract] OR "metabolic syndrome"[Title/Abstract] OR "syndrome x"[Title/Abstract])

#11 Search (alcohol*[Title/Abstract] OR beer[Title/Abstract] OR spirit[Title/Abstract] OR wine[Title/Abstract] OR fat[Title/Abstract] OR “trans-fat”[Title/Abstract] OR carbohydrate[Title/Abstract] OR sugar[Title/Abstract] OR antioxidant*[Title/Abstract] OR vitamin E[Title/Abstract] OR Vitamin C[Title/Abstract] OR Ascorbic acid[Title/Abstract] OR ubiquinone[Title/Abstract] OR coenzyme Q[Title/Abstract] OR bioflavonoids[Title/Abstract] OR “beta-carotene”[Title/Abstract] OR selenium[Title/Abstract] OR “folic acid”[Title/Abstract] OR “vitamin B6”[Title/Abstract] OR “vitamin B12”[Title/Abstract] OR “omega-3”[Title/Abstract] OR “n-3 fat*”[Title/Abstract] OR salt[Title/Abstract] OR sodium[Title/Abstract] OR fibre[Title/Abstract] OR fruit?[Title/Abstract] OR vegetable?[Title/Abstract] OR diet[Title/Abstract] OR “physical activity”[Title/Abstract] OR PAL[Title/Abstract] OR “physically active”[Title/Abstract] OR sedentary[Title/Abstract] OR inactiv*[Title/Abstract] OR “energy expenditure”[Title/Abstract] OR “metabolic equivalent?”[Title/Abstract] OR MET?[Title/Abstract] OR “physical activity ratio”[Title/Abstract] OR PAR?[Title/Abstract])

#12 Search (β-carotene[Title/Abstract] OR ω-3 fat*[Title/Abstract] OR omega?3 fat*[Title/Abstract])

#13 Search (((((#7) OR #8) OR #9) OR #10) OR #11) OR #12

#14 Search Developing Countries/[MeSH Terms]

#15 Search (“Low-and middle income countr*”[Title/Abstract] OR lmic*[Title/Abstract] OR “third world”[Title/Abstract] OR “lami countr*”[Title/Abstract] OR “transitional countr*”[Title/Abstract] OR “developing econom*”[Title/Abstract] OR “less* developed econom*”[Title/Abstract] OR “under-developed econom*”[Title/Abstract] OR “underdeveloped econom*”[Title/Abstract] OR “middle income econom*”[Title/Abstract] OR “low* income econom*”[Title/Abstract] OR “developing countr*”[Title/Abstract] OR “less* developed countr*”[Title/Abstract] OR “under-developed countr*”[Title/Abstract] OR “underdeveloped countr*”[Title/Abstract] OR “middle income countr*”[Title/Abstract] OR “low* income countr*”[Title/Abstract] OR “underserved countr*”[Title/Abstract] OR “under served countr*”[Title/Abstract] OR “deprived countr*”[Title/Abstract] OR “poor* countr*”[Title/Abstract] OR “developing nation?”[Title/Abstract] OR “less* developed nation?”[Title/Abstract] OR “under-developed nation?”[Title/Abstract] OR “underdeveloped nation?”[Title/Abstract] OR “middle income nation?”[Title/Abstract] OR “low* income nation?”[Title/Abstract] OR “underserved nation?”[Title/Abstract] OR “under served nation?”[Title/Abstract] OR “deprived nation?”[Title/Abstract] OR “poor* nation?”[Title/Abstract] OR “developing population?”[Title/Abstract] OR “less* developed population?”[Title/Abstract] OR “under-developed population?”[Title/Abstract] OR “underdeveloped population?”[Title/Abstract] OR “middle income population?”[Title/Abstract] OR “low* income population?”[Title/Abstract] OR “underserved population?”[Title/Abstract] OR “under served population?”[Title/Abstract] OR “deprived population” [Title/Abstract] OR “poor* population?”[Title/Abstract] OR “developing world”[Title/Abstract] OR “low* gdp”[Title/Abstract] OR “low* gnp”[Title/Abstract] OR “low* gross domestic”[Title/Abstract] OR “low* gross national”[Title/Abstract] OR Asia[Title/Abstract] OR India[Title/Abstract] OR “Andhra Pradesh”[Title/Abstract])

#16 Search (#14) OR #15

#17 Search ((#6) AND #13) AND #16

#18 Search ((#6) AND #13) AND #16 Filters: Humans

S3 Table. Characteristics of studies excluded from results synthesis (n=11)

Author and Year	Country	Population and site	Study design, sampling (n)	Relevant outcomes	Relevant outcome(s) measured but not reported by employment	Covariates	Exposure and comparator(s)	Participants						Age range, years	Comments
								n (%)			Age, mean (SD)				
								Men	Women	Total	Men	Women	Total		
Atkinson et al. 2016	47 low- and middle-income countries (LMIC)	Urban and rural areas	Cross-sectional (on data from World Health Organization (WHO) Health Surveillance Information Study (n=70 countries)), selected on health surveillance needs (n=242,753)	**Excluded:** Physically inactive (International Physical Activity Questionnaire Score)	None	**Country level:** Human Development Index, economic development, urbanisation **Individual level:** age, gender, education, income, rural/urban residence	Agriculture White-collar Blue-collar Homemaker Unemployed Total			196,742				18, 69	
Barker et al. 2006	India	Rural village in agricultural community near Pune (Paba)	Cross-sectional, families from the Maternal Nutrition and Fetal Growth Study (n=797), non-random sample of married women in the village (n=180)	**Excluded:** Body Mass Index (BMI) (kilograms (kg)/metres (m)^2)	Fruit consumption	None	Farming Non-farming	63 30	79 19					Reproductive age	Husband and wife pairs of child bearing age with at least one son and one daughter aged three to eight years (mainly agricultural castes)
Gupta et al. 1997	Rajasthan, India	Three villages in rural communities (Bagoth, Badoo and Janjila in Parbatsar Tehsil, county of Nagaur)	Cross-sectional, entire communities in randomly selected villages (n=3,148)	**Excluded:** Coronary Heart Disease (CHD)	Alcohol, diastolic blood pressure (DBP), family history of diabetes and CHD and High-Density Lipoprotein (HDL), history of diabetes, tobacco, physical activity (PA); hypertension, Low-Density Lipoprotein (LDL), obesity, systolic blood pressure (SBP), total cholesterol (TC), triglycerides (TG), waist-to-hip-ratio (WHR)	Age, gender, education, anger/grief, prayer habit, family structure, religion, marital status, vegetarianism and intake of Ghee, (depression, stressful life events, housing and Yoga practice)	Agriculture Business Professional Government Household Total	1,303 377 4 298 0 1,982	180 19 3 31 933 1,166	1,483 396 7 329 933 3,148	39.87 (15)	36.81 (3)	38.73 (14)	>20	Desert population. Engage in farming activities for two to three months per year Low response rate for women (59.2%) due to Purdah custom (religious seclusion of women)

Author and Year	Country	Population and site	Study design, sampling (n)	Relevant outcomes	Relevant outcome(s) measured but not reported by employment	Covariates	Exposure and comparator(s)	Participants						Age range, years	Comments
								n (%)			Age, mean (SD)				
								Men	Women	Total	Men	Women	Total		
Panwar & Punia 1998a	Orissa, India	Pregnant women from six rural farming/non-farming villages in Haryana State	Cross-sectional, sampling methods not available (N/A) (n=90)	**Excluded:** fruit consumption (grams (g)/day (d)), vegetable consumption (g/d)	None	None	Farming Non-farming	0 0	45 45					Reproductive age	Pregnant women
Panwar & Punia 1998b	Orissa, India	Pregnant women from six rural farming/non-farming villages in Haryana State	Cross-sectional, sampling methods N/A (n=90)	Excluded: ascorbic acid (milligrams (mg)/d), folic acid (micrograms (µg)/d), riboflavin (mg/d)	None	None	Farming Non-farming	0 0	45 45					Reproductive age	Pregnant women
Poulter et al. 1984	Nyanza Province, Kenya	35 rural villages in Siaya District	Cross-sectional, random sample of Lou tribe (rural) and volunteers of potential (Lou) migrants (rural) (n=2,334)	**Excluded:** BMI (kg/m^2), DBP (millimetre mercury (mmHg)) (age-adjusted), SBP (mmHg) (age-adjusted)	None	Age, education, sodium level, potassium level, sodium:potasium ratio, weight	Land work Other work (teaching, machinery maintenance, lorry driving etc.) Total			2,334				≥17	Occupation data missing for 37% of participants Combines participants sampled by two methods (861 men and women from census and 1473 male volunteers)
Pritchard et al. 2016	Himachal Pradesh and Uttarakhand, India	Two (anonymous) rural villages, on each side of the boundary between Himachal Pradesh and Uttarakhand	Cross-sectional (of cohort), three-stage households (HH) baseline random survey strategy (purpose sampling of villages) (n=123 HHs)	**Excluded:** fruits (kilojoules (KJ)/d), vegetables (KJ/d)	None	None	HH farming own plot HH farming other's land/ share-cropping HH with one or more non-agricultural livelihood sources HH farming own plot and one or more non-farm livelihood sources Total	123 HHs						Adult HH members, adjusted for children	Men, women and children (results are adjusted for infants and children) Villages differ markedly in history of landholding, social patterning of land and food insecurity
Sengupta 2014	Andhra Pradesh and West Bengal, India	Fishermen from rural Araku valley of Visakhapatnam District, Andhra	Cross-sectional, random sample (site selection	**Excluded:** BMI (kg/m^2), DBP (mmHg), energy expenditure (EE)	None	Age	Fishermen College students	25 25	0 0		22.8 (1.92) 21.9 (2.25)				Non-smokers Students are different

Author and Year	Country	Population and site	Study design, sampling (n)	Relevant outcomes	Relevant outcome(s) measured but not reported by employment	Covariates	Exposure and comparator(s)	Participants						Age range, years	Comments
								n (%)			Age, mean (SD)				
								Men	Women	Total	Men	Women	Total		
		Pradesh and urban college students from Kolkata, West Bengal	N/A) (n=50)	(kcal minute (min)$^{-2}$), % Body fat (% BF) (%) (age adjusted), SBP (mmHg), waist circumference (WC) (centimetres (cm)), WHR											samples from the same college as other studies by Sengupta and colleagues
Sengupta & Sahoo 2013	Orissa and West Bengal, India	Fishermen slum in rural Puri, Orissa and urban college students from Kolkata, West Bengal	Cross-sectional, random sample (site selection N/A) (n=30)	**Excluded:** BMI (Kg/m^2), DBP (mmHg), EE (kcal min^{-2}), % BF (%) (age adjusted), SBP (mmHg)	None	Age	Fishermen (slum) College students	15 15	0 0		22.2 (2.70) 21.0 (2.25)			18, 25	Non-smokers Students are different samples from the same college as other studies by Sengupta and colleagues
Sengupta & Sahoo 2011	West Bengal, India	Fishermen slum in rural East Midnapore and urban college students from Kolkata, West Bengal	Cross-sectional random sampling (site selection N/A) (n=30)	**Excluded:** BMI (kg/m^2), DBP (mmHg), EE (Kcal min-2), % BF (%) (age adjusted), SBP (mmHg), WC (cm), WHR	None	Age	Fishermen (slum) College students	15 15	0 0		22.5 (2.97) 21.9 (2.16)			18, 25	Non-smokers Students are different samples from the same college as other studies by Sengupta and colleagues
Van Minh et al. 2003	Hatay Province, Vietnam	Rural areas in Bavi District	Cohort (longitudinal demographic surveillance system FilaBavi (n=49,543 person-years)), Village cluster design (n=15,193 person-years)	**Excluded:** Cardiovascular Disease mortality (CVD) (for subgroup)	CHD mortality, CVD mortality, pulmonary heart disease mortality, stroke mortality	Age, gender, education, economic condition	Farmer Government Other Non-pension retired Total			15,193 person-years				50, >70	Analysis restricted to subsample of participants aged ≥50 years (n=15 193/49 543 person-years)

BMI – Body Mass Index; CHD – Coronary Heart Disease; cm – centimetre(s); CVD Cardiovascular Disease; d – day; DBP – diastolic blood pressure; EE – energy expenditure; g – gram; HDL – High-Density Lipoprotein cholesterol; HH – household; kcal – kilocalories; kg – kilograms; KJ - kilojoule; LDL – Low-Density Lipoprotein cholesterol; LMIC – low- and middle-income

country; m – metre(s); mg – milligram; min – minute; mmHg – millimetre mercury; n – sample size; N/A – not available; PA – physical activity; SBP – systolic blood pressure; TC – total cholesterol; TG – triglycerides; WC – waist circumference; WHO – World Health Organization; WHR – waist-to-hip ratio; % BF – percent body fat; μg – microgram

S4 Table. Detailed characteristics of included studies (n=13)

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
Addo et al. 2006	Ghana, four rural farming communities (Sarpeiman, Opah, Ayikai Doblo and Amamoley)	Cross-sectional, purpose sampling (n=362)	Hypertension (≥140/90 millimetre mercury (mmHg))	Age, gender, education, smoking, alcohol, contraceptive use, work-related physical activity (PA), occupation, Body Mass Index (BMI), diabetes	Farmer			107			42.4 (18.6)	18, 99	73% of respondents have physically active employment and do other activities several times/ week
					Trader			152					
					Other			103					
					Total	107	255	362					
Arlappa et al. 2009	India, rural areas in nine states	Cross-sectional (of a rapid population-based cross-sectional study by the National Institute of Nutrition and the National Nutrition Monitoring Bureau), multistage random sampling (n=1,569)	Chronic energy deficiency (CED) (BMI <18.5 kilograms (kg)/metres $(m)^2$)	Caste, age pension, Annapurna, food for work program, acres of land	Agriculture			399				60, 70+	Subgroup (n=1,569/3,147 participants aged 18-70 years+); severely drought-affected districts; co-interventions offered to the most vulnerable
					Non-agriculture			1,170					

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
Asgary et al. 2013	Jamkhed, India, six villages in rural farming community	Cross-sectional, proportional random sampling by village (n=224)	Hypertension (≥140/90 mmHg)	Age, gender, occupation, income, abdominal girth, tobacco, alcohol	Farmer			112 (52.8)				40, 85	The farming area received health interventions
					Housekeeper			100 (47.1)					
					Total	95	129						
Balagopal et al. 2012	Gujarat, India, rural community	Cross-sectional (baseline of cohort), exhaustive (all village residents except migrants) (n=1,638)	Hypertension (systolic blood pressure (SBP) ≥140 mmHg); underweight, overweight, obese (BMI <18.5; 23-24.99; ≥25 kg/m^2); Tobacco	None	Agrarian (low socio-economic status)	362	402	764	43.3 (16.1)	43.3 (15.8)	43.4 (15.9)	18+	Migrant workers (1/5 of population) were not present and thus excluded
					Business (high socio-economic status)	404	470	874	40.1 (15.5)	40.4 (15.9)	40.2 (15.7)		
Gregory et al. 2007	Guatemala, participants of the Institute of Nutrition of Central America and Panama (INCAP) Longitudinal Study (1969–1977) (n=2392), born in four rural villages	Cross-sectional, non-random sample of available INCAP participants (n=360)	hypertension (≥130/85 mmHg); overweight (BMI ≥25 kg/m^2), obese (BMI ≥30 kg/m^2), smoker	Age, gender	Agriculture rural	88 (24.4)			31.7 (4.4)				High rates of infectious disease mortality and migration
					Non-agriculture rural	153 (42.5)			31.4 (4.2)				
					Urban	119 (33.1)			33.6 (4.3)				
Hazarika et al. 2004	Assam, India, native rural population from 25 villages	Cross-sectional, cluster random sampling of households (HHs) (districts selected on	Hypertension (≥140/90 mmHg)	Age, sex, marital status, extra salt intake, alcohol intake, smoking, BMI,	Service							≥30	
					Business								
					Cultivator								

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
		geographical location) in one state (n=3,180)		WHR, tobacco chewing, type of work	Daily wager Unemployed Others Total	1,441 (45)	1,739 (55)	3,180 (100)					
He et al. 1991	Sichuan Province, China, Yi farmers from remote mountain areas, Yi migrants and Han people living ≥ five years in Xichang City and county seats of Butuo, Meigu, and Zhaojue	Cross-sectional (of Yi Migrant Study (YMS)), Cluster randomised with probability proportional to size (n=14,505)	Hypertension I (140-159/90-94 mmHg) (age standardised), hypertension II (≥160/95) (age standardised); smoking	Age, BMI, smoking, alcohol, heart rate	Farmer Migrant Urban	5,023 1,656 2,173	3,218 919 1,516		30.9 34.8 34.8	31.8 31.3 32.9		15, 89	Married women usually moved to husband's village; Yi people are an ethnic minority. Farmers reside in remote mountain areas (altitude ≥1,500 m) with primitive life-styles
Norboo et al. 2015	Jammu and Kashmir, India, two groups from Leh town subdivision, 41 villages representative of six rural subdivisions: Leh block (n=12), Nubra (n=7), Kargil (n=6), Sham (Khalse), Zanskar (n=10),	Cross-sectional, two-stage stratified sampling (on urban/rural) (n=2,800)	Hypertension (≥140/90 mmHg); overweight (BMI ≥25 kg/m2)	Age, gender, obesity, rural/urban residence	Farmer Nomad Sedentary worker **Other, including:** *Housewife* *Manual labourer* *Monk*			1,247 220 549 784 *325* *63* *157*				20, 94	Rural participants are volunteers

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
	Changthang (n=6)				*No job*			*138*					
					Retired sedentary			*101*					
					Total			*2,800*			53.8 (15.0)		
Olugbile & Oyemade 1982	Two states, Nigeria, workers from two rural areas in two states: an agricultural company and a cement factory	Cross-sectional, random sampled from Agricultural Production and Supply Company (farmers), Factory workers sampled by 'stratified method' (n=276)	Hypertension (≥140/90 mmHg)		Agriculture company			112				20, 59	Working population; farmer, but not cement workers, may have received health interventions; farmers >59 years were excluded
					Factory worker			136					
Subasinghe et al. 2014	Andhra Pradesh, India, 12 rural villages surrounding the Rishi Valley Rural Health Centre, North Western region of Chittoor District	Cross-sectional, purpose sampling (those who presented at health centre after contact) (n=1,169)	CED (BMI <18 kg/m^2)	Age education, HH income, dietary energy	Non-government, government employee	170	206		.			18, 55+	(Partially) disadvantaged population; primarily low income subsistence farmers (excluded landowners not working on land)
					Self-employed	65	100						
					Farming and livestock	156	170						
					Homemaker	208	1						
					Unemployed, student, retired	43	50						
					Not working	0	48,160					15, 49	

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
Subramanian & Davey Smith 2006	26 states, India, rural and urban areas	Cross-sectional (subsample of National Family Health Survey from 26 Indian states (1998-99) (n=90,303)) (n=77,220)	Underweight (BMI <16; 16-16.9; 17-18.49; <18.5 kg/m^2), overweight, obese (BMI 23-24.9; 25-29.9; ≥30 kg/m^2)	Age, SES, caste, education, living environment, religion, parity, tobacco, alcohol; treatment of morbidities from asthma, malaria, tuberculosis	Non-manual	0	4,433						Non-pregnant women not attending school
					Agricultural	0	17,758						
					Manual	0	6,869						
Wang et al. 2010	South-western China, Yi farmers from remote mountain areas (Butuo, Zhaojue, Jinyang, Puge, and Xide counties), Yi migrants and Han people from the county seats and Xichang city	Cross-sectional (of the YMS), stratified random cluster sampling (all individuals in clusters sampled) (n=4,971)	Hypertension (≥130/85 mmHg); overweight and obesity (BMI ≥24 kg/m^2); smoking	None	Farmer	675	860		39.2 (12.3)	40.1 (11.5)		≥20	
					Migrant	760	546		40.3 (11.7)	37.4 (11.7)			
					Urban	1,080	1,050		43.6 (13.0)	45.0 (13.3)			
Zhou et al. 2003	Rural areas near big cities in Beijing, Northern China, and Guangzhou, Southern China	Cohort, sub-sample (<50%) of China-United States Collaborative Study on Cardiovascular and Cardiopulmonary Epidemiology (non-random sample)	Smoking, total cholesterol (millimoles (mmol)/ Litre (l)), triglycerides (mmol/L)	None	Agriculture 1983–84	326	0	326				35, 54 (baseline)	Rural areas near big cities
					Remained in agriculture 1993–94								
					Agriculture 1983–84	102	0	102					

Author and Year	Population and site	Study design and sampling (n)	Outcomes	Covariates	Exposure and comparators	Participants						Age range, years	Comments
						n (%)			Age, mean (SD)				
						Men	Women	Total	Men	Women	Total		
		(n=633)			Shifted out of agriculture 1993-94								
					Factory work 1983–84	135	0	135					
					Remained in factory work 1993-94								
					Office work 1983–84	70	0	70					
					Remained in office work 1993-94								

BMI – Body Mass Index; CED – Chronic Energy Deficiency; HH – household(s); INCAP – Nutrition of Central America and Panama longitudinal study; kg – kilograms; l – litre; m – metre(s); mmHg – millimetre mercury; mmol – millimoles; n – sample size; N/A – not available; PA – physical activity; SBP – systolic blood pressure; YMS – Yi Migrant Study

S5 Table. Quality of evidence rating[a] of included studies from five countries (n=13)

Author	Confounding	Selection of participants	Measurement of exposure	Departures from intended intervention	Missing data	Measurement of outcome	Selection of reported results	Overall bias
Addo et al 2006	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Well covered	Not described adequately to classify	Poorly addressed	Poorly addressed
Arlappa 2009	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described adequately to classify	Not described adequately to classify	Poorly addressed	Poorly addressed
Asgary et al 2013	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Poorly addressed	Not described adequately to classify	Poorly addressed	Poorly addressed
Balagopal et al 2012	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Well covered	Not described	Adequately addressed	Poorly addressed
Gregory et al 2007	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described adequately to classify	Not described adequately to classify	Poorly addressed	Poorly addressed
Hazarika et al 2004	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described	Not described adequately to classify	Poorly addressed	Poorly addressed
He et al 1991	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described	Not described adequately to classify	Poorly addressed	Poorly addressed
Noboro et al 2015	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described	Not described	Poorly addressed	Poorly addressed
Olugbile & Oyemade 1982	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Well covered	Not described	Not described adequately to classify	Poorly addressed
Subasinghe et al 2014	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described adequately to classify	Not described adequately to classify	Not described adequately to classify	Poorly addressed
Subramanian & Davey Smith 2006	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described	Not described adequately to classify	Poorly addressed	Poorly addressed
Wang C et al 2010	Poorly addressed	Poorly addressed	Not described adequately to classify	Not applicable	Not described	Not described	Poorly addressed	Poorly addressed
Zhou et al 2003	Poorly addressed	Poorly addressed	Not described adequately to classify	Well covered	Not described	Not described adequately to classify	Not described adequately to classify	Poorly addressed

[a] The risk of bias assessment was based on adapted 'A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions

CHAPTER 4 NIGHT-TIME LIGHT INTENSITY CONSISTENTLY ASSOCIATED WITH CARDIOVASCULAR DISEASE RISK FACTORS: A CROSS SECTIONAL STUDY OF THE ANDHRA PRADESH CHILDREN AND PARENTS STUDY, SOUTH INDIA

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Tina Bonde Sorensen
Principal Supervisor	Professor Alan Dangour
Thesis Title	Influences of place of residence on risk factors for atherosclerotic cardiovascular diseases in South India

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	BMJ Open
Please list the paper's authors in the intended authorship order:	Tina B Sørensen, Robin Wilson, John Gregson, Bhavani Shankar, Alan D Dangour and Sanjay Kinra
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	All authors contributed to study conception, revised the manuscript critically, approved the final manuscript for publication and agreed to be accountable for the work. Further contributions: Tina Bonde Sorensen and Dr John Gregson performed data management on the Andhra Pradesh Children and Parents Study (APCAPS) data. Tina

	Bonde Sorensen analysed the data and produced tables and figures, interpreted analyses and drafted the manuscript. Dr Robin Wilson extracted the night-time light intensity (NTLI) data and calibrated and prepared the NTLI data for statistical analysis. Professor Alan D Dangour provided expertise on epidemiology and health and Professor Sanjay Kinra provided expertise on epidemiology and cardiovascular diseases. Dr John Gregson advised on the analysis strategy.

Student Signature: [redacted] **Date:** 17 October 2019

Supervisor Signatu [redacted] **Date:** 17/10/19

Night-time light intensity consistently associated with cardiovascular disease risk factors: a cross sectional study of the Andhra Pradesh Children and Parents Study, South India

Tina B Sørensen,[1*] Robin Wilson,[2] John Gregson,[1] Bhavani Shankar,[3,4] Alan D Dangour[1,4] and Sanjay Kinra[1]

*Corresponding author. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine (LSHTM), London WC1E 7HT, United Kingdom. E-mail: Tina.Sorensen@lshtm.ac.uk. Telephone +44 (0) 2072994723. Orcid: https://orcid.org/0000-0003-1517-6159

[1]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

[2]Geography & Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK

[3]Centre for Development, Environment and Policy, School of Oriental and African Studies, 10 Thornhaugh Street, Russell Square, London WC1H 0XG, UK

[4]Leverhulme Centre for Integrative Research in Agriculture and Health (LCIRAH), 36 Gordon Square, London WC1H 0PD, UK

E-mails: Tina.Sorensen@lshtm.ac.uk, robin@rtwilson.com, John.Gregson@lshtm.ac.uk, b.shankar@soas.ac.uk, Alan.Dangour@lshtm.ac.uk, Sanjay.Kinra@lshtm.ac.uk

Abstract

Objectives: Explore associations of night-time light intensity (NTLI), a novel proxy for urbanisation levels, with mean levels of systolic blood pressure (SBP), body mass index (BMI), fasting serum low-density lipoprotein (LDL), and fasting plasma glucose (FPG), among adults at early stages of urbanisation in Telangana, South India.

Design: Cross-sectional analysis of cluster-randomised cohort data (third wave of the Andhra Pradesh Children and Parents Study)

Setting: The study was set in 28 villages representing a continuum of urbanisation levels, ranging from rural settlement to medium sized town, in Telangana, South India

Participants: Data were available from 6944 participants, 6236 of whom were eligible after excluding pregnant women, participants younger than 18 years of age and participants with missing data for age. Eligible participants were excluded if they were not fasting, had implausible or missing outcome values, were medicated for hypertension or diabetes, or had triglyceride levels invalidating derived LDL. The analysis included 5924 participants for BMI, 5752 participants for SBP, 5287 participants for LDL and 5328 participants for FPG.

Results: Increasing NTLI was positively associated with BMI, SBP and LDL. Adjusted mean differences across the range of village-level NTLI were 1.0 kg/m^2 (95% CI 0.01, 1.9, p=0.05) for BMI; 4.2 mmHg (95% CI 1.0, 7.4, p=0.01) for SBP; 0.3 mmol/l (95% CI -0.01, 0.7, p=0.06) for LDL; and -0.01 mmol/l (95% CI -0.4, 0.4, p=0.97) for FPG. Associations of NTLI with BMI and SBP were stronger in older age groups.

Conclusion: The association of NTLI with CVD risk factors identify NTLI as a potentially important tool for exploring urbanisation-related health. Consistent associations of moderate increases in urbanisation level with important CVD markers warrant prevention strategies to curb expected large public health impacts from continued rapid urbanisation in India.

Key words: Urbanisation, night-time light intensity (NTLI), environment, non-communicable disease (NCD), cardiovascular disease (CVD), CVD risk factors, epidemiology, Andhra Pradesh Children and Parents Study

Strengths and limitations of this study

We used a novel, standardised proxy of urbanisation level, remote sensing data on night-time light intensity measured at the village level

The villages had experienced rapid asymmetric urbanisation during decades preceding the study and represented a continuum of urbanisation levels ranging from rural settlement to medium sized town

The study used a large cluster-randomised population-based sample from 28 urbanising villages at early stages of urbanisation

Differences between excluded and included participants warranted some caution of generalising the findings to the general population

The cross-sectional design prevented us from making causal inferences or exploring the impact of urbanisation rate on CVD risk factors

Background

Cardiovascular diseases (CVDs), primarily coronary heart disease, are the leading causes of death (28%) and disability adjusted life years (14%) in India.[1] They occur 5-10 years earlier than in Western populations[2] and the age-standardised mortality rate (272 per 100000 persons) exceeds the global average and rates of some high-income countries.[3, 4] Individuals and societies in India therefore continue to experience substantial losses of productivity and income.[1, 5-7] Urbanisation is often described as a key driver of CVDs in low- and middle-income (LMICs) acting on CVD and CVD risk factors through interrelated changes in social, physical and build environments, socio-demographics, lifestyle and mental health. Some of these change are associated with harmful impacts on cardiovascular health, e.g. due to reduced physical activity and nutrition transitions, while others may be beneficial, e.g. through more and better education and access to health and social services.[8-12] A number of simple rural-urban comparisons and migration studies provide insight to the association of urbanisation level with CVD risk factors, such as overweight and obesity, hypertension, diabetes, and dyslipidaemia.[8, 13-21] The broad rural-urban dichotomies are useful, however insufficient for exploring which factors of urbanisation are most important for health in LMICs.[22] In part because no universal definition exists of 'rural' and 'urban' environments, and data on neighbourhood characteristics are often not available from these settings.[22] Furthermore, residents' characteristics tend to be more similar within than between urban and rural areas. For example, urban residents may on average enjoy wider access to goods and services, higher levels of education, occupation, affluence etc., whereas the opposite may be the case in rural areas.[22, 23] Such limited discordance at either end of the urbanisation continuum makes it difficult to tease out the underlying mechanisms of observed differences in CVD risk between the two extremes. Urbanisation factors may change at different rates with urbanisation,[22] however, and thus vary widely across the continuum from rural to urban areas. It may therefore be more useful and informative to study the impact of urbanisation level on CVD risk factors in transitioning populations, where the underlying mechanisms may be easier to separate. Unfortunately, it is challenging to accurately measure urbanisation as a continuum, whether using single- (e.g. population density) or multi-component classification systems (e.g. housing types and densities, economic activities, physical environment, services, etc.).[22] This is particularly true in LMICs where these types of data are often scarce. As an alternative to physical data based scales, remote sensing data have been widely used to characterise urban landscapes and scale of urbanisation.[24, 25] The night-time light intensity (NTLI) data, obtained by the United States' Defence Meteorological Satellite Programme's Operational Linescan System (DMSP-OLS), are suggested a valid proxy for urbanisation level and dynamics due to their strong and consistent relationships with demographic and economic indicators of urbanisation such as urban build up area, population

density, economic activity, and energy consumption at global, regional and local levels.[24-27] The raw satellite data is processed into annual averages of light intensity from persistent sources associated with human settlement, i.e. excluding transient light sources such as lightning, fire; aurora, solar and lunar light. The processed data are aggregated into 30 arc seconds grids[28] and published as unitless digital number ranging from 0 (no light) to 63 (light saturation).[24, 27] The NTLI data, which have near global coverage and a freely available historical data archive dating back to 1992,[29] offer a unique opportunity to study the association of urbanisation, measured by a single standard measure world-wide, with CVD risk. In the current study we used a continuum of NTLI data to explore associations of urbanisation level and mean levels of four leading risk factors for CVDs, systolic blood pressure (SBP), body mass index (BMI), fasting serum low-density lipoprotein (LDL), and fasting plasma glucose (FPG), in adults from 28 villages in rapidly urbanising Telangana, South India. Improved understanding of the association of urbanisation level and CVD risk factors in transitioning areas will help inform policy for disease prevention and control. Such policies are imperative in India, where economic growth and urbanisation continues, to avoid further adverse economic and social impacts of (premature) CVD morbidity and mortality.

Materials and methods

Study design, setting and participants

Data from the third survey wave of the Andhra Pradesh Children and Parents Study (APCAPS)[30] were used for the current cross-sectional study. The APCAPS is a prospective, intergenerational cohort study set in 29 villages (currently 28, as two villages have merged over time), which are located 30-80 km from Hyderabad in the Ranga Reddy district of Telangana, India. The APCAPS was established in 2003-05 as a long term follow-up of the cluster-randomised Hyderabad Nutrition Trial (HNT) (1987-90), in which all pregnant women and young children received (15 villages) or awaited (14 villages) a nutrition intervention.[23] In 2010-12, a third APCAPS survey wave included all available HNT index children, their parents and siblings (n=6944).[31, 32]

Sample size and selection

All adults (18 years or older) from the third APCAPS survey wave, excluding pregnant women (n=80), were eligible for inclusion (n=6236). Participants who reported taking medication for outcome-related conditions were excluded to reduce the risk of biasing results, e.g. towards the null if participants with medically controlled blood pressure sustained high levels of risk factors for hypertension. Participants were excluded from analyses of glucose if they reported fasting less than eight hours at the time of giving the blood (n=366) or currently taking medication

(tablets or insulin) for diabetes (n=94). Participants were excluded from analysis of LDL if they reported fasting less than nine hours (n=364) or had triglyceride levels above 400 mg/dl[33] (n=92). Participants who reported taking antihypertensive medication were excluded from analyses of SBP (n=197) (Fig. 1). One village was significantly larger and had a considerably higher NTLI value than the other villages and its participants were excluded (n=297) to avoid overreliance of results from this outlier village.

Data

Determining urbanisation levels of APCAPS villages

Aerial tracing of Bing satellite imagery, using Open Street Map software and GPS-based ground surveying during 2012-13, defined APCAPS village boundaries. The village boundary data were overlaid with the NTLI grids from 2011 (technical issues with satellite sensors prevented the use of NTLI data from 2012).[29] Twenty-four villages were intersected by two or more pixels, which in some cases had different NTLI values. Therefore, to obtain one NTLI value (urbanisation level) per village, we used a super-resolution method to resample the NTLI images to a higher-resolution, before summing the pixels intersecting with the village boundary to produce an overall village level NTLI value. This ensured giving the appropriate weighting to each of the pixel values present over each village. Finally the data were calibrated using the ridgeline sampling regression methods[29] and oversaturation of light was removed manually. Since the first survey wave in 2003-5, the 27 included APCAPS villages had changed from rural agricultural communities with similar NTLI levels to represent a continuum of NTLI levels with growing differentiation at higher levels.

Selection of outcomes, CVD risk factors

We pre-specified the four outcomes, BMI, SBP, LDL and FPG, from a number of CVD risk factors available in the APCAPS, due to their well-established strong relationships with CVDs and their modifiable nature.[33, 34] Continuous outcomes were selected for analysis to enable exploration of linear relationships between increasing level of NTLI and the CVD risk factors.

Health and socio-demographic data

Trained field staff collected data on health and socio-demographic information in the local language, Telugu, using semi-structured questionnaires. Trained field staff and medical doctors performed the clinical assessments in local clinics and at the National Institute of Nutrition in Hyderabad. Weight was measured twice to the nearest 0.1 kg with digital weight scales (Seca Leicester 899; Chasmors Ltd, London, United Kingdom (UK)), while wearing light clothing and no shoes. Standing height was measured twice to the nearest 0.1 cm with a portable stadiometer

(Seca Leicester height measure; Chasmors Ltd, London, UK). Blood pressure was measured three times at the right arm in the sitting position using a digital device and appropriate sized cuff (Omron hem 7300; Matsusaka Co., Japan).

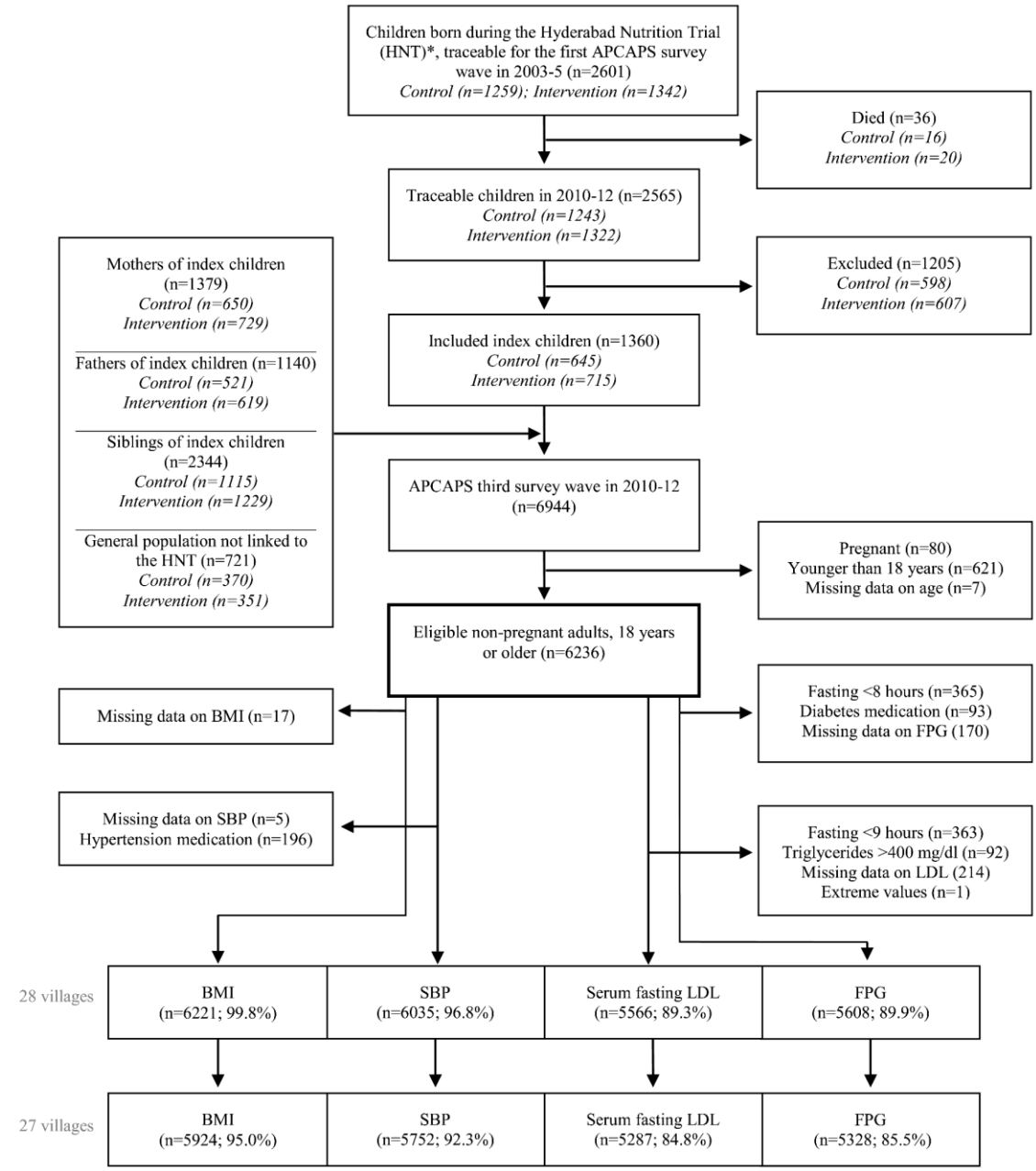


Fig. 1 Sampling and participant flow of the current study

BMI – body mass index; FPG – fasting Plasma glucose; HNT – Hyderabad Nutrition Trial; LDL – low-density lipoprotein; SBP – systolic blood pressure

* The Hyderabad Nutrition Trial (1987-1990) included 15 intervention and 14 control villages from the Integrated Child Development Services Scheme (stepped-wedge cluster randomised nutrition supplementation trial)

Biochemical assays

Venous fasting (eight to 12 hours) blood samples were collected, separated within 30 minutes, and stored locally at -20°C. Biochemical assays were performed using the Cobas311 auto-analyser at the Genetics and Biochemistry Laboratory at The South Asia Network for Chronic Diseases of the Public Health Foundation of India, New Delhi. Plasma glucose was assessed on the day of sampling by enzymatic Glucose Oxidase/Peroxidase-4-Aminophenazone-pPenol (GOD-PAP) method (Randox Laboratories; London, UK). Serum total cholesterol and triglycerides were assessed by enzymatic methods using Cholesterol Oxidase-Peroxidase-4-Aminophenazone-Phenol (CHOD-PAP) and Glycerol Phosphate Oxidase-Peroxidase-4-Aminophenazone-Phenol (GPO-PAP) (Roche Diagnostics GmbH; Mannheim, Germany). High-density lipoprotein (used for derivation of LDL) was derived by direct method using Polyethylene Glycol (PEG) modified Cholesterol Esterase, Cholesterol Oxidase and Dextran Sulphate (Roche Diagnostics GmbH; Mannheim, Germany). Low-density lipoprotein level was derived indirectly using the Fridewald-Fredrickson formula: LDL=total cholesterol–HDL–triglycerides/5 (with all measures in mg/dl).[33, 35]

Quality assurance

Trained field staff and medical doctors followed standardised procedures and anthropometric equipment were calibrated at the start of every clinic. Reproducibility of clinical measurements were evaluated by repeating measures on a 5% random subsample after one to three weeks. The consistency was high for all measures with intra-class correlation coefficient for anthropometric measurements >0.98; blood pressure >0.85; and biochemical assays >0.94.[36] The quality of biochemical assays were monitored by the Cardiac Biochemistry laboratory at the All India Institute of Medical Sciences, which participates in the UK National External Quality Assessment Programme and the External Quality Assessment Scheme of Randox International.

Statistical methods

The two weight and height measurements were averaged and BMI calculated for the analyses as weight (kg)/height $(m)^2$. The last two measurements of SBP were averaged for the analysis. Variables were assessed for outliers, cleaned accordingly, and examined for normality. Night-time light intensity data were non-normally distributed and log-transformed for analysis. To preserve study power, we allowed sample sizes to vary by outcomes. Mixed effects linear regression models with clustering by household and village were fitted in all analyses to account for the hierarchical nature of the data and the sampling strategy. Night-time light intensity (log transformed), BMI, SBP, LDL, FPG, age, and room temperature at clinical assessment (the latter for analysis of SBP) were included as continuous variables. Gender (women and men),

caste (General caste, Scheduled caste, Scheduled tribe, Other backward class, and other), religion (Muslim, Hindu, Christian, and other), current marital status (married and unmarried), season of survey and season of birth (summer (March-May), South West monsoon (June-October), and winter (November-February)) were included as categorical variables. Study characteristics were presented by thirds of ranked NTLI, referred to as low, medium, and high levels. We evaluated potential confounders for each outcome using the double-selection method.[37] One final set of covariates, identified as potential confounders for at least one exposure-outcome relationship, were used in analyses of all four outcomes. Level of urbanisation may affect CVD risk profiles of different genders and age groups differently due to e.g. gender inequality, employment patterns, and readiness to adopt new lifestyles and behavious.[23, 38-40] We therefore tested for interaction by gender and age. We did not consider socio-demographic factors such as education, occupation, and socio-economic status as potential confounders or effect modifiers due to their importance on the causal pathway from urbanisation level to CVD risk factors. A sensitivity analysis among HNT index children (n=1245 (21% of overall sample); intervention n=622, control n=623) did not indicate confounding or interaction by intervention status. Results of a supplementary analysis for room temperature-adjusted SBP were presented separately, due to a large proportion of missing data (Additional Fig. 3 and Additional Tab. 2-3). A sensitivity analysis was performed to evaluate whether borderline outlier villages (with higher or lower observed mean outcomes than the majority of villages) influenced regression estimates considerably. Differences between included and excluded participants were examined with Chi square, and linear, and logistic mixed effects models with clustering by village and household. All analyses were performed in Stata 14.

Results

A total of 6236 non-pregnant adults were eligible for inclusion in the current study (90% of the third APCAPS survey wave). After removing extremes and participants with missing outcome data for specific analyses, the final samples came to 5924 participants for BMI; 5752 participants for SBP; 5287 participants for LDL; and 5328 participants for FPG (Fig. 1). Approximately half of participants were women (Tab. 1). The overall median age was 32 years (interquartile range (IQR) 24, 48), with women on average older than men. Men and women most frequently identified with other backward caste and Hindu religion. Women were more likely to be married (76%) than men (59%). Village-level NTLI ranged from 61.7 (equivalent to 4.1 on the log scale) to 1081.1 (equivalent to 7.0 on the log scale). The distribution of socio-demographic characteristics were similar across low, medium and high level of NTLI. None of

Tab. 1 Participant characteristics of APCAPS adults from 27 villages, 2010-12 (n=5932)

	NTLI [a], women (n=2753 [46 %])			NTLI [a], men (n=3179 [54 %])		
	Low	**Medium**	**High**	**Low**	**Medium**	**High**
NTLI (unitless), median (IQR)	138.2 (121, 149)	202.4 (185, 273)	492.8 (485, 753)	138.2 (129, 149)	202.4 (185, 273)	492.8 (396, 753)
Age, median (IQR)	40 (25, 46)	40 (25, 46)	38 (25, 45)	28 (23, 51)	29 (24, 50)	28 (23, 49)
Age group, n (%)						
<30 years	299 (38.7)	372 (38.2)	416 (41.3)	489 (54.5)	595 (52.7)	636 (55.2)
30-39 years	85 (11.0)	106 (10.9)	134 (13.3)	57 (6.4)	74 (6.6)	91 (7.9)
40-49 years	261 (33.8)	348 (35.8)	348 (34.5)	98 (10.9)	151 (13.4)	154 (13.4)
50+ years	127 (16.5)	147 (15.1)	110 (10.9)	254 (28.3)	309 (27.4)	271 (23.5)
Total	772 (100)	973 (100)	1008 (100)	898 (100)	1129 (100)	1152 (100)
Caste, n (%)						
General caste	51 (6.6)	79 (9.2)	76 (7.5)	51 (5.7)	89 (8.9)	63 (5.5)
Schedule caste	277 (35.9)	375 (43.6)	292 (29.0)	331 (36.9)	428 (42.6)	335 (29.1)
Schedule tribe	9 (1.2)	3 (0.4)	5 (0.5)	11 (1.2)	3 (0.3)	7 (0.6)
Other backward	425 (55.1)	374 (43.5)	628 (62.3)	488 (54.3)	455 (45.3)	739 (64.2)
Other	9 (1.2)	29 (3.4)	7 (0.7)	17 (1.9)	30 (3.0)	8 (0.7)
Total	771 (100)	860 (100)	1008 (100)	898 (100)	1005 (100)	1152 (100)
Religion, n (%)						
Muslim	19 (2.5)	37 (4.3)	71 (7.0)	16 (1.8)	45 (4.5)	64 (5.6)
Hindu	742 (96.2)	786 (91.4)	919 (91.2)	875 (97.4)	931 (92.6)	1079 (93.7)
Christian	9 (1.2)	37 (4.3)	16 (1.6)	7 (0.8)	29 (2.9)	8 (0.7)
Other	1 (0.1)	0 (0)	2 (0.2)	0 (0)	0 (0)	1 (0.1)
Total	771 (100)	860 (100)	1008 (100)	898 (100)	1005 (100)	1152 (100)
Marital status, n (%)						
Not married	196 (25.4)	238 (24.5)	229 (22.7)	381 (42.4)	459 (40.7)	487 (42.3)
Married	576 (74.6)	734 (75.5)	779 (77.3)	517 (57.6)	669 (59.3)	665 (57.7)
Total	772 (100)	972 (100)	1008 (100)	898 (100)	1128 (100)	1152 (100)

	NTLI [a], women (n=2753 [46 %])			NTLI [a], men (n=3179 [54 %])		
	Low	**Medium**	**High**	**Low**	**Medium**	**High**
Hyderabad Nutrition Trial, n (%)						
Control	643 (83.3)	826 (84.9)	835 (82.8)	667 (74.3)	852 (75.5)	865 (75.1)
Intervention	129 (16.7)	147 (15.1)	173 (17.2)	231 (25.7)	277 (24.5)	287 (24.9)
Total	772 (100)	973 (100)	1008 (100)	898 (100)	1129 (100)	1152 (100)
Room temperature (^{0}C), mean (95% CI)	28.7 (3.4)	30.2 (3.4)	30.1 (2.6)	28.3 (3.5)	29.8 (3.2)	29.6 (2.6)

CI – confidence interval; IQR – Inter quartile range; n – sample size; NTLI – night-time light intensity

a Characteristics are presented by thirds (low, medium and high level) of ranked NTLI

the included (non-medicated) participants reported ever being diagnosed with heart disease, including stroke.

Model predicted mean BMI, SBP and LDL increased linearly with rising NTLI in crude and adjusted models (Fig. 2 and Tab. 2). Across the range of NTLI, the fully adjusted predicted mean increased 1.0 kg/m^2 (95% CI 0.01, 1.9, p=0.05) for BMI; and 4.2 mmHg (95% CI 1.0, 7.4, p=0.01) for SBP. The pattern was similar for LDL with a fully adjusted predicted mean increase across the range of NTLI of 0.3 mmol/l (95% CI -0.01, 0.7), although the evidence was weak from crude (p=0.06) and fully adjusted models (p=0.06). There were no linear associations of NTLI with FPG (fully adjusted mean difference across the range of NTLI: -0.01 mmol/l (95% CI -0.4, 0.4, p=0.97).

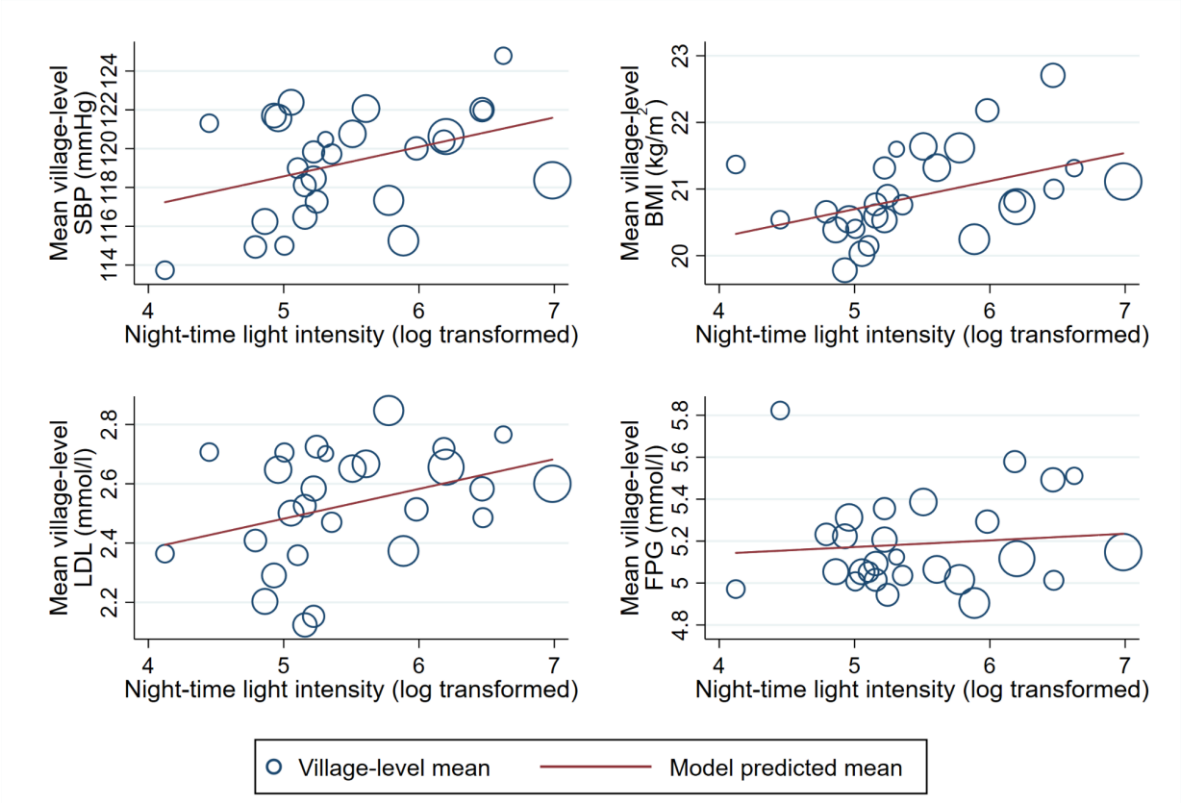


Fig. 2 Crude associations of night-time light intensity with CVD risk factors among APCAPS adults, 2010-12 (n=5937)

APCAPS – Andhra Pradesh Children and Parents Study; BMI – body mass index; FPG – fasting plasma glucose; LDL – low-density lipoprotein; NTLI – night-time light intensity; SBP – systolic blood pressure

Model predicted means were derived from multilevel linear regression with clustering by household and village, using individual-level outcome data

Marker size proportional to village size

Tab. 2 Crude and adjusted associations of NTLI with CVD risk factors among APCAPS adults, 2010-12 (n=5937)

		Model predicted crude mean (95% CI) at the lowest NTLI (61.7 (4.1 on the log scale))	Model predicted crude mean (95% CI) at the highest NTLI (1081.1 (7.0 on the log scale))		Model predicted age and gender adjusted mean change (95% CI) with increasing NTLI[c]			Model predicted fully adjusted mean change (95% CI) with increasing NTLI[c] (adjusted for age, gender, caste, religion, marital status and survey season)		
	n			p-value	n	β (95% CI)	p-value	n	β (95% CI)	p-value
SBP (mmHg)[a]	5752	117.2 (115.1, 119.4)	121.6 (119.2, 123.9)	0.04	5747	1.8 (0.4, 3.2)	0.01	5514	1.5 (0.3, 2.6)	0.01
BMI (kg/m2)	5924	20.3 (19.8, 20.8)	21.5 (21, 22.1)	0.01	5920	0.4 (0.1, 0.8)	0.01	5682	0.3 (0.03, 0.7)	0.05
Fasting LDL (mmol/l)	5287	2.4 (2.2, 2.6)	2.7 (2.5, 2.9)	0.06	5285	0.1 (0.01, 0.2)	0.04	5051	0.1 (-0.003, 0.2)	0.06
Fasting glucose (mmol/l)[b]	5328	5.1 (5.0, 5.3)	5.2 (5.0, 5.4)	0.61	5326	0.04 (-0.1, 0.2)	0.52	5094	-0.002 (-0.1, 0.1)	0.97

BMI – body mass index; CI - confidence interval; LDL – low-density lipoprotein; n – sample size; NTLI – night-time light intensity; SBP – systolic blood pressure; β – beta-coefficient

Model predicted means (95% CIs), β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression models with clustering by household and village (using individual-level outcome data)

Participants were excluded if medicated for hypertension[a] or diabetes[b]

c Mean change per whole number increase in log transformed NTLI

Model predicted mean SBP with increasing NTLI was on average 6.1 mmHg (95% CI 5.42, 6.9, p<0.001) higher among men than women, whereas levels of BMI, LDL and FPG did not differ by gender. However, there was no statistical evidence of gender modifying any of the associations between NTLI and the CVD risk factors. There was strong evidence of age modifying the associations of NTLI with BMI (p<0.001) and SBP (p<0.001), but not LDL, or FPG. Night-time light intensity was not associated with BMI or SBP in the younger age groups (up to 39 years), whereas the evidence and magnitudes of associations from both crude and fully adjusted models were stronger in older age groups (from 40 years) than predicted in the overall analysis (Additional Fig. 1 and Additional Tab. 1). Additional Fig. 2 illustrates at which levels of NTLI the fully adjusted models predicted the greatest probability of associations differing by age group. Overall and age-specific results of a supplementary analysis for SBP adjusted for room temperature at clinical assessment (n=3160) were similar to the main and age-specific SBP analyses (Additional Tab. 2-3 and Additional Fig. 3).

Sensitivity analysis

A sensitivity analysis removing three borderline outlier villages from analysis of BMI, SBP and FPG (none identified for LDL) were consistent with the overall and age-stratified main analyses (Additional Fig. 4-5 and Additional Tab. 4-5). However, results for SBP (not adjusted for room temperature) were not statistically significant at the 95% level. Additional Fig. 6 illustrates at which levels of NTLI the fully adjusted sensitivity analysis predicted the greatest probability of differences in associations of NTLI and CVD risk factors by age group

Discussion

Among 6236 participants, model predicted mean BMI, SBP and serum fasting LDL, but not FPG, increased linearly with rising level NTLI after controlling for confounders. We expected younger individuals to more readily adopt new lifestyles and behaviours associated with urbanisation and CVD risk factors than older individuals.[9, 39] This was however not apparent from the data. The observed stronger associations of NTLI with BMI and SBP in older age groups were likely attributable to advancing age.[33] Gender did not modify associations.

Comparison with studies using night-time light intensity

Few studies have used NTLI data to explore relationships of urbanisation and CVD risk factors. A Korean cross-sectional study (n=8526)[41] and a multi-country study (n=130 counties),[42] investigated associations of DMSP NTLI with BMI. Consistent with our results on BMI, increasing level of NTLI was associated with rising levels of overweight and obesity in both studies. A cross-sectional study from India, which analysed NTLI data from NASA's Visible

Infrared Imaging Radiometer Suite[43], corroborates our adjusted results for SBP, although the included set of covariates differed between our studies. We did not find any studies on relationships between NTLI and LDL or FPG.

Comparison with studies using multi-component urbanisation scores

Our findings on BMI and SBP agree with cross-sectional studies from South India[15] (n=3705) and Sri Lanka (n=4485),[18] which used multi-component scores to estimate urbanisation. Although we cannot compare our results directly to studies that use urbanisation scores (as opposed to NTLI) the direction of effects are consistent with our findings and support evidence of an association. We found no studies exploring relationships between urbanisation score and FPG, whereas findings from Sri Lanka[18] and China[44] suggest a greater likelihood of diabetes as urbanisation score increased. Restricting our analysis to non-diabetics may have contributed to the null results for FPG, particularly as excluded individuals were on average older and had higher BMI.

Comparison with rural-urban and migrant studies

Measuring urbanisation is challenging, particularly in settings where population and environment data are sparse.[22] As a result, relationships of urbanisation and CVD risk factors are often inferred from rural-urban comparisons and migrant studies. Several of these studies focus on or include Asian populations. An Indian migration study of 3537 sibling pairs suggests that mean BMI is lower among rural than migrant and urban populations. Patterns were similar for mean SBP, LDL, and FPG among men but not women.[40] In a different analysis of the same study, cumulative exposure to urban environments additionally increased mean BMI of both genders as well as SBP and fasting glucose of men, while LDL remained unaffected.[13] Systolic blood pressure (as the only outcome) was on average higher among men than women in our sample, however gender did not modify associations. A cross-sectional study from Southern India suggests that mean BMI, SBP, and the prevalence of diabetes were lower among peri-urban residents of both genders than among town and city dwellers.[45] Urban versus rural residence was also associated with higher prevalence of diabetes[44] and mean LDL in China[46] after adjusting for important confounders. Two large studies from 56[8] and 36[47] countries (n=878000 and 148579 respectively) across Africa, the Middle East, Asia, Americas, the Caribbean and USA suggest that overweight (BMI $\geq$25 kg/m^2) is considerably more common among urban than rural women, particularly in lower income countries.[8] In a subsample from India (n=7608), the prevalence of overweight was almost five fold among urban (26.4 %) compared to rural women (5.6 %).[47]

Strengths and limitations

A major contribution of our study to the existing evidence was exploring CVD risk factors in a transitioning population, i.e. one that has experienced rapid and uneven urbanisation over time and thus represented a continuum of urbanisation levels at the time of data collection. The village-level NTLI has been validated as a standardised and comparable proxy of urbanisation level, reflecting extent of build-up area, population density, economic activity, and energy consumption world-wide.[24-27] It may further be a more objective measure than the survey based data used for multiple-component urbanisation scales, which are at risk of both interviewer and reporting biases. The NTLI data have gained popularity for exploring associations of urbanisation level and artificial light at night with human health, particularly cancer,[48-56] overweight and obesity.[41, 42, 57] The NTLI time series has further been suggested a valid proxy for urbanisation dynamics over time,[24, 27] especially in rapidly developing countries with high urban growth rates such as China, India, and Brazil.[27] Recent efforts to calibrate the NTLI time series has shown great promise for studying urbanisation dynamics in these settings[29] and offer a unique resource for health research worldwide. We calibrated the NTLI data using a robust and reliable semiautomatic method to reduce systematic bias from satellite discrepancies and enable more efficient and precise comparisons of results between studies from different regions and time points.[29] However, the semiautomatic method could introduce new bias from manual processing.[29] The overall agreement of our result with existing evidence (using different measures of urbanisation level) supports the utility of NTLI to predict urbanisation-related changes in CVD risk factors. The large sample size and use of hierarchical statistical methods should reduce overestimation of the magnitude of associations from risk factor clustering by sampling units (villages and households). The comparisons of excluded and included participants did however warranted some caution of generalising the results to the general population. Trained field and clinical staff followed standardised protocols and collected information in the local language. A comprehensive list of factors were measured and explored as potential confounders and effect modifiers in our study, however, some residual confounding from unmeasured factors may remain. Data for a number of covariates were self-reported, and some recall or reporting bias cannot be rule out. The cross-sectional design of the present study further prevented us from inferring causality of the results as well as from exploring the relative contribution of different urbanisation elements to CVD risk.

Possible mechanisms

There are a number of plausible pathways through which urbanisation may deteriorate cardiovascular health. Physical and build environments change with urbanisation and in turn influence the level of exposure to noise and air pollution, green spaces, infrastructure.[22, 43, 58, 59]

Urbanisation is also associated with changes in occupation patterns and socio-economic status[15, 58]; social norms, cohesion, and support; as well as personal tastes and preferences.[22, 43, 59] Together with increased availability, access and advertising of commodities, these transitions promote lifestyle changes that are strongly associated with CVD risk, e.g. 'nutrition transition' towards diets high in fat, sugar and salt as well as processed and convenience foods; physical inactivity during work, leisure time, transport and household chores; tobacco use and alcohol consumption.[9, 18, 22, 58, 60] Cumulative exposure to urban environments additionally appears to increase CVD risk.[13, 17] Recent studies further suggest that increased exposure to artificial light at night observed with increasing urbanisation level supresses melatonin production, disrupts circadian rhythm, and leads to physiological and behavioural changes associated with CVD.[42] A number of studies focusing on artificial light at night (other than NTLI) without accounting for urbanisation support this mechanism.[49, 61-66] At the same time, greater level of urbanisation is also associated with number of changes that may improve CVD health. Greater level of urbanisation is for example associated with greater levels and higher quality of education; improved opportunities for paid work, particularly for women; higher socio-economic status and living conditions; better access to health, social services and safe food, which have great potential to improve cardiovascular health.[67-72] Although these pathways have been studies extensively, no study has to our knowledge attempted to quantify their relative harmful and beneficial contributions.

Implications

Our findings indicate that the harmful impacts of increasing urbanisation level on SBP and BMI, well-established leading modifiable risk factors for CVDs, outweigh any beneficial impacts even with moderate increases in urbanisation level. These findings suggest an opportunity for policy makers and urban planners to help curb expected increases in urbanisation-related CVDs by addressing underlying and immediate risk factors for elevated SBP and BMI during early stages of urbanisation. We will however need to understand better and quantify the implicated pathways in different settings in order to identify cost-effective prevention strategies.

Conclusion

We found that the NTLI data, a novel measure of urbanisation, is a useful tool to predict changes in important urbanisation-related CVD risk factors where objective data on urbanisation factors are not readily available. The observed consistent associations of moderate increases in urbanisation level (measured by NTLI) with CVD risk factors provide an important early warning for the potential progression of the CVD burden in South India with continued

urbanisation. To curb expected large public health impacts from continued and rapid urbanisation in India, further research is warranted to identify the most important mechanisms. To enable targeting of cost-effective prevention strategies, we suggest that future studies focus on transitioning populations and use causal methods to better understand potentially important pathways.

Abbreviations

APCAPS	Andhra Pradesh Children and Parents Study
BMI	body mass index
CI	confidence interval
CVD	cardiovascular disease
DMSP-OLS	Defence Meteorological Satellite Programme's Operational Linescan System
FPG	fasting plasma glucose
HNT	Hyderabad Nutrition Trial
IQR	inter quartile range
LDL	low-density lipoprotein
LMIC	low- and middle-income country
LSHTM	London School of Hygiene and Tropical Medicine
n	sample size
NTLI	night-time light intensity
NCD	non-communicable disease
SBP	systolic blood pressure
UK	United Kingdom
β	Beta-coefficient

Declarations

Ethics

Ethics approvals were obtained from Public Health Foundation of India, New Delhi, India; National Institute of Nutrition, Hyderabad, India; and the London School of Hygiene and Tropical Medicine (LSHTM), London, UK. Approvals were obtained from all village heads and their committees. Study participants provided written informed consent or a witnessed thumbprint if illiterate prior to study start.

Consent for publication

Study participants provided written informed consent, or a witnessed thumbprint if illiterate, for the use and publication of their anonymised data in accordance with the study objectives.

Availability of data and materials

The dataset analysed during the current study is available from the APCAPS (http://apcaps.lshtm.ac.uk/)[30] upon reasonable request.

Competing interests

The authors declare that they have no competing interests

Role of the funding source

This work was supported by Wellcome Trust (www.wellcome.ac.uk/) Grant number 083707 and the Bloomsbury PhD Studentship (http://www.bloomsbury.ac.uk/) through LSHTM, London, UK

Author contributions

All authors contributed substantially to study conception and manuscript revision; and approved the final manuscript for publication. TBS and JG cleaned and managed the Andhra Pradesh Children and Parents Study (APCAPS) data in collaboration with the APCAPS team. RW extracted, calibrated and managed the night-time light intensity data; JG advised on statistical methods. TBS conducted the literature search, data analysis and interpretation, and drafted the manuscript; SK is the director of the APCAPS.

Acknowledgements

We would like to thank the APCAPS team and support staff, in particular (in alphabetical order by surname) Srivalli Addanki, Christopher Baker, Santhi Bhogadi, Naveen Chittaluri, Ruby

Gupta, Bharati Kulkarni, Poppy Mallinson, Sandeep Masapeta, Subhash Masapeta, and Chitra Sharma.

References

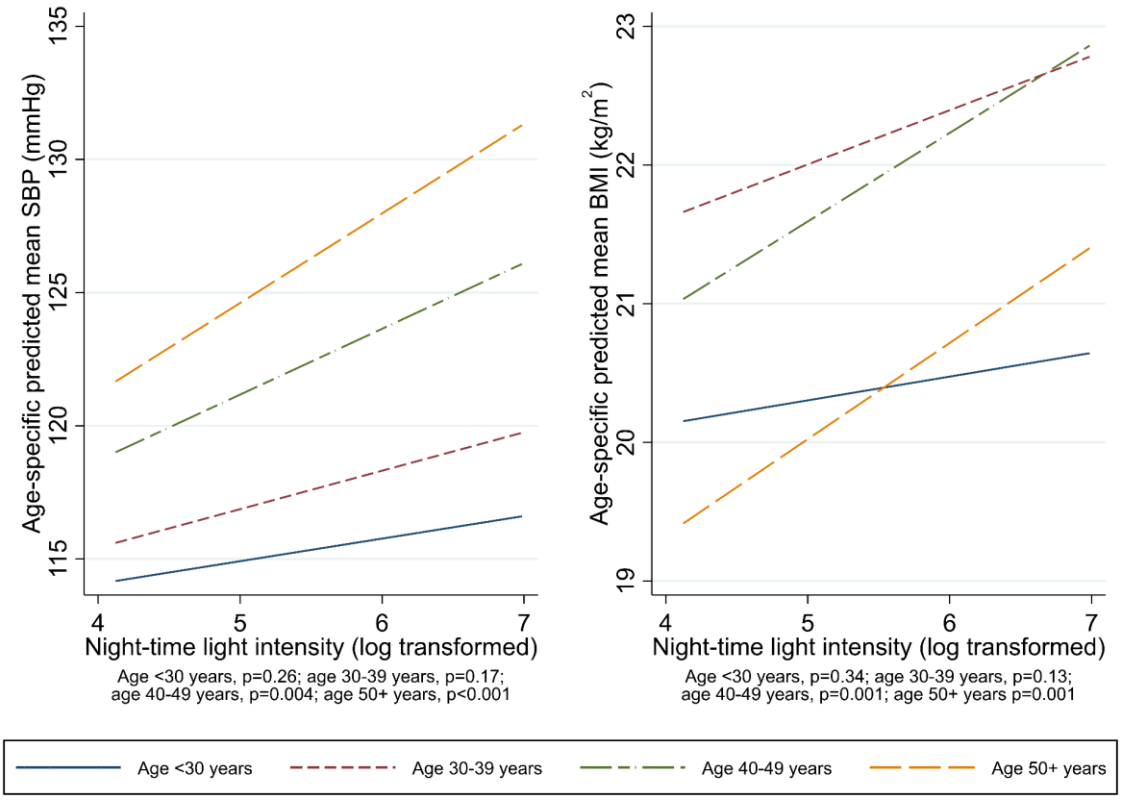
1. India State-level Disease Burden Initiative Collaborators. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet. 2017 Nov 13
2. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India: Current Epidemiology and Future Directions. Circulation. 2016 Apr 19;133(16):1605-20
3. Patel V, Chatterji S, Chisholm D, Ebrahim S, Gopalakrishna G, Mathers C, et al. Chronic diseases and injuries in India. Lancet. 2011 Jan 29;377(9763):413-28
4. Yusuf S, Rangarajan S, Teo K, Islam S, Li W, Liu L, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. The New England journal of medicine. 2014 Aug 28;371(9):818-27
5. Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart (British Cardiac Society). 2008 Jan;94(1):16-26
6. Sharma M, Ganguly NK. Premature coronary artery disease in Indians and its associated risk factors. Vascular health and risk management. 2005;1(3):217-25
7. Prabhakaran D, Singh K. Premature coronary heart disease risk factors & reducing the CHD burden in India. The Indian journal of medical research. 2011 Jul;134:8-9
8. Goryakin Y, Suhrcke M. Economic development, urbanization, technological change and overweight: what do we learn from 244 Demographic and Health Surveys? Economics and human biology. 2014 Jul;14:109-27
9. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012 Jan;70(1):3-21
10. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008 Nov;85(6):938-51
11. Vlahov D, Galea S. Urbanization, urbanicity, and health. Journal of urban health : bulletin of the New York Academy of Medicine. 2002 Dec;79(4 Suppl 1):S1-s12
12. Moore M, Gould P, Keary BS. Global urbanization and impact on health. International journal of hygiene and environmental health. 2003 2003/01/01/;206(4):269-78
13. Kinra S, Andersen E, Ben-Shlomo Y, Bowen L, Lyngdoh T, Prabhakaran D, et al. Association between urban life-years and cardiometabolic risk: the Indian migration study. American journal of epidemiology. 2011 Jul 15;174(2):154-64
14. Bowen L, Ebrahim S, De Stavola B, Ness A, Kinra S, Bharathi AV, et al. Dietary intake and rural-urban migration in India: a cross-sectional study. PloS one. 2011;6(6):e14822
15. Allender S, Lacey B, Webster P, Rayner M, Deepa M, Scarborough P, et al. Level of urbanization and noncommunicable disease risk factors in Tamil Nadu, India. Bulletin of the World Health Organization. 2010 Apr;88(4):297-304
16. Subramanian SV, Smith GD. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. The American journal of clinical nutrition. 2006 Sep;84(3):633-40
17. Sobngwi E, Mbanya JC, Unwin NC, Porcher R, Kengne AP, Fezeu L, et al. Exposure over the life course to an urban environment and its relation with obesity, diabetes, and hypertension in rural and urban Cameroon. International journal of epidemiology. 2004 Aug;33(4):769-76
18. Allender S, Wickramasinghe K, Goldacre M, Matthews D, Katulanda P. Quantifying urbanization as a risk factor for noncommunicable disease. Journal of urban health : bulletin of the New York Academy of Medicine. 2011 Oct;88(5):906-18
19. Torun B, Stein AD, Schroeder D, Grajeda R, Conlisk A, Rodriguez M, et al. Rural-to-urban migration and cardiovascular disease risk factors in young Guatemalan adults. International journal of epidemiology. 2002 Feb;31(1):218-26
20. Poulter NR, Khaw KT, Hopwood BE, Mugambi M, Peart WS, Rose G, et al. The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. BMJ (Clinical research ed). 1990 Apr 14;300(6730):967-72
21. Al-Moosa S, Allin S, Jemiai N, Al-Lawati J, Mossialos E. Diabetes and urbanization in the Omani population: an analysis of national survey data. Population health metrics. 2006 Apr 24;4:5

22. Dahly DL, Adair LS. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. Social science & medicine (1982). 2007 12/29;64(7):1407-19
23. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.
24. Yi K, Tani H, Li Q, Zhang J, Guo M, Bao Y, et al. Mapping and evaluating the urbanization process in Northeast China using DMSP/OLS nighttime light data. Sensors (Basel, Switzerland). 2014;14(2):3207-26
25. Ma T, Zhou C, Pei T, Haynie S, Fan J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. Remote Sensing of Environment. 2012 2012/09/01/;124(Supplement C):99-107
26. Imhoff M, Lawrence W, Stutzer D, Elvidge C. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. Remote Sensing of Environment. 1997 1997/09/01/;61(3):361-70
27. Zhang Q, Seto KC. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sensing of Environment. 2011 2011/09/15/;115(9):2320-9
28. Elvidge CD, Imhoff ML, Baugh KE, Hobson VR, Nelson I, Safran J, et al. Nighttime lights of the world: 1994-95 ISPRS Journal of Photogrammetry and Remote Sensing. 2001;56:81-99
29. Zhang Q, Pandey B, Seto KC. A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data. IEEE Transactions on geoscience and remote sensisng. 2016
30. Andhra Pradesh Children and Parents Study (APCAPS). Andhra Pradesh Children and Parents Study: London School of Hygiene and Tropical Medicine; 2018 [cited 2018 12 July]. Available from: http://apcaps.lshtm.ac.uk/.
31. Kinra S, Radha Krishna K, Kuper H, Rameshwar Sarma K, Prabhakaran P, Gupta V, et al. Cohort Profile: Andhra Pradesh Children and Parents Study (APCAPS). International journal of epidemiology. 2013 Sep 9
32. Kinra S, Rameshwar Sarma KV, Ghafoorunissa, Mendu VV, Ravikumar R, Mohan V, et al. Effect of integration of supplemental nutrition with public health programmes in pregnancy and early childhood on cardiovascular risk in rural Indian adolescents: long term follow-up of Hyderabad nutrition trial. BMJ (Clinical research ed). 2008;337:a605
33. National Cholesterol Education Program, National Heart L, and Blood Institute,, National Institutes of Health. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). NIH Publication: 2002.
34. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct 08;388(10053):1659-724
35. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clinical chemistry. 1972 Jun;18(6):499-502
36. Kinra S, Johnson M, Kulkarni B, Rameshwar Sarma K, Ben-Shlomo Y, Smith G. Socio-economic position and cardiovascular risk in rural indian adolescents: evidence from the Andhra Pradesh children and parents study (APCAPS). Public health. 2014 Sep;128(9):852-9
37. Belloni A, Chernozhukov V, Hansen C. Inference on Treatment Effects after Selection among High-Dimensional Controls. The Review of Economic Studies. 2014;81(2):608-50
38. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March.
39. Poobalan A, Aucott L. Obesity Among Young Adults in Developing Countries: A Systematic Overview. Current Obesity Reports. 2016;5:2-13
40. Ebrahim S, Kinra S, Bowen L, Andersen E, Ben-Shlomo Y, Lyngdoh T, et al. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. PLoS medicine. 2010 Apr;7(4):e1000268
41. Koo YS, Song JY, Joo EY, Lee HJ, Lee E, Lee SK, et al. Outdoor artificial light at night, obesity, and sleep health: Cross-sectional analysis in the KoGES study. Chronobiology international. 2016;33(3):301-14
42. Rybnikova NA, Haim A, Portnov BA. Does artificial light-at-night exposure contribute to the worldwide obesity pandemic? International journal of obesity (2005). 2016 May;40(5):815-23

43. Lane KJ, Stokes EC, Seto KC, Thanikachalam S, Thanikachalam M, Bell ML. Associations between Greenness, Impervious Surface Area, and Nighttime Lights on Biomarkers of Vascular Aging in Chennai, India. Environmental health perspectives. 2017 Aug 2;125(8):087003
44. Attard SM, Herring AH, Mayer-Davis EJ, Popkin BM, Meigs JB, Gordon-Larsen P. Multilevel examination of diabetes in modernising China: what elements of urbanisation are most associated with diabetes? Diabetologia. 2012 Dec;55(12):3182-92
45. Ramachandran A, Mary S, Yamuna A, Murugesan N, Snehalatha C. High prevalence of diabetes and cardiovascular risk factors associated with urbanization in India. Diabetes care. 2008 May;31(5):893-8
46. He J, Klag MJ, Wu Z, Qian MC, Chen JY, Mo PS, et al. Effect of migration and related environmental changes on serum lipid levels in southwestern Chinese men. American journal of epidemiology. 1996 Nov 1;144(9):839-48
47. Mendez MA, Monteiro CA, Popkin BM. Overweight exceeds underweight among women in most developing countries. The American journal of clinical nutrition. 2005 Mar;81(3):714-21
48. Kloog I, Haim A, Stevens RG, Portnov BA. Global co-distribution of light at night (LAN) and cancers of prostate, colon, and lung in men. Chronobiology international. 2009 Jan;26(1):108-25
49. Cho Y, Ryu SH, Lee BR, Kim KH, Lee E, Choi J. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. Chronobiology international. 2015;32(9):1294-310
50. Rybnikova NA, Haim A, Portnov BA. Is prostate cancer incidence worldwide linked to artificial light at night exposures? Review of earlier findings and analysis of current trends. Archives of environmental & occupational health. 2017 Mar 4;72(2):111-22
51. Kloog I, Haim A, Stevens RG, Barchana M, Portnov BA. Light at night co-distributes with incident breast but not lung cancer in the female population of Israel. Chronobiology international. 2008 Feb;25(1):65-81
52. Hurley S, Goldberg D, Nelson D, Hertz A, Horn-Ross PL, Bernstein L, et al. Light at night and breast cancer risk among California teachers. Epidemiology (Cambridge, Mass). 2014 Sep;25(5):697-706
53. James P, Bertrand KA, Hart JE, Schernhammer ES, Tamimi RM, Laden F. Outdoor Light at Night and Breast Cancer Incidence in the Nurses' Health Study II. Environmental health perspectives. 2017 Aug 17;125(8):087010
54. Kloog I, Stevens RG, Haim A, Portnov BA. Nighttime light level co-distributes with breast cancer incidence worldwide. Cancer causes & control : CCC. 2010 Dec;21(12):2059-68
55. Bauer SE, Wagner SE, Burch J, Bayakly R, Vena JE. A case-referent study: light at night and breast cancer risk in Georgia. International Journal of Health Geographics. 2013 Apr 17;12:23
56. Rybnikova N, Haim A, Portnov BA. Artificial Light at Night (ALAN) and breast cancer incidence worldwide: A revisit of earlier findings with analysis of current trends. Chronobiology international. 2015;32(6):757-73
57. Touitou Y, Reinberg A, Touitou D. Association between light at night, melatonin secretion, sleep deprivation, and the internal clock: Health impacts and mechanisms of circadian disruption. Life sciences. 2017 Mar 15;173:94-106
58. Tacoli C. Urbainzation, gender and poverty. New York, USA: United Nations Population Fund, International Institute for Environment and Development, 2012 March.
59. de Vries S, van Dillen SM, Groenewegen PP, Spreeuwenberg P. Streetscape greenery and health: stress, social cohesion and physical activity as mediators. Soc Sci Med. 2013 Oct;94:26-33
60. Shetty PS. Nutrition transition in India. Public health nutrition. 2002 Feb;5(1A):175-82
61. Hurley S, Nelson DO, Garcia E, Gunier R, Hertz A, Reynolds P. A cross-sectional analysis of light at night, neighborhood sociodemographics and urinary 6-sulfatoxymelatonin concentrations: implications for the conduct of health studies. International Journal of Health Geographics. 2013;12:39
62. McFadden E, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. The relationship between obesity and exposure to light at night: cross-sectional analyses of over 100,000 women in the Breakthrough Generations Study. American journal of epidemiology. 2014 Aug 1;180(3):245-50
63. Gangwisch JE. Invited commentary: nighttime light exposure as a risk factor for obesity through disruption of circadian and circannual rhythms. American journal of epidemiology. 2014 Aug 1;180(3):251-3
64. Stevens RG. Light-at-night, circadian disruption and breast cancer: assessment of existing evidence. International journal of epidemiology. 2009 Aug;38(4):963-70
65. Obayashi K, Saeki K, Iwamoto J, Ikada Y, Kurumatani N. Association between light exposure at night and nighttime blood pressure in the elderly independent of nocturnal urinary melatonin excretion. Chronobiology international. 2014 Jul;31(6):779-86

66. Obayashi K, Saeki K, Iwamoto J, Okamoto N, Tomioka K, Nezu S, et al. Effect of exposure to evening light on sleep initiation in the elderly: a longitudinal analysis for repeated measurements in home settings. Chronobiology international. 2014 May;31(4):461-7
67. Ezzati M, Webster CJ, Doyle YG, Rashid S, Owusu G, Leung GM. Cities for global health. 2018;363:k3794
68. World Health Organization. Urbanization and health 2019 [cited 2019 16 July]. Available from: https://www.who.int/bulletin/volumes/88/4/10-010410/en/.
69. Vlahov D, Freudenberg N, Proietti F, Ompad D, Galea S. A conceptual framework for organizing determinants of urban health. Thematic paper 1 for Knowledge Network on Urban Settings WHO Centre for Health Development, 2008.
70. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
71. Estache A, Garsous G. The scope for an impact of infrastructure investments on jobs in developing countries International Finance Corporation, 2012.
72. Eckert S, Kohler S. Urbanization and health in developing countries: a systematic review. World health & population. 2014;15(1):7-20

Additional figures

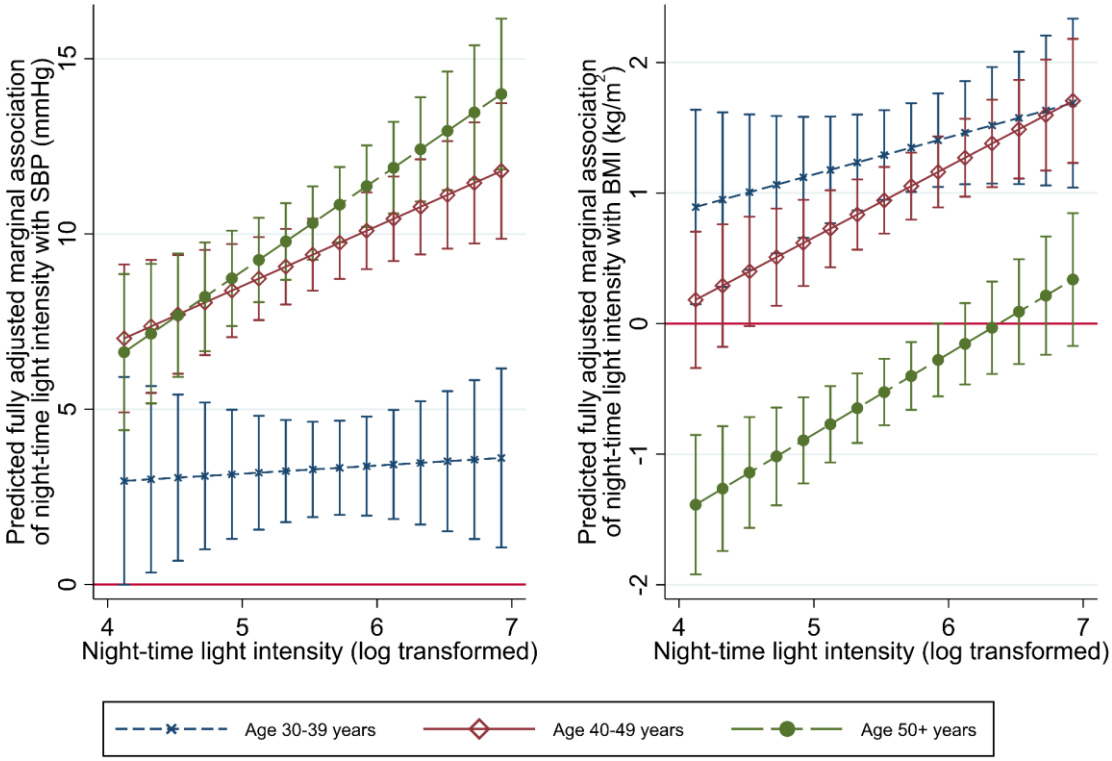


Additional Fig. 1 Age-specific crude associations NTLI with mean BMI and SBP, APCAPS, 2010-12 (n=5937)

APCAPS – Andhra Pradesh Children and Parents Study; BMI – body mass index; NTLI – night-time light intensity; SBP – systolic blood pressure

The fitted lines (model predicted means) were derived from crude multilevel linear regression models with clustering by household and village, using individual-level outcome data

Participants were excluded from analysis of SBP if medicated for hypertension

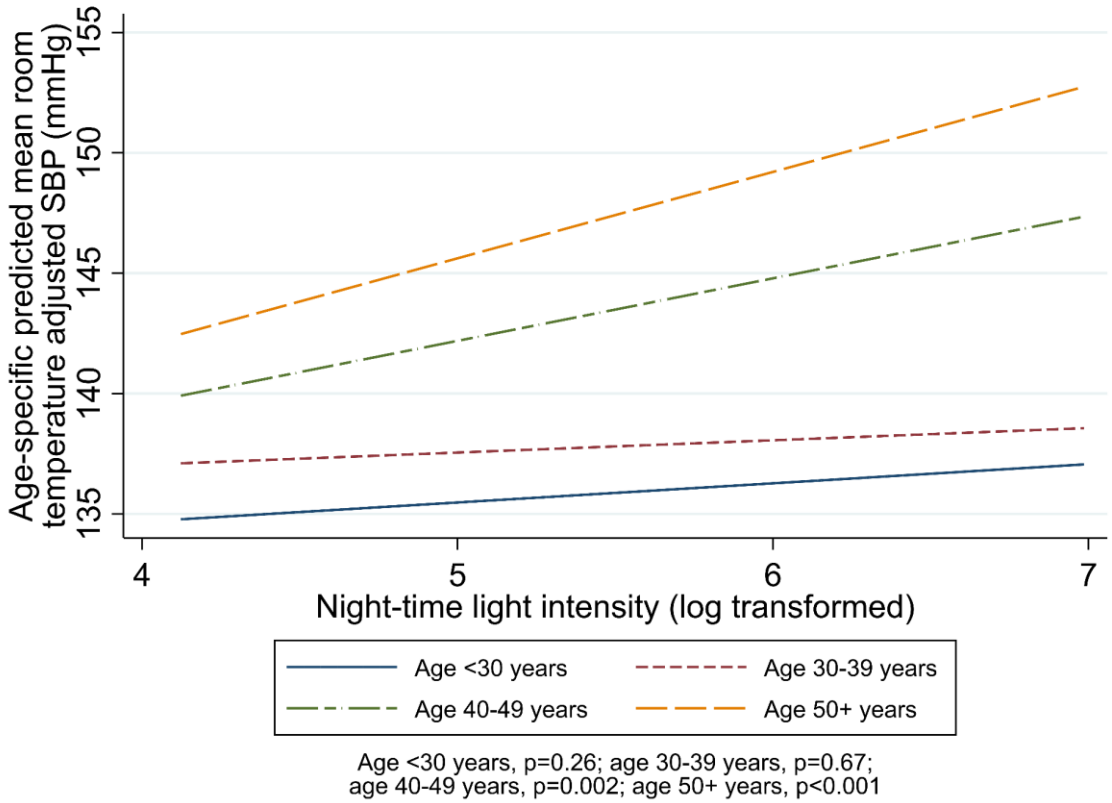


Additional Fig. 2 Fully adjusted marginal associations of NTLI with BMI and SBP by age groups (n=5937)

BMI – body mass index; CI – confidence interval; NTLI – night-time light intensity; SBP – systolic blood pressure

Marginal associations, including 95% CIs, were obtained from contrasting age-specific results from multilevel linear regression models with clustering by household and village, and adjusted for gender, caste, religion, marital status and survey season, using individual-level outcome data

Participants were excluded from analysis of SBP if medicated for hypertension

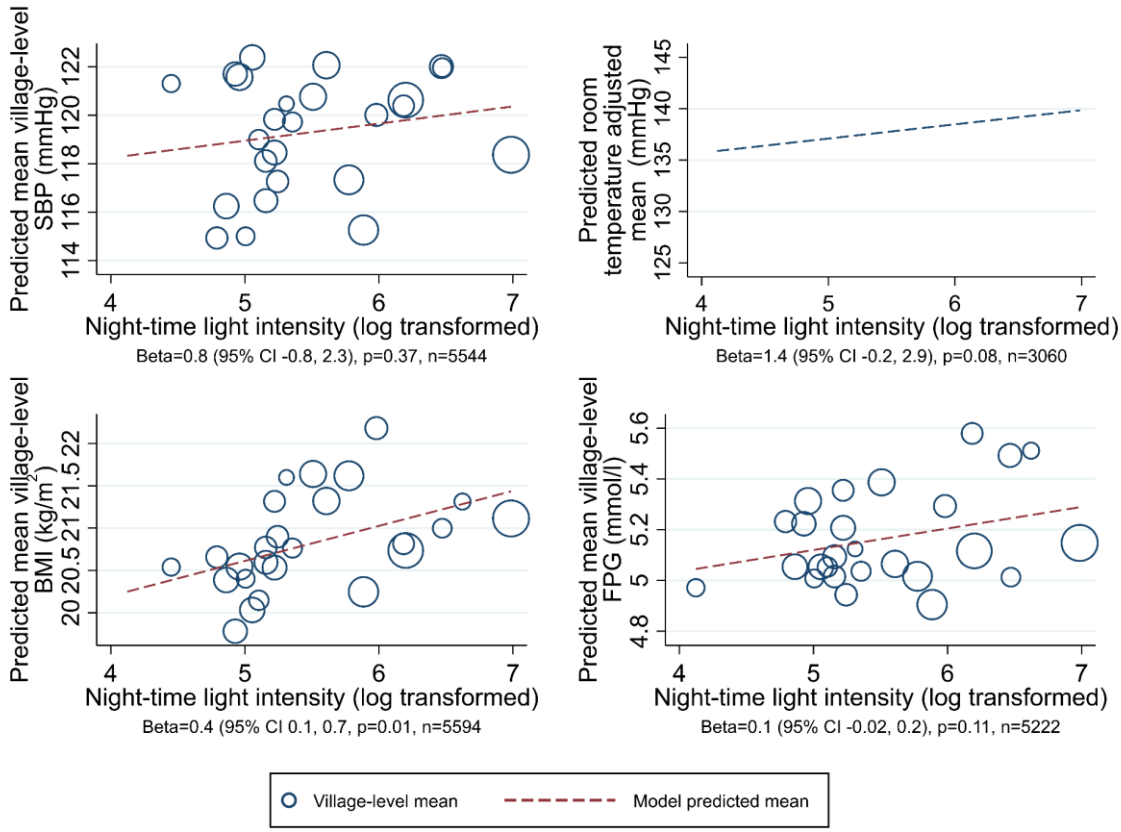


Additional Fig. 3 Age-specific room temperature adjusted association of NTLI and SBP among APCAPS adults, 2010-12 (n=3160)

APCAPS – Andhra Pradesh Children and Parents Study; NTLI – night-time light intensity; SBP – systolic blood pressure

The fitted lines (model predicted means) were derived from a crude multilevel linear regression models with clustering by household and village, using individual-level outcome data

Participants were excluded if medicated for hypertensions

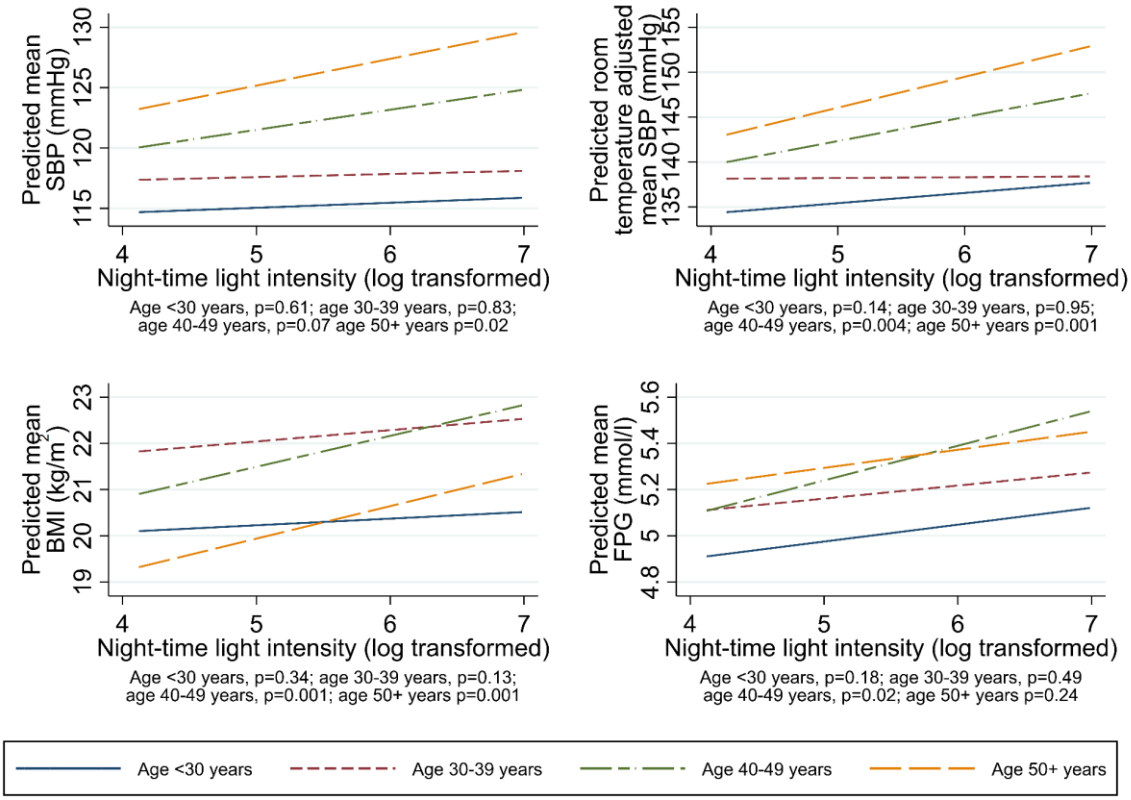


Additional Fig. 4 Sensitivity analysis: Crude association of NTLI with SBP, BMI and FPG among APCAPS adults, 2010-12

APCAPS – Andhra Pradesh Children and Parents Study; BMI – body mass index; FPG – fasting plasma glucose; NTLI – night-time light intensity; SBP – systolic blood pressure

Model predicted means were derived from crude and (for SBP) room temperature adjusted multilevel linear regression models with clustering by household and village, using individual-level outcome data

Marker size proportional to village size. Participants from borderline outlier villages were excluded.

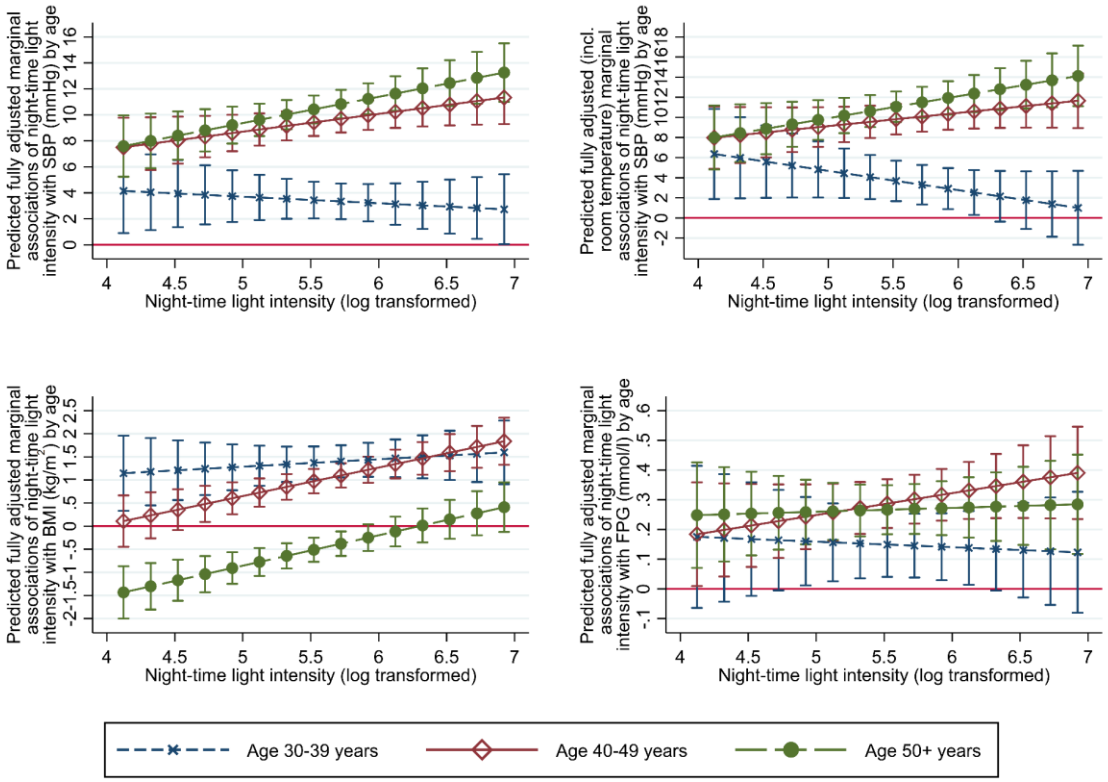


Additional Fig. 5 Sensitivity analysis: age-specific crude associations of NTLI with SBP, BMI and FPG

BMI – body mass index; FPG – fasting glucose; NTLI – night-time light intensity; SBP – systolic blood pressure

The fitted lines (model predicted means) were derived from a crude multilevel linear regression models with clustering by household and village, using individual-level outcome data

Participants from borderline outlier villages were excluded. Participants were excluded from analysis of SBP if medicated for hypertension and from analysis of FPG if medicated for diabetes.



Additional Fig. 6 Sensitivity analysis: fully adjusted marginal associations of NTLI with SBP, BMI and FPG by age groups

BMI – body mass index; CI – Confidence Interval; FPG – fasting glucose; SBP – systolic blood pressure

Marginal associations (95% CIs) were obtained from contrasting age-specific results from multilevel linear regression models with clustering by household and village, and adjusted for gender, caste, religion, marital status and survey season, using individual level outcome data

Participants from borderline outlier villages were excluded. Participants were excluded from analysis of SBP if medicated for hypertension and from analysis of glucose if medicated for diabetes

Additional tables

Additional Tab. 1 Age-specific crude and adjusted associations of NTLI with mean BMI and SBP, APCAPS, 2010-12 (n=5937)

CVD risk factors	Age group	n (%)	Age-specific crude model predicted mean (95% CI) at the lowest NTLI (61.7 (4.1 on the log scale))	Age-specific crude model predicted mean (95% CI) at the highest NTLI (1081.1 (7.0 on the log scale))	p-value	Age-specific fully adjusted model predicted mean change with increasing NTLI[b] (adjusted for gender caste, religion, marital status and survey season)	
						β (95% CI)	p-value
SBP (mmHg)[a]	<30 years	2806 (48.8)	114.2 (111.9, 116.4)	116.6 (114.2, 119.0)	0.26	0.5 (-0.7, 1.6)	0.45
	30-39 years	544 (9.5)	115.6 (112.2, 119)	119.7 (116.5, 123.0)	0.17	0.7 (-1.2, 2.5)	0.47
	40-49 years	1306 (22.7)	119.0 (116.4, 121.6)	126.1 (123.4, 128.8)	0.004	2.2 (0.8, 3.6)	0.003
	≥50 years	1096 (19.1)	121.7 (118.9, 124.4)	131.3 (128.4, 134.2)	<0.001	3.1 (1.6, 4.6)	<0.001
BMI (kg/m^2)	<30 years	2804 (47.3)	20.2 (19.6, 20.7)	20.6 (20.1, 21.2)	0.34	0.04 (-0.3, 0.4)	0.84
	30-39 years	545 (9.2)	21.7 (20.8, 22.5)	22.8 (22.0, 23.6)	0.13	0.3 (-0.2, 0.8)	0.22
	40-49 years	1358 (22.9)	21.0 (20.4, 21.7)	22.9 (22.2, 23.5)	0.001	0.6 (0.2, 1.0)	0.005
	≥50 years	1217 (20.5)	19.4 (18.8, 20.0)	21.4 (20.7, 22.1)	0.001	0.7 (0.2, 1.1)	0.002

APCAPS – Andhra Pradesh Children and Parents Study; BMI – body mass index; NTLI – night-time light intensity; SBP – systolic blood pressure

Model predicted means (95% CIs), β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression models with clustering by household and village, using individual-level outcome data

[a] Participants were excluded if medicated for hypertension. [b] Mean change per whole number increase in log transformed NTLI

Additional Tab. 2 Room temperature adjusted association of NTLI and SBP among APCAPS adults, 2010-12 (n=3160)

		Model predicted room temperature adjusted mean (95% CI) at the lowest NTLI (61.7 (4.1 on the log scale))	Model predicted room temperature adjusted mean (95% CI) at the highest NTLI (1081.1 (7.0 on the log scale))		Model predicted age, gender and room temperature adjusted mean change (95% CI) with increasing NTLI[b]			Model predicted fully adjusted mean change with increasing NTLI[b] (adjusted for gender caste, religion, marital status, survey season and room temperature)		
	n			**p-value**	**n**	**β (95% CI)**	**p-value**	**n**	**β (95% CI)**	**p-value**
SBP (mmHg)[a]	3160	136.4 (129.0, 143.8)	140.7 (133.5, 148.0)	<0.001	3159	1.9 (0.7, 3.0)	0.002	3122	1.8 (0.6, 3.0)	0.002

APCAPS – Andhra Pradesh Children and Parents Study; CI – confidence interval; n – sample size; NTLI – night-time light intensity; SBP – systolic blood pressure; β – beta-coefficient

Model predicted means (95% CIs), β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression with clustering by household and village, using individual-level outcome data

[a] Participants were excluded if medicated for hypertension. [b] Mean change per whole number increase in log transformed NTLI

Additional Tab. 3 Age-specific room temperature adjusted association of NTLI and SBP among APCAPS adults, 2010-12 (n=3160)

SBP (mmHg)[a]	n (%)	Model predicted room temperature adjusted mean (95% CI) at the lowest NTLI (61.7 (4.1 on the log scale))	Model predicted room temperature adjusted mean (95% CI) at the highest NTLI (1081.1 (7.0 on the log scale))	p-value	Model predicted fully adjusted mean change with increasing NTLI[b] (adjusted for gender caste, religion, marital status, survey season and room temperature at clinical assessment)	
					β (95% CI)	p-value
<30 years	1347 (42.6)	134.8 (128.1, 141.4)	137.1 (130.4, 143.7)	0.26	0.8 (-0.5, 2.1)	0.24
30-39 years	305 (9.7)	137.1 (129.6, 144.6)	138.6 (131.3, 145.8)	0.67	-0.2 (-2.5, 2.0)	0.83
40-49 years	838 (26.5)	139.9 (133.0, 146.8)	147.3 (140.6, 154.1)	0.002	2.4 (0.8, 4.0)	0.003
≥50 years	670 (21.2)	142.5 (135.6, 149.4)	152.7 (145.8, 159.7)	<0.001	3.5 (1.8, 5.3)	<0.001

APCAPS – Andhra Pradesh Children and Parents Study; CI – confidence interval; n – sample size; NTLI – night-time light intensity; SBP – systolic blood pressure; β – beta coefficient

Model predicted means (95% CIs), β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression with clustering by household and village, using individual-level outcome data

p<0.001 for effect modification by age group

[a] Participants were excluded if medicated for hypertension. [b] Mean change per whole number increase in log transformed NTLI

Additional Tab. 4 Sensitivity analysis: crude and adjusted associations of NTLI with SBP, BMI and FPG

	Model predicted crude mean change with increasing NTLI[c]			Model predicted age and gender adjusted mean change with increasing NTLI[c]			Model predicted fully adjusted mean change with increasing NTLI[c] (adjusted for age, gender, caste, religion, marital status and survey season)		
CVD risk factors	**n**	**β (95% CI)**	**p-value**	**n**	**β (95% CI)**	**p-value**	**n**	**β (95% CI)**	**p-value**
SBP (mmHg)[a]	5544	0.8 (-0.8, 2.3)	0.37	5539	1 .0 (-0.5, 2.5)	0.19	5306	0.8 (-0.4, 2.1)	0.18
SBP (mmHg), room temperature adjusted[a]	3060	1.4 (-0.2, 2.9)	0.08	3059	1.8 (0.5, 3.1)	0.01	3022	1.7 (0.4, 3.0)	0.01
BMI (kg/m^2)	5594	0.4 (0.1, 0.7)	0.01	5590	0.4 (0.1, 0.7)	0.01	5352	0.3 (0.03, 0.7)	0.03
Fasting glucose (mmol/l)[b]	5222	0.1 (-0.02, 0.2)	0.11	5220	0.1 (-0.01, 0.2)	0.08	4988	0.1 (-0.06, 0.2)	0.37

BMI – body mass index; CI – confidence interval; LDL – low-density lipoprotein; n – sample size; NTLI – night-time light intensity; SBP – systolic blood pressure; β – beta coefficient

β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression with clustering by household and village, using individual-level outcome data

Participants from borderline outlier villages were excluded. Participants were excluded if medicated for hypertension[a] or diabetes[b]. [c] Mean change per whole number increase in log transformed NTLI

Additional Tab. 5 Sensitivity analysis: age-specific crude and adjusted associations of NTLI with SBP, BMI and FPG

CVD risk factors		n	Age-specific model predicted mean (95% CI) at lowest NTLI	Age-specific model predicted mean (95% CI) at highest NTLI	p-value	Age-specific fully adjusted model predicted mean change with increasing NTLI[c] (adjusted for gender caste, religion, marital status and survey season)		
						n	β (95% CI)	p-value
SBP (mmHg)[a]	<30 years	2706	114.8 (112.8, 116.8)	115.9 (113.3, 118.4)	0.61	2615	0.2 (-1.1, 1.5)	0.79
	30-39 years	520	117.4 (114.4, 120.5)	118.1 (114.6, 121.5)	0.83	494	-0.3 (-2.4, 1.7)	0.76
	40-49 years	1252	120.6 (118.3, 122.9)	124.8 (121.9, 127.7)	0.07	1196	1.5 (0.02, 3.1)	0.05
	≥50 years	1066	123.9 (121.6, 126.3)	129.6 (126.5, 132.6)	0.02	1001	2.2 (0.5, 3.8)	0.009
SBP (mmHg)[a]	<30 years	1308	134.8 (127.9, 141.7)	137.7 (130.3, 145.1)	0.14	1294	1.0 (-0.5, 2.5)	0.20
adjusted for room	30-39 years	292	138.2 (130.6, 145.7)	138.4 (130.4, 146.4)	0.95	285	-0.8 (-3.4, 1.8)	0.54
temperature at	40-49 years	809	140.9 (133.8, 148.0)	147.6 (140.1, 155.2)	0.004	799	2.3 (0.6, 4.1)	0.01
clinical assessment	≥50 years	651	144.2 (137.1, 151.2)	152.9 (145.2, 160.6)	0.001	644	3.2 (1.3, 5.1)	0.001
BMI (kg/m2)	<30 years	2653	20.1 (19.7, 20.6)	20.5 (20.0, 21.1)	0.42	2563	0.02 (-0.3, 0.4)	0.92
	30-39 years	508	21.9 (21.2, 22.6)	22.5 (21.7, 23.3)	0.37	482	0.2 (-0.3, 0.7)	0.51
	40-49 years	1272	21.1 (20.6, 21.6)	22.8 (22.2, 23.5)	0.001	1213	0.6 (0.2, 1.0)	0.002
	≥50 years	1161	19.6 (19.0, 20.1)	21.3 (20.7, 22.0)	<0.001	1094	0.7 (0.3, 1.1)	0.001
Fasting plasma	<30 years	2537	4.9 (4.7, 5.1)	5.1 (4.9, 5.3)	0.20	2450	0.03 (-0.1, 0.2)	0.59
glucose[b]	30-39 years	491	5.1 (4.9, 5.4)	5.3 (5.0, 5.5)	0.49	465	0.01 (-0.2, 0.2)	0.88
	40-49 years	1170	5.1 (4.9, 5.3)	5.5 (5.3, 5.7)	0.02	1114	0.1 (-0.03, 0.2)	0.13
	≥50 years	1024	5.2 (5.0, 5.4)	5.4 (5.2, 5.7)	0.25	959	0.05 (-0.1, 0.2)	0.52

β – beta coefficient; BMI – body mass index; CI - confidence interval; n – sample size; NTLI – night-time light intensity; SBP – systolic blood pressure (continued next page)

Model predicted means (95% CIs), β-coefficients (95% CIs) and p-values were obtained from multilevel linear regression models with clustering by household and village, using individual-level outcome data

Participants from borderline outlier villages were excluded. Participants were excluded if medicated for hypertension[a] or diabetes[b]. [c] Mean change per whole number increase in log transformed NTLI

CHAPTER 5 QUANTIFYING THE INFLUENCE OF LOCATION OF RESIDENCE ON BLOOD PRESSURE IN URBANISING SOUTH INDIA: A PATH ANALYSIS WITH MULTIPLE MEDIATORS

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Tina Bonde Sørensen
Principal Supervisor	Professor Alan Dangour
Thesis Title	Influences of place of residence on risk factors for atherosclerotic cardiovascular diseases in South India

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Epidemiologic Methods
Please list the paper's authors in the intended authorship order:	Tina B Sørensen, Stijn Vansteelandt, Robin Wilson, John Gregson, Bhavani Shankar, Sanjay Kinra and Alan D Dangour
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	All authors contributed to study conception, revised the manuscript critically, approved the final manuscript for publication and agreed to be accountable for the work. Further contributions: Professor Stijn Vansteelandt advised on causal methodology and inference and prepared the mathematical

	appendix. Dr John Gregson provided additional advice on the statistical analysis. Professor Stijn Vansteelandt and Tina Bonde Sorensen developed the analysis strategy. Tina Bonde Sorensen and Dr John Gregson performed data management on the Andhra Pradesh Children and Parents Study data. Tina Bonde Sorensen analysed the data and produced tables and figures, interpreted analyses and drafted the manuscript. Dr Robin Wilson extracted the night-time light intensity (NTLI) data and calibrated and prepared the NTLI data for statistical analysis. Professor Alan D Dangour provided expertise on epidemiology, nutrition and health and Professor Sanjay Kinra provided expertise on epidemiology and cardiovascular diseases.

Student Signature: [redacted] **Date:** 25 May 2020

Supervisor Signatur [redacted] **Date:** 25 May 2020

Quantifying the influence of location of residence on blood pressure in urbanising South India: A path analysis with multiple mediators

Tina B Sørensen,[1*] Stijn Vansteelandt,[1, 2] Robin Wilson,[3] John Gregson,[1] Bhavani Shankar,[4, 5] Sanjay Kinra[1] and Alan D Dangour[1, 5]

*Corresponding author. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine (LSHTM), London WC1E 7HT, United Kingdom. E-mail: tina.sorensen@lshtm.ac.uk. Telephone +44 (0) 2072994723

[1]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

[2]Department of Applied Mathematics, Computer Science and Statistics, University of Ghent, Krijgslaan 281, S9 9000 Gent, Belgium

[3]Geography & Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK

[4]Centre for Development, Environment and Policy, School of Oriental and African Studies, 10 Thornhaugh Street, Russell Square, London WC1H 0XG, UK

[5]Leverhulme Centre for Integrative Research in Agriculture and Health (LCIRAH), 36 Gordon Square, London WC1H 0PD, UK

E-mails: tina.sorensen@lshtm.ac.uk, stijn.vansteelandt@ugent.be, robin@rtwilson.com, john.gregson@lshtm.ac.uk, b.shankar@soas.ac.uk, sanjay.kinra@lshtm.ac.uk, alan.dangour@lshtm.ac.uk

Abstract

Background

A large body of evidence suggests that increasing levels of urbanisation are associated with rising systolic blood pressure (SBP) in low- and middle-income countries, however the implicated pathways are poorly quantified. The current study aims to estimate the causal effect of increasing levels of urbanisation on mean SBP, and to decompose the direct and indirect effects via hypothesised mediators.

Methods

We analysed health and demographic data from 5840 adults (≥ 18years) from the Andhra Pradesh Children and Parents study (APCAPS) conducted in 27 villages in Telangana, South India. The villages experienced asymmetric urbanisation during preceding decades and ranged from a rural village to a medium sized town at the time of survey (2012). We estimated urbanisation levels of surveyed villages by combining remote sensing data of night-time light intensity (NTLI), measured by unitless digital numbers, with satellite imagery and ground surveying of village boundaries. We performed mediation analysis using linear mixed-effects models with SBP as the outcome, log-transformed continuous NTLI as the exposure, and three composite mediators summarising information on (i) socio-demographics (e.g. occupation and education); (ii) lifestyle and mental health (e.g. diet and depression); (iii) metabolic factors (e.g. fasting glucose and triglycerides). All models fitted random intercepts to account for clustering by villages and households, and adjusted for confounders.

Findings

The NTLI range across the 27 villages was 62 to 1081 (4.1 to 7.0 on the log scale). Mean SBP was 122.7mmHg (±15.7) among men and 115.8mmHg (±14.2) among women. One unit (integer) log-NTLI increase was associated with a rise in mean SBP of 2.0mmHg (95% CI 0.4, 3.5) among men and 1.3mmHg (95% CI 0.006, 2.6) among women. We identified a positive indirect effect of log-NTLI on SBP via the metabolic pathway, where one log-NTLI increase elevated SBP by 4.6mmHg (95% CI 2.0, 7.3) among men and by 0.7mmHg (95% 0.1, 1.3) among women. There was also a positive indirect effect of log-NTLI on SBP via the lifestyle and mental health pathway among men, where one log-NTLI increase elevated SBP by 0.7mmHg (95% CI 0.1, 1.3). There was no evidence of negative direct effects of log-NTLI on SBP or positive indirect effects via the socio-demographic pathway among both genders; or a positive indirect effect via the lifestyle and mental health pathway among women. The size of effects were approximately doubled among participants ≥40 years of age.

Conclusion

Our findings offer new insights into the pathways via which urbanisation level may act on blood pressure. Large indirect effects via metabolic factors, independent of socio-demographic, lifestyle and mental health factors identify a need to understand better the indirect effects of environmental cardiovascular disease (CVD) risk factors that change with urbanisation. We encourage researchers to use causal methods in further quantification of path-specific effects of place of residence on CVDs and risk factors. Available evidence-based, cost-effective interventions that target upstream determinants of CVDs, such as those outlined in the WHO Global Action Plan for the Prevention and Control of Noncommunicable Diseases, should be implemented across all socio-demographic gradients in India.

Key words: Mediation analysis; path analysis; causal inference; pathways; epidemiology; urbanisation level; blood pressure; SBP; cardiovascular disease (CVD); CVD risk factor

Introduction

Background

The hypertensive heart diseases (HHD) ischemic heart disease (IHD) and stroke are currently the leading causes of mortality in India, accounting for more than 25% of all deaths.[1] They are common in middle age and result in substantial losses of productivity and income for individuals and societies.[1-3] Systolic blood pressure (SBP) is a major modifiable risk factor for HHDs[4, 5] and evidence suggests that increasing level of urbanisation may elevate blood pressure in India and other low- and middle-income countries, particularly among men.[6-9] Location of residence may influence blood pressure through a complex web of pathways that have proved difficult to disentangle,[10, 11] and while intermediate links are often well-established,[12] the complete pathways remain poorly quantified.

Residing in an area of greater urbanisation level is hypothesised to act on blood pressure, often in opposite directions, via influences on environmental, socio-demographic, lifestyle, mental health and metabolic factors. For example, upgraded road and transport networks coupled with improved opportunities for education and employment at higher levels of urbanisation can benefit productivity and income with positive impacts on living conditions, diet, nutritional status and mental health.[8, 13, 14] At the same time, the combination of increased purchasing power, modernisation of food systems and extensive advertising at greater urbanisation levels promote nutrition transitions that are characterised by diets high in fat, sugar, salt and (convenience) foods eaten away from home.[15, 16] Nutrition transitions and reductions in physical activity during work, leisure, transport and household chores (e.g. cooking and child care) promote weight gain and influence metabolic risk factors for hypertension, such as plasma glucose and blood lipids.[6, 16-21] The interlinkages between pathways make it difficult to determine the extent to which urbanisation is beneficial or harmful for blood pressure and HHDs (Appendix I illustrates the full web of hypothesised pathways).

Improving our understanding of the pathways by which urbanisation levels may act on blood pressure might reveal cost-effective opportunities for policy makers and urban planners to mitigate expected large burdens of HHDs and related premature deaths and disability with continued urbanisation in India. In this study we used remote sensing night-time light intensity (NTLI) data to estimate urbanisation levels of 27 villages from the Andhra Pradesh Children and parents study (APCAPS), which represented a continuum of urbanisation levels. The aim of the study was to estimate the total effect of increasing level of urbanisation (measured by NTLI) on SBP, and to decompose the direct effect and indirect effects via three groups of hypothesised mediators: (i) socio-demographics, (ii) lifestyle and mental health, (iii) metabolic factors.

Methods

Population and setting

The current cross-sectional mediation analysis included data from the third wave of the APCAPS, an intergenerational cohort study of CVD determinants in Telangana, South India.[22] The study was conducted in 28 villages (originally 29 villages, but two merged over time) near Hyderabad in the Ranga Reddy district of Telangana (formerly Andhra Pradesh). The villages have experienced asymmetric urbanisation during the decades preceding the third data wave and this offered a good opportunity to study blood pressure across a continuum of urbanisation levels, ranging from a rural village to a medium sized town. Further details of the APCAPS cohort are published elsewhere.[23, 24] Data were available from 6944 female and male participants. Participants were excluded from the current analysis if they were pregnant (n=80), younger than 18 years of age (n=621), missing most data (n=7) or missing data on gender (n=6) (Figure 1). Eligible participants residing in one large village with considerably higher NTLI than other villages were excluded from analysis to avoid overreliance on this outlier village (n=297). Participants for whom derived (and imputed) LDL levels were invalid due to high serum triglyceride levels (>400 mg/dl)[25] were also excluded (n=93).

Exposure

There are no standard definitions of 'urban', 'urbanisation' or 'urbanisation levels'. Most of the evidence linking urbanisation levels with SBP and mediators stems from rural-urban studies that potentially compare inherently different populations. A growing number of studies quantify urbanisation levels from multi-component scales that combine multi-level data relating to e.g. built environment, services, housing types and densities, economic activities and education levels. These scales are likely more informative for studying health outcomes than rural-urban comparisons and single component classification systems such as population density.[26, 27] However, there is a lack of consensus on which components best capture urbanisation levels, which has led to calls for identifying improved standardised measures of urbanisation levels.[28, 29] Remote sensing data has long been of interest to health researchers. A growing evidence base supports the utility and validity of remotely sensed NTLI as a globally consistent proxy indicator of urbanisation levels and processes.[30-32] In the current analysis we used NTLI measured in 2011 (due to technical issues with satellite sensors in 2012[33]) as a proxy for urbanisation levels of APCAPS villages. The village boundaries were traced using Open Street Map software and GPS-based ground surveying in 2012-13. Urbanisation levels were estimated by overlaying the village boundaries with NTLI pixels obtained from the United States' Defence Meteorological Satellite Programme's Operational Linescan System (DMSP-OLS).

The NTLI data is published annually as 30 arc seconds pixels[34] containing unitless digital number ranging from 0 (no light) to 63 (light saturation).[30, 31] To obtain one NTLI value for villages that were intersected by two or more pixels, a super resolution method was used to resample the NTLI images to a higher resolution, before summing the pixels intersecting a village. We calibrated the data with the ridgeline sampling regression method[33] and removed any oversaturation of light.

Health and socio-demographic data

Trained interviewers collected socio-demographic, lifestyle and health information using semi-structured questionnaires in Telugu. Information were collected on date of birth, gender, education, occupation, caste, religion, marital status, household size, assets, living conditions, and own and parental history of obesity, diabetes, hypertension, stroke and CVD. Socio-economic status (SES) was estimated (by RP) using principal component analysis of 19 assets identified by the Standard of Living Index scale, a household level asset-based scale devised for use in Indian surveys.[35, 36]

Diet and physical activity

Questionnaires on diet and physical activity were adapted and validated for use in South India.[37-40] Semi-quantitative food frequency questionnaires with one-year recall covered average portion sizes and frequency of consumption of 98 common foods and drinks. The nutrient contents of consumed foods and drinks were calculated from Indian food composition tables, supplemented by international databases.[41, 42] A seasonality fraction was applied to seasonal foods, accounting for the number of months per year a food was in season.[39] Physical activity data were obtained from the survey, where average frequency and time spent in common daily activities in the week preceding the interview were ascertained. The data were converted to metabolic equivalent tasks (METs), or physical activity ratios (PARs) when METs were not available. When METs or PARs were not available from the Indian setting, international estimates were used.[43, 44] Physical activity levels (PAL) were calculated as the sum of daily METs or PARs divided by 24 hours. If the sum of average daily activities exceeded 24 hours, the value was reduced proportionally to the amount overreported.[40] If the average daily activities accounted for less than 24 hours, a standard value of 1.4 was applied to the residual time.[40] Adjustments were made for natural breaks during strenuous activities using the Integrated Energy Index.[45, 46]

Clinical measures

Medical doctors and field staff carried our clinical assessments in local clinics and at the National Institute of Nutrition in Hyderabad. Weight was measured twice to the nearest 0.1 kg

wearing minimal clothing and no shoes (digital Seca Leicester 899; Chasmors Ltd, London, United Kingdom (UK)). Standing height was measured twice to the nearest 0.1 cm (portable Seca Leicester stadiometer; Chasmors Ltd, London, UK). Waist and hip circumferences were measured twice with metallic tape measures (ADE, Germany). The two measurements of weight, height, waist and hip circumferences were averaged and BMI (weight (kg)/height $(m)^2$) and waist-hip-ratio (waist (cm)/hip (cm)) were calculated for analysis. Blood pressure was measured trice at the right arm while seated (digital Omron hem 7300; Matsusaka Co., Japan). An average of the last two measures were used for analysis. Depression was diagnosed using the Telugu version of the nine item depression module of The Brief Patient Health Questionnaire,[47] validated for use in South India.[48, 49] Due to a limited variation in depression severity, a binary measure was used in the analysis (none vs. mild, moderate or severe depression).

Biochemical assays

Fasting venous blood samples were separated within 30 minutes of collection, stored locally at -20°C, and analysed with the Cobas311 auto-analyser and reagents from Roche Diagnostics GmbH, Mannheim, Germany and Randox Laboratories, London, UK. Biochemical assays were performed at the Genetics and Biochemistry Laboratory at The South Asia Network for Chronic Diseases of the Public Health Foundation of India, New Delhi (Table 1).

Quality assurance

Trained field staff and medical doctors followed standard protocols for data collection. Anthropometric equipment were calibrated daily. One to three weeks into the study, anthropometric measurements, blood pressure and biochemical assays were reproduced in a 5% random sample with high consistency for all measures (intra-class correlation coefficients 0.85 to >0.98).[50] The Cardiac Biochemistry Laboratory at the All India Institute of Medical Sciences monitored the quality of biochemical assays.

Statistical methods

Numerical and visual examinations of data distributions guided data cleaning (12 outliers were set to missing). We performed mediation analyses under linear mixed-effects models (with random intercepts) for continuous SBP and composite mediators, allowing for clustering by sampling units (village and household). The exposure, continuous NTLI, had a skewed distribution and was log-transformed for analysis. We focused the analysis to a limited number of pathways to avoid the more restrictive assumptions of traditional linear structural equation models.[51, 52] In brief, our approach combined individual mediators into three 'conceptual groups': (i) socio-demographic factors (M1), (ii) lifestyle and mental health factors (M2), and

(iii) metabolic factors (M3) (Box 1). In the first step we built a fully conditional model for SBP, including all individual mediators and potential confounders of exposure-mediator, mediator-mediator, mediator-outcome, and exposure-outcome relationships. Hypothesised mediators and potential confounders were identified *a priori* from the urbanisation, CVD and CVD risk factor literature. We expanded the double selection method[53] for identification of covariates (confounders, effect modifiers and mediators) to consider covariate-mediator relationships in addition to covariate-exposure and covariate-outcome relationships. A covariate was retained in the model if it was associated at the 5% level with either exposure, outcome or mediator(s). All *a priori* identified mediators were retained. Collinearity of confounders (including confounding by mediators) with exposure and mediators was assessed by examining the standard errors of exposure and mediators in a model of exposure, mediators and outcome with and without the confounder. A considerable increase in standard error was assumed to be suggestive of collinearity. None of the confounders were excluded as a result of collinearity. The list of confounders, that were not also mediators, included in the final causal models were: Age (10 year bands); marital status (unmarried, married); caste (General caste, Schedule caste, Schedule tribe, other backward class, other); religion (Christian, Hindu, Muslim, Other); season of survey and season of birth (summer (March-June), South West monsoon (July-September), winter (October-February)); maternal and paternal history of obesity, diabetes, hypertension, stroke and coronary heart disease (CHD) (yes, no, don't know); and room temperature (^{0}C) at blood pressure measurement. Interaction terms were included between NTLI or age with relevant mediators; and age with NTLI when there was evidence at the 5% level of their importance. In the second step, we aggregated the individual socio-demographic; lifestyle and mental health; and metabolic factors into three composite scalar mediator summaries. These were defined as the parts of the linear predictor from the outcome model related to these factors. After hypothesising unidirectional relationships between the resulting three composite mediators, the product of coefficients method was used to decompose the total effect of increasing urbanisation levels (log-transformed NTLI) on mean SBP into four causal pathways. One direct effect (X → Y acting around all included mediators) and three indirect (mediated) effects described below and illustrated in Figure 2.

1. Socio-demographic factors: X → M1 → Y, X → M1 → M2 → Y, X → M1 → M3 → Y and X → M1 → M2 →M3 → Y
2. Lifestyle and mental health factors: X → M2 → Y and X → M2 → M3 → Y
3. Metabolic factors: X → M3 → Y
4.

Figure 3 illustrates the Directed Acyclic Graph of the expected total, direct and indirect effects of urbanisation level on SBP. Further (mathematical) details on the analysis strategy and the calculation of bootstrap-based confidence intervals are provided in Appendix II.

Study characteristics and descriptive statistics

Study characteristics, including the distribution of mediators, in the observed data were presented by thirds of ranked NTLI, referred to as low, medium, and high urbanisation levels. Means with standard deviations (sd) described normally distributed data, medians with interquartile ranges (IQR) described skewed data, and numbers with percentages described categorical data. Associations between mediators and SBP were estimated from multivariable linear mixed-effects models adjusting for confounders and interaction between age and mediators, and with random intercepts to allow for clustering by village and household. All analyses were conducted in Stata MP v15.

Missing data

Fifty-five of 61 variables had missing data ranging from 0.1 to 9.5%, with one exception: room temperature, which was missing for 45% of participants (Appendix III). Plasma fasting glucose and fasting insulin were set to missing if participants had fasted less than eight hours at the time of giving blood ($n_{glucose}$= 351; $n_{insulin}$=351). Fasting LDL was set to missing if participants had fasted less than 9 hours (n=348). The missing-data pattern was unstructured (non-monotone) and multiple imputations with chained equations were used to avoid considerable reductions of sample sizes in multivariable models. Ten multiple imputations datasets were generated from a fully conditional model, i.e. a model including all known measured predictors of the variables with missing data. Data were imputed by strata of age and gender to allow for effect modification and were restricted to the range of cleaned observed data by predictive mean matching with five known nearest neighbours. Stata's 'mi estimate' and manual computations were used to apply Rubin's combination rule[54] when combining results from individual multiple imputations datasets into overall estimates (see Appendix IV for Stata code example).

Results

Sample and participant characteristics and descriptive statistics

Data were available from 6236 eligible men and non-pregnant women aged 18 years or older. After exclusions and removal of outliers (Figure 1), we analysed data from 5840 participants from 27 villages (Table 2). Women (47% of the sample) were on average older and more often married than men. Socio-demographic characteristics were otherwise similar among women and men. Night-time light intensity ranged from 62 to 1081 (4.1 to 7.0 on the log scale). There were

small variations in the crude distribution of participant characteristics across low, medium and high urbanisation levels among both genders.

Crude distribution of mediators across urbanisation levels

Household SES appeared to increase across urbanisation levels among men, whereas it was similar at low and medium levels and greatest at high urbanisation level among women (Table 3). Women were predominantly daily wage earners at low urbanisation level and unskilled manual workers (mainly in housework) at high urbanisation level. Men were predominantly daily wage earners at low level of urbanisation and mainly self-employed at high level of urbanisation. Agricultural work was common among both genders and at all levels of urbanisation (17-23% of men and 14-17% of women). Current use of solid fuel for cooking was common, particularly at lower levels of urbanisation, however almost all participants had been exposed to indoor open fire with solid fuels for at least six months during their life. Tobacco use was common among men at all levels of urbanisation and less common among women, particularly at high level of urbanisation. Median fasting insulin appeared to increase with rising urbanisation level, particularly among women. There were some indications of gender differences in education, alcohol consumption, shift work, depression and diet, however there were little or no indications of differences across urbanisation levels of these or the remaining mediators.

Adjusted associations of mediators with SBP

All associations between mediators and SBP were adjusted for potential confounders that were not also considered mediators in the main analysis: Age, marital status, caste, religion, season of survey, season of birth, room temperature at measurement of blood pressure, and parental history of obesity, diabetes, hypertension, stroke and CHD. Mean SBP was 115.8mmHg (sd 14.2) among women and 122.7mmHg (sd 15.7) among men. Increasing SES was associated with increasing SBP among women, with a linear trend across categories. Only high vs. low SES was associated with higher SBP among men (Table 4). Having secondary education or beyond vs. no education was estimated to decrease mean SBP by -1.9mmHg (95% CI -3.8, -0.1) among women. The pattern was in the opposite direction among men, where having primary education was estimated to increase SBP by 1.9mmHg (95% CI 0.2, 3.6) and having secondary education or beyond by 2.5mmHg (95% CI 0.7, 4.2). Unskilled manual work was associated with higher SBP than agricultural work (estimated mean increase of 2.3mmHg (95% CI 0.7, 3.9)) among women. Among men, unskilled manual work, self-employment, studying or being unemployed, and in particular salaried work were estimated to increase SBP compared to agricultural work. Increasing number of household members and current household use of solid fuel for cooking were associated with lower SBP among men, but not women. Increasing

physical activity was associated with decreased SBP, particularly among men (mean change per one MET increase -7.9mmHg (95% CI -10.6, -5.3), p<0.001). Increasing daily sleep duration was associated with greater SBP among both genders. Increasing WHR was associated with increasing SBP among both women (mean change per 1.0 increase in WHR 37.2mmHg (95% CI 29.4, 45.0)) and men (mean change per 1.0 increase in WHR 69.8mmHg (95% CI 60.6, 79)). There were evidence of associations between increasing BMI, LDL, fasting insulin, fasting glucose and serum triglyceride levels and higher SBP among both genders. Associations of CRP and dietary variables with SBP were marginal or insignificant at the 5% level. Age modified associations of SES, occupation, shift work, indoor open fire, solid fuel use, number of household members, tobacco use and sleep duration with SBP among women; and associations of occupation, sleep duration, percentage of total energy consumed from carbohydrates, vitamin B6, BMI, HDL, fasting insulin and triglycerides with SBP among men (Appendix V).

Gender-specific total, direct and indirect effects of urbanisation levels on mean SBP

There was evidence of a considerable total effect of increasing NTLI on mean SBP, where one unit (integer) log-NTLI increase was associated with a rise in mean SBP of 2.0mmHg (95% CI 0.4, 3.5) among men and 1.3mmHg (95% CI 0.01, 2.6) among women (Table 5). The total effect predicted a rise in mean SBP of 5.9mmHg (95% CI 1.8, 10.0) among men and 3.7mmHg (95% CI 0.02, 7.4) among women across the NTLI range (from rural village to medium sized town). Level of NTLI was estimated to predominantly elevate SBP via the metabolic pathway, independent of socio-demographic, lifestyle and mental health factors, among both genders. The contribution of the metabolic pathway was substantial among men, accounting for an estimated rise in mean SBP of 4.6mmHg (95% CI 2.0, 7.3) per one log-NTLI increase, and an estimated average SBP rise across the NTLI range of 13mmHg. The indirect effect on SBP via the metabolic pathway was smaller among women, accounting for an estimated rise in SBP of 0.7mmHg (95% CI 0.1, 1.3) per one log-NTLI increase and an estimated average SBP rise across the NTLI range of 2mmHg. Among men but not women, there was additionally evidence of a sizable indirect effect via the lifestyle and mental pathway contributing to an increase in SBP of 0.7mmHg (95% CI 0.1, 1.3) per one log-NTLI increase (2mmHg across the NTLI range). There was no evidence supporting the negative direct effects of NTLI on SBP or the positive indirect effects via the socio-demographic pathway that made up the remaining parts of the total effects among both genders.

Varying strengths of evidence suggested that the total effects of NTLI on SBP were greater among participants aged 40-49y (women p=0.002, men p=0.09) and 50+y (women p=0.004, men p<0.001) than in the total sample, but not among younger participants (data not shown). We collapsed the two oldest age bands in further analysis to preserve study power. Among

participants aged 40 years or older, the estimated total effects were more than double those observed in the total sample. One (integer) log-NTLI increase was associated with a rise in SBP of 4.2mmHg (95% CI 1.8, 6.7) among men and 2.8mmHg (95% CI 0.9, 4.6) among women (Table 5). The indirect effect via the metabolic pathway was also considerably greater among men in this age group, where one log-NTLI increase elevated mean SBP by 8.3mmHg (95% CI 2.5, 14.1). There was no evidence of any of the remaining path-specific effects.

Discussion

Key findings

In this large cross-sectional mediation study including 5840 participants, we used a continuum of NTLI to describe urbanisation levels of 27 villages, that had urbanised asymmetrically prior to survey. There was evidence of a considerable rise in mean SBP across the NTLI range by 5.6mmHg among men and 3.7mmHg among women. The indirect effect via the metabolic pathway accounted for a substantial part of the rise in SBP among men (13mmHg) across the NTLI range, and a smaller but considerable part of the rise in SBP among women (2mmHg) across the NTLI range. Among men but not women, there was additionally evidence of a sizable indirect effect via the lifestyle and mental health pathway accounting for a rise in SBP of 2mmHg across the NTLI range. There was no evidence to support observed positive indirect effects via the socio-demographic pathway or negative direct effects of NTLI on SBP. The total effect among both genders and the indirect effect via metabolic profiles among men approximately doubled among participants aged ≥40 years.

To the best of our knowledge, this study is the first to use causal methods to explore the relationship between urbanisation levels and SBP. Our findings corroborate a number of studies exploring associations of residing in an area of greater urbanisation level with SBP in LMIC.[9, 55-72] Our findings extend existing evidence by showing that, in villages at relatively early stages of urbanisation, even moderate increases in urbanisation levels across a continuum elevate SBP among men and women in South India. The considerable indirect effect via the metabolic pathway in our study, particularly among men, support previous reports from India of links between greater levels of urbanisation and BMI, overweight, obesity; blood lipids and dyslipidaemia; FPG and diabetes.[55, 56, 59, 66, 69] These findings support hypothesised pathways via environmental factors further upstream in the causal hierarchy that we were unable to include in analysis (Appendix I).[10, 11] Most of the potential upstream mediators we identified in the literature (e.g. market forces and advertising) are hypothesised to act via the socio-demographic or lifestyle and mental health pathways, e.g. by promoting cheap ‘unhealthy’ foods, sedentary transport and inactive leisure activities.[15, 16] Air pollution is however consistently associated

with considerable rises in blood pressure, via metabolic mechanisms, including abnormal CRP, triglycerides, (oxidised) LDL and obesity.[11, 73-77] While indoor air pollution from solid fuel use has decreased considerably in India since the 1990s, levels of outdoor air pollution have increased markedly.[78] Our analysis adjusted for indoor air pollution and the large indirect effects via metabolic factors thus support calls to better understand and quantify how blood pressure is affected by rising outdoor air and other types of pollution and environmental factors that change with urbanisation.[73-76]

Another important finding was that increasing urbanisation level elevated mean SBP via the lifestyle and mental health pathway among men, but not women. One explanation might be that men more readily adopt risk behaviours as urbanisation levels increase, e.g. tobacco use, alcohol drinking, low leisure time physical activity, and greater consumption of snacks or convenience foods outside the home. Alternatively, the indirect effect via the lifestyle and mental health pathway among men might reflect an interaction of air pollution with tobacco use.[73, 74] That is, greater levels of air pollution at greater levels of urbanisation might have magnified the rise in SBP associated with tobacco use as NTLI increased. Women would be less affected by this interaction because of their infrequent tobacco use.

A small number of studies conducted in Korea,[79] India[80] and multiple countries[81] suggest that artificial light and light pollution associated with greater levels of NTLI may promote obesity[79, 81] and elevate blood pressure[80] by supressing melatonin production, disrupting circadian rhythm and altering physiology and behaviours.[79-88] Our study was not designed to assess the effect of light *per se* and it is possible that effects of light at night on SBP are reflected in all estimated pathways, e.g. via shift work; sleep patterns or BMI.

We could not determine whether the greater total and indirect effects observed among older participants were due to biological mechanisms of aging or perhaps due to cumulative exposure to urban environments as suggested by other authors.[29, 89] One study from North India suggests that the prevalence of hypertension and diabetes, low HDL and high LDL and triglycerides increase exponentially after the age of the 30years.[90] Included age interactions with mean HDL, BMI, fasting insulin and triglycerides among men would explain some of the urbanisation-related rise in SBP in the older men (compared to men younger than 40 years) in our study. There were however no indications of age interactions with metabolic mediators among women. The doubling of the indirect effects via metabolic factors after the age of 40 years in both genders, despite gender differences in the included interaction terms of age with metabolic factors, indicate that urbanisation-related SBP rises with age have different mechanisms in women and men. It would be interesting to explore these mechanisms in future longitudinal studies.

Two counter-intuitive associations of NTLI with mediators and mediators with SBP deserve notice. Firstly, indoor use of solid fuel for cooking was associated with decreased SBP among men but not women. This association was likely confounded by strong links of type of cooking fuel with education ($p<0.001$) and SES ($p<0.001$) after adjustment for age and other confounders, which appeared to act differently on SBP among men and women. Women were also more likely to do housework than men, and thus probably had greater exposure to the poor air quality from combustion of the solid fuels that are hypothesised to increase blood pressure.[74] Secondly, among both genders, longer daily sleep duration was associated with greater SBP after adjusting for age and other confounders. This association could reflect impacts on SBP from unmeasured ill health (e.g. chronic kidney disease) or sleep disorders among participants who had longer sleep duration. We did not have data on chronic kidney disease, however we controlled for several (depression, physical inactivity, and markers of inflammation), but not all (e.g. anaemia) factors associated with fatigue among chronic kidney disease patients in the main analysis.[91] As such, we cannot rule out residual confounding from unmeasured comorbidities.

Strengths and limitations

A major strength of our study was the use of causal path analysis methods, proposed by Vanderweele and Vansteelandt,[51] under mixed-effects models, which flexibly model multiple mediators simultaneously. These methods are robust to excluding mediators acting further upstream in the causal hierarchy.[51] A great advantage is that we avoided double counting shared pathways via some mediators, which is a common problem when using simpler causal methods, e.g. 'mediation one at a time', in situations with multiple interrelated mediators.[51] Constructing the three composite mediators enabled us to handle a large number of interrelated mediators while rendering underlying model assumptions more plausible. That is, we could relax identification assumptions of 'known structural dependency' and 'unidirectional flow of effect' between certain mediators nested in a composite mediator. Other strengths of the current study include: the large population-based sample; the comprehensive set of data available from villages that had undergone asymmetric urbanisation during preceding decades; and the use of a novel standardised measure of urbanisation level, NTLI, to define a continuum of urbanisation levels. We further calibrated the NTLI data using the ridgeline sampling regression methods to reduce bias from satellite discrepancies and improve comparability of our results to those from other settings and time points.[33] Multiple imputation of missing data enabled us to include a large number of mediators with an unstructured missing data pattern. This approach further prevented the omission of some categories with very small cell sizes from a limited number of categorical variables (e.g. religion) in fully adjusted models in the observed data.

The pathways were not easily separated into testable components and our findings reflect the complex, multifaceted and at times contradictory nature of the implicated pathways. Great advances have been made in the field of causal analysis in recent years. However, limitations to handling complex interrelations of variables in situations with multiple mediators, limited the number of pathways we could test under current model assumptions. Interpreting joint indirect effects via sets of mediators summarised in the composite mediators was not straightforward in view of the strong interrelations between them. The fact that variables within each composite mediator were allowed to act in opposite directions, further added complexity to the interpretation of our results. For example, the absence of evidence (at the 5% level) of an indirect effect via the socio-demographic pathway could have several interpretations. The first would suggest that increasing NTLI did not act on mean SBP via this pathway. Another interpretation that aligns better with aforementioned hypotheses, would suggest that coinciding protective and deleterious indirect effects via this pathway outweighed each other. It is also possible that we did not have sufficient power to detect small indirect effects via some pathways. Similarly, the evidence of indirect effects via the metabolic pathway among both genders and the lifestyle and mental health pathway among men may reflect different scenarios. Either the indirect effects via most or all metabolic factors consistently elevated SBP, or a larger positive indirect effect via one or some metabolic factors outweighed negative effects via others. Further methodological advances are needed to address these uncertainties in causal analysis in situations with multiple mediators. While our applied method overcame some of the methodological limitations previously highlighted, the novelty of the approach limited our ability to compare directly our findings to those of existing literature. The cross-sectional design further limited the causal interpretation.

It is important note that estimates based on the self-reported and sometimes retrospective socio-demographic, lifestyle and mental health data were more susceptible to measurement error than estimates that were based the more objective anthropometric and biochemically derived metabolic data. The lifestyle and mental health pathway might be particularly affected by this type of error due to challenges of accurately measuring diet and physical activity, especially from recall data.[92] Over reporting of perceived desirable behaviours, such as consumption of fruits and vegetables, and underreporting of perceived undesirable behaviours, such as physical inactivity and smoking, are not uncommon in epidemiological studies of chronic diseases.[92-94] It is possible that recall or reporting of risk factors for raised SBP differed systematically between participants with different (perceived) health status in our study, from which bias in positive or negative directions could have arised.[94] Measurement error could explain some of the greater imprecisions (wide CIs) of estimated indirect effects via the socio-demographic and lifestyle and mental health pathways. Most of any true effect masked by these or other data impressions

would be absorbed in the direct effect and could explain , at least to some extent, the large CIs of the estimated direct effect of urbanisation level on SBP, particularly observed among men.

Other types of risk of bias also need consideration when interpreting our results. For example, 'The white coat effect' or 'white coat hypertension' has been well documented for measurement of blood pressure.[94] The white coat effect is a type of apprehension bias whereby blood pressure rises as a result of anxiety related to being in a clinical setting or being assessed.[94] Our study took several steps to reduce apprehension bias from the white coat effect on blood pressure readings by ensuring i) that participants had not undertaken the following activities during the 30 minutes preceding blood pressure measurement: strenuous exercise, eating, drinking of anything other than water, smoking or taking drugs that affect blood pressure; ii) five minutes of rest after fitting the cuff and before taking the first blood pressure reading; iii) minimum one minute between readings for blood circulation to resume; iv) averaging the last two of three blood pressure measurements for analysis. Nevertheless, we cannot rule out that the white coat effect increased blood pressure readings of some participants. We would however expect that this type of bias is randomly distributed across urbanisation levels, and as such did not bias beta coefficients, even if mean blood pressure was affected. The white coat effect could be reduced in future follow-ups by introducing self-monitoring of blood pressure if allowed by available study resources.[94] Finally, interviewers and clinicians were not blinded to exposure or outcomes of interest in our study. Lack of blinding of interviewers might result in differential probing for CVD risk factors and behaviours among participants residing in areas perceived as associated with higher CVD risk (i.e. higher urbanisation level). Although it is difficult to blind investigators in observational studies of fee living populations, this type of bias might be reduced by limiting disclosure of study aims, primary exposures and outcomes of interest, to the extent it is possible.

Attrition might also have affected our results in different ways. The APCPAS cohort was based on the Hyderabad Nutrition Trial (HNT) (1987-1990). Approximately 60% of the children born by mothers participating in the HNT (n=2601) were traceable in 2003-2005 when the APCAPS was established. However, the concurrence of child death rates among HNT children (10% in intervention villages and 13% in control villages) and Indian children under five (approximately 11.5% in 1990) provided some reassurance on the completeness of the final sample.[23] The third survey wave of the APCAPS included 1360 of the HNT children and their available parents and siblings. It was not possible to assess whether traceable individuals from the HNT differed from untraceable individuals at the establishment of the APCAPS (and thus assess the samples' representativeness of the original source population), due to a lack of a reliable identifier for linking HNT participants to data collected during the HNT.[23] The proportion of attrition across the three APCAPS follow-ups did not appear to differ by HNT trial status but we cannot rule

out that differences existed according to other characteristics, including exposure and outcome status. Differences in attrition according to sample characteristics could not be determined from the third survey wave data which was used in the current analysis. Initial exploration of the third survey wave data suggested that participants with incomplete data for major CVD risk factors (SBP, BMI, LDL and FPG) differed from participants with complete data. We used multiple imputations for missing values to retain study participants with incomplete data, and thus improve representativeness of the sample. However, this approach did not overcome potential attrition bias form any differential loss to follow up over time.

The use of remote sensing data also had limitations. NTLI is strongly associated with indicators of urbanisation levels, such as population density, built-up area, economic activity and energy consumption,[30-32, 95] however, it may capture other aspects of urbanisation levels with less accuracy. Using village-level NTLI could further mask 'urbanisation of rural lives' resulting from time spent away from one's area of residence, e.g. dietary changes influenced by access to new commodities available elsewhere.[96, 97] This can be adjusted for in future studies by e.g. including measures of distance and mobility between settlements. Other limitations may also have affected our results. There may be an increased risk of model-misspecification due to the inclusion of a large number of mediators in composite mediators. We can also not rule out some residual mediation or confounding not captured by the included variables. For example, early life nutrition and growth; which is linked to human capital, glucose and lipid metabolism, obesity, diabetes and hypertension later in life.[16, 98] Finally, attrition across APCAPS data waves, e.g. from migration or death, could have introduced selection bias and reduced generalisability of the results.

Implications for further research and policy

The current study was an ambitious first step to formally disentangle and quantify how increasing urbanisation levels (measured by a NTLI continuum) may drive rises in mean SBP during initial stages of urbanisation in South India. We hope that this study will encourage researchers to use causal methods in future explorations of urbanisation related health and disease. Future studies would benefit from including environmental risk factors that change with urbanisation and act on SBP independently of socio-demographic, lifestyle and mental health factors. There is a particular need for longitudinal causal studies to more firmly establish the flow of effect via closely related mediators and evaluate the potential incremental effect of rate of urbanisation and cumulative effects of location of residence. It will also be important to determine the drivers of age and gender differences observed in our study, particularly in the light of aging populations.

A large systematic review and meta-analysis of 26 drug trials (152,290 participants wold-wide showed that a reduction in SBP by 5mmHg could reduce the risk of major CVD events by 17%.[99] Another large international systematic review and meta-analysis of 42 randomised controlled trial (144,220 individuals) reported linear associations between reduced mean SBP and decreased risk of major CVD events (CHD, stroke, heart failure, CVD mortality), stroke, CHD, CVD mortality and all-cause mortality.[100] This review highlighted the potential far reaching health benefits of SBP levels below current recommended targets. For example, a reduction in mean SBP from 130-334mmHg to 120-124mmHg was associated with a 29% reduction in major CVD and a 27% reduction in all-cause mortality.[100] Targeting the upstream determinants of the large indirect effects on SBP via metabolic factors in our study (as high as 13mmHg across the urbanisation continuum), could potentially prevent a large proportion of expected CVD events with continued urbanisation in South India as well as provide additional benefit for all-cause mortality. Across socio-demographic gradients and levels of urbanisation in India, it is therefore recommended to implement available evidence-based and cost-effective interventions[101, 102] that target environmental determinants of lifestyle, mental health and metabolic CVD risk factors, including intervention outside of the health sector.

Declarations

Ethics

We received ethical approvals from Public Health Foundation of India, New Delhi, India; National Institute of Nutrition, Hyderabad, India; and the London School of Hygiene and Tropical Medicine (LSHTM), London, UK. Approvals were also obtained from all village heads and their committees. Study participants provided written informed consent or a witnessed thumbprint if illiterate prior to study start.

Author contributions

All authors contributed to study conception, revised the manuscript critically, approved the final manuscript for publication and agreed to be accountable for the work. Professor Stijn Vansteelandt advised on causal methodology and inference and prepared the mathematical appendix. Dr John Gregson provided additional advice on the statistical analysis. Professor Stijn Vansteelandt and Tina Bonde Sorensen developed the analysis strategy. Dr Robin Wilson extracted the night-time light intensity (NTLI) data and calibrated and prepared the NTLI data for statistical analysis. Tina Bonde Sorensen and Dr John Gregson performed data management of the Andhra Pradesh Children and Parents Study data. Tina Bonde Sorensen analysed the data and produced tables and figures, interpreted analyses and drafted the manuscript. Professor Alan

D Dangour provided expertise on epidemiology, nutrition and health and Professor Sanjay Kinra provided expertise on epidemiology and cardiovascular diseases.

Conflict of interest

The authors declare that they have no competing interests

Acknowledgements

We would like to thank the APCAPS team and support staff. We would like to acknowledge the contributions of Rashmi Pant who constructed the socio-economic status variable; Liza Bowen, Poppy Mallinson and Lukasz Aleksandrowicz who contributed to deriving nutrition and other variables. We extend our gratitude to statisticians Karla Diaz-Ordaz for advising on multiple imputations and Daniel Altmann for advising on matrix expressions. Also a big thanks to Stephen Nash and Ruth Farmer for helpful discussions on statistics and epidemiology as well as manuscript reviews. This work was supported by Wellcome Trust (www.wellcome.ac.uk/) Grant number 083707 and the Bloomsbury PhD Studentship (http://www.bloomsbury.ac.uk/) to the corresponding author through LSHTM, London, UK.

Data availability statement

The data used for this analysis are available from http://apcaps.lshtm.ac.uk[22] upon request.

Abbreviations

APCAPS	Andhra Pradesh Children and Parents Study
BMI	Body mass index
C	Set of confounders
CHD	Coronary heart disease
CI	Confidence interval
CRP	C-reactive protein
CVD	Cardiovascular disease
d	Day
DMSP-OLS	Defence Meteorological Satellite Programme's Operational Linescan System
FPG	Fasting plasma glucose

H	Hour(s)
HDL	High-density lipoprotein
HHD	Hypertensive heart disease
HNT	Hyderabad Nutrition Trial
IHD	Ischemic heart disease
IQR	Inter quartile range
LDL	low-density lipoprotein
LMIC	low- and middle-income country
LSHTM	London School of Hygiene and Tropical Medicine
M	Composite mediators
MET	Metabolic equivalent tasks
n	Sample size
NTLI	Night-time light intensity
PAL	Physical activity level
PAR	Physical activity ratio
SBP	Systolic blood pressure
sd	Standard deviation
SES	Socio-economic status
UK	United Kingdom
y	Year(s)
β	Beta-coefficient

References

1. Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart (British Cardiac Society). 2008 Jan;94(1):16-26
2. Sharma M, Ganguly NK. Premature coronary artery disease in Indians and its associated risk factors. Vascular health and risk management. 2005;1(3):217-25

3. Prabhakaran D, Singh K. Premature coronary heart disease risk factors & reducing the CHD burden in India. The Indian journal of medical research. 2011 Jul;134:8-9
4. O'Donnell MJ, Chin SL, Rangarajan S, Xavier D, Liu L, Zhang H, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet. 2016 Aug 20;388(10046):761-75
5. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004 Sep 11-17;364(9438):937-52
6. Ibrahim MM, Damasceno A. Hypertension in developing countries. Lancet. 2012 Aug 11;380(9841):611-9
7. Gupta R, Gaur K, Mohan I, Khedar RS. Urbanization, Human Development and Literacy and Syndemics of Obesity, Hypertension and Hyperglycemia in Rajasthan: National Family Health Survey-4. The Journal of the Association of Physicians of India. 2018 Dec;66(12):20-6
8. Eckert S, Kohler S. Urbanization and health in developing countries: a systematic review. World health & population. 2014;15(1):7-20
9. Zhao J, Seubsman SA, Sleigh A, Thai Cohort Study Team T. Timing of urbanisation and cardiovascular risks in Thailand: evidence from 51 936 members of the thai cohort study, 2005-2009. Journal of epidemiology. 2014;24(6):484-93
10. Godfrey R, Julien M. Urbanisation and health. Clin Med (Lond). 2005 Mar-Apr;5(2):137-41
11. Moore M, Gould P, Keary BS. Global urbanization and impact on health. International journal of hygiene and environmental health. 2003 2003/01/01/;206(4):269-78
12. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.
13. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
14. Estache A, Garsous G. The scope for an impact of infrastructure investments on jobs in developing countries International Finance Corporation, 2012.
15. Goryakin Y, Suhrcke M. Economic development, urbanization, technological change and overweight: what do we learn from 244 Demographic and Health Surveys? Economics and human biology. 2014 Jul;14:109-27
16. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012 Jan;70(1):3-21
17. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March. Report No.
18. UNDP/UNFPA, UNICEF, WFP. Agenda Item 1: Population Growth and Rapid Urbanization: Food Insecurity on the Rise in Urban Settings. UNDP/UNFPA, UNICEF, WFP, 2009.
19. Popkin BM, Horton S, Kim S, Mahal A, Shuigao J. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutrition reviews. 2001 Dec;59(12):379-90
20. Sullivan R, Kinra S, Ekelund U, Bharathi AV, Vaz M, Kurpad A, et al. Socio-demographic patterning of physical activity across migrant groups in India: results from the Indian Migration Study. PloS one. 2011;6(10):e24898
21. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. The American journal of clinical nutrition. 2006;84(2):289-98
22. Andhra Pradesh Children and Parents Study (APCAPS). Andhra Pradesh Children and Parents Study: London School of Hygiene and Tropical Medicine; 2018 [cited 2018 12 July]. Available from: http://apcaps.lshtm.ac.uk/.
23. Kinra S, Radha Krishna K, Kuper H, Rameshwar Sarma K, Prabhakaran P, Gupta V, et al. Cohort Profile: Andhra Pradesh Children and Parents Study (APCAPS). International journal of epidemiology. 2013 Sep 9
24. Kinra S, Rameshwar Sarma KV, Ghafoorunissa, Mendu VV, Ravikumar R, Mohan V, et al. Effect of integration of supplemental nutrition with public health programmes in pregnancy and early childhood on cardiovascular risk in rural Indian adolescents: long term follow-up of Hyderabad nutrition trial. BMJ (Clinical research ed). 2008;337:a605
25. National Cholesterol Education Program, National Heart L, and Blood Institute,, National Institutes of Health. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). NIH Publication: 2002.

26. Dahly DL, Adair LS. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. Social science & medicine (1982). 2007 12/29;64(7):1407-19
27. Jones-Smith JC, Popkin BM. Understanding community context and adult health changes in China: development of an urbanicity scale. Soc Sci Med. 2010 Oct;71(8):1436-46
28. Cyril S, Oldroyd JC, Renzaho A. Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales. BMC public health. 2013 May 28;13:513
29. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008 Nov;85(6):938-51
30. Yi K, Tani H, Li Q, Zhang J, Guo M, Bao Y, et al. Mapping and evaluating the urbanization process in Northeast China using DMSP/OLS nighttime light data. Sensors (Basel, Switzerland). 2014;14(2):3207-26
31. Zhang Q, Seto KC. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sensing of Environment. 2011 2011/09/15/;115(9):2320-9
32. Ma T, Zhou C, Pei T, Haynie S, Fan J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. Remote Sensing of Environment. 2012 2012/09/01/;124(Supplement C):99-107
33. Zhang Q, Pandey B, Seto KC. A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data. IEEE Transactions on geoscience and remote sensisng. 2016
34. Elvidge CD, Imhoff ML, Baugh KE, Hobson VR, Nelson I, Safran J, et al. Nighttime lights of the world: 1994-95 ISPRS Journal of Photogrammetry and Remote Sensing. 2001;56:81-99
35. Subramanian SV, Nandy S, Irving M, Gordon D, Lambert H, Davey Smith G. The mortality divide in India: the differential contributions of gender, caste, and standard of living across the life course. Am J Public Health. 2006 May;96(5):818-25
36. Sciences IIfP. National Family Health Survey (NFHS-2), 1998-99. Mumbai: IIPS, 2000.
37. Sudha V, Radhika G, Sathya RM, Ganesan A, Mohan V. Reproducibility and validity of an interviewer-administered semi-quantitative food frequency questionnaire to assess dietary intake of urban adults in southern India. International journal of food sciences and nutrition. 2006 Nov-Dec;57(7-8):481-93
38. Bansal D, Satija A, Khandpur N, Bowen L, Kinra S, Prabhakaran D, et al. Effects of migration on food consumption patterns in a sample of Indian factory workers and their families. Public health nutrition. 2010 Dec;13(12):1982-9
39. Bowen L, Ebrahim S, De Stavola B, Ness A, Kinra S, Bharathi AV, et al. Dietary intake and rural-urban migration in India: a cross-sectional study. PloS one. 2011;6(6):e14822
40. Sullivan R, Kinra S, Ekelund U, Bharathi AV, Vaz M, Kurpad A, et al. Evaluation of the Indian Migration Study Physical Activity Questionnaire (IMS-PAQ): a cross-sectional study. The international journal of behavioral nutrition and physical activity. 2012;9:13
41. United States Department of Agriculture. USDA Food Composition Databases 2017 [cited 2017 7 December]. Available from: https://ndb.nal.usda.gov/ndb/.
42. Welch AAUI, Buss DH, Paul AA, Southgate DAT. McCane and Widdowson's The Composition of Foods. Cambridge: Royal Society of Chemistry1995.
43. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Medicine and science in sports and exercise. 2011 Aug;43(8):1575-81
44. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett JDR, C. T-L, et al. Compendium of physical activity 2011 [cited 2014 22 August]. Available from: https://sites.google.com/site/compendiumofphysicalactivities/.
45. James WPT, Schofield EC. Human Energy Requirements. A manual for planners and nutritionists. Oxford: Oxford University Press; 1990.
46. FAO/WHO/UNU Expert Consultation. Human energy requirements. Rome: Food and Agriculture Organization of the United Nations (FAO), 2004.
47. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA : the journal of the American Medical Association. 1999 Nov 10;282(18):1737-44
48. Kochhar PH, Rajadhyaksha SS, Suvarna VR. Translation and validation of brief patient health questionnaire against DSM IV as a tool to diagnose major depressive disorder in Indian patients. Journal of postgraduate medicine. 2007 Apr-Jun;53(2):102-7
49. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine. 2001 Sep;16(9):606-13

50. Kinra S, Johnson M, Kulkarni B, Rameshwar Sarma K, Ben-Shlomo Y, Smith G. Socio-economic position and cardiovascular risk in rural indian adolescents: evidence from the Andhra Pradesh children and parents study (APCAPS). Public health. 2014 Sep;128(9):852-9
51. VanderWeele T, Vansteelandt S. Mediation Analysis with Multiple Mediators. Epidemiologic methods. 2014 Jan;2(1):95-115
52. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015 Mar;71(1):1-14
53. Belloni A, Chernozhukov V, Hansen C. Inference on Treatment Effects after Selection among High-Dimensional Controls. The Review of Economic Studies. 2014;81(2):608-50
54. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons; 2004
55. Kinra S, Andersen E, Ben-Shlomo Y, Bowen L, Lyngdoh T, Prabhakaran D, et al. Association between urban life-years and cardiometabolic risk: the Indian migration study. American journal of epidemiology. 2011 Jul 15;174(2):154-64
56. Subramanian SV, Smith GD. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. The American journal of clinical nutrition. 2006 Sep;84(3):633-40
57. Poulter NR, Khaw KT, Hopwood BE, Mugambi M, Peart WS, Rose G, et al. The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. BMJ (Clinical research ed). 1990 Apr 14;300(6730):967-72
58. Al-Moosa S, Allin S, Jemiai N, Al-Lawati J, Mossialos E. Diabetes and urbanization in the Omani population: an analysis of national survey data. Population health metrics. 2006 Apr 24;4:5
59. Ebrahim S, Kinra S, Bowen L, Andersen E, Ben-Shlomo Y, Lyngdoh T, et al. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. PLoS medicine. 2010 Apr;7(4):e1000268
60. He J, Klag MJ, Wu Z, Qian MC, Chen JY, Mo PS, et al. Effect of migration and related environmental changes on serum lipid levels in southwestern Chinese men. American journal of epidemiology. 1996 Nov 1;144(9):839-48
61. He J, Klag MJ, Whelton PK, Chen JY, Mo JP, Qian MC, et al. Migration, blood pressure pattern, and hypertension: the Yi Migrant Study. American journal of epidemiology. 1991 Nov 15;134(10):1085-101
62. Wang C, Wei D, Wang B, Zhang J, Zhang K, Ma M, et al. Effect of lifestyle on the prevalence of the metabolic syndrome among farmers, migrants with Yi ethnicity and Han population in Sichuan province of China. Asia Pacific journal of clinical nutrition. 2010;19(2):266-73
63. Gregory CO, Dai J, Ramirez-Zea M, Stein AD. Occupation is more important than rural or urban residence in explaining the prevalence of metabolic and cardiovascular disease risk in Guatemalan adults. The Journal of nutrition. 2007 May;137(5):1314-9
64. Norboo T, Stobdan T, Tsering N, Angchuk N, Tsering P, Ahmed I, et al. Prevalence of hypertension at high altitude: Cross-sectional survey in Ladakh, Northern India 2007-2011. BMJ Open. 2015;5 (4) (no pagination)(e007026)
65. Patel JV, Vyas A, Cruickshank JK, Prabhakaran D, Hughes E, Reddy KS, et al. Impact of migration on coronary heart disease risk factors: comparison of Gujaratis in Britain and their contemporaries in villages of origin in India. Atherosclerosis. 2006 Apr;185(2):297-306
66. Ramachandran A, Mary S, Yamuna A, Murugesan N, Snehalatha C. High prevalence of diabetes and cardiovascular risk factors associated with urbanization in India. Diabetes care. 2008 May;31(5):893-8
67. Sobngwi E, Mbanya JC, Unwin NC, Porcher R, Kengne AP, Fezeu L, et al. Exposure over the life course to an urban environment and its relation with obesity, diabetes, and hypertension in rural and urban Cameroon. International journal of epidemiology. 2004 Aug;33(4):769-76
68. Sodjinou R, Agueh V, Fayomi B, Delisle H. Obesity and cardio-metabolic risk factors in urban adults of Benin: relationship with socio-economic status, urbanisation, and lifestyle patterns. BMC public health. 2008 Mar 4;8:84
69. Allender S, Lacey B, Webster P, Rayner M, Deepa M, Scarborough P, et al. Level of urbanization and noncommunicable disease risk factors in Tamil Nadu, India. Bulletin of the World Health Organization. 2010 Apr;88(4):297-304
70. Attard SM, Herring AH, Zhang B, Du S, Popkin BM, Gordon-Larsen P. Associations between age, cohort, and urbanization with SBP and DBP in China: a population-based study across 18 years. Journal of hypertension. 2015 May;33(5):948-56

71. Allender S, Wickramasinghe K, Goldacre M, Matthews D, Katulanda P. Quantifying urbanization as a risk factor for noncommunicable disease. Journal of urban health : bulletin of the New York Academy of Medicine. 2011 Oct;88(5):906-18
72. Van de Poel E, O'Donnell O, Van Doorslaer E. Urbanization and the spread of diseases of affluence in China. Economics and human biology. 2009 Jul;7(2):200-16
73. Meo SA, Suraya F. Effect of environmental air pollution on cardiovascular diseases. European review for medical and pharmacological sciences. 2015 Dec;19(24):4890-7
74. Yamamoto SS, Phalkey R, Malik AA. A systematic review of air pollution as a risk factor for cardiovascular disease in South Asia: limited evidence from India and Pakistan. International journal of hygiene and environmental health. 2014 Mar;217(2-3):133-44
75. Uzoigwe JC, Prum T, Bresnahan E, Garelnabi M. The emerging role of outdoor and indoor air pollution in cardiovascular disease. N Am J Med Sci. 2013 Aug;5(8):445-53
76. Passchier-Vermeer W, Passchier WF. Noise exposure and public health. Environmental health perspectives. 2000 Mar;108 Suppl 1:123-31
77. Ezzati M, Webster CJ, Doyle YG, Rashid S, Owusu G, Leung GM. Cities for global health. 2018;363:k3794
78. India State-level Disease Burden Initiative Collaborators. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet. 2017 Nov 13
79. Koo YS, Song JY, Joo EY, Lee HJ, Lee E, Lee SK, et al. Outdoor artificial light at night, obesity, and sleep health: Cross-sectional analysis in the KoGES study. Chronobiology international. 2016;33(3):301-14
80. Lane KJ, Stokes EC, Seto KC, Thanikachalam S, Thanikachalam M, Bell ML. Associations between Greenness, Impervious Surface Area, and Nighttime Lights on Biomarkers of Vascular Aging in Chennai, India. Environmental health perspectives. 2017 Aug 2;125(8):087003
81. Rybnikova NA, Haim A, Portnov BA. Does artificial light-at-night exposure contribute to the worldwide obesity pandemic? International journal of obesity (2005). 2016 May;40(5):815-23
82. Hurley S, Nelson DO, Garcia E, Gunier R, Hertz A, Reynolds P. A cross-sectional analysis of light at night, neighborhood sociodemographics and urinary 6-sulfatoxymelatonin concentrations: implications for the conduct of health studies. International Journal of Health Geographics. 2013;12:39
83. McFadden E, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. The relationship between obesity and exposure to light at night: cross-sectional analyses of over 100,000 women in the Breakthrough Generations Study. American journal of epidemiology. 2014 Aug 1;180(3):245-50
84. Gangwisch JE. Invited commentary: nighttime light exposure as a risk factor for obesity through disruption of circadian and circannual rhythms. American journal of epidemiology. 2014 Aug 1;180(3):251-3
85. Stevens RG. Light-at-night, circadian disruption and breast cancer: assessment of existing evidence. International journal of epidemiology. 2009 Aug;38(4):963-70
86. Cho Y, Ryu SH, Lee BR, Kim KH, Lee E, Choi J. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. Chronobiology international. 2015;32(9):1294-310
87. Obayashi K, Saeki K, Iwamoto J, Ikada Y, Kurumatani N. Association between light exposure at night and nighttime blood pressure in the elderly independent of nocturnal urinary melatonin excretion. Chronobiology international. 2014 Jul;31(6):779-86
88. Obayashi K, Saeki K, Iwamoto J, Okamoto N, Tomioka K, Nezu S, et al. Effect of exposure to evening light on sleep initiation in the elderly: a longitudinal analysis for repeated measurements in home settings. Chronobiology international. 2014 May;31(4):461-7
89. Attard SM, Herring AH, Mayer-Davis EJ, Popkin BM, Meigs JB, Gordon-Larsen P. Multilevel examination of diabetes in modernising China: what elements of urbanisation are most associated with diabetes? Diabetologia. 2012 Dec;55(12):3182-92
90. Gupta R, Misra A, Vikram NK, Kondal D, Gupta SS, Agrawal A, et al. Younger age of escalation of cardiovascular risk factors in Asian Indian subjects. BMC Cardiovasc Disord. 2009 Jul 5;9:28
91. Zalai D, Bohra M. Fatigue in chronic kidney disease: Definition, assessment and treatment. Cannt j. 2016 Jan-Mar;26(1):39-44
92. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct 08;388(10053):1659-724

93. Lacombe J, Armstrong MEG, Wright FL, Foster C. The impact of physical activity and an additional behavioural risk factor on cardiovascular disease, cancer and all-cause mortality: a systematic review. BMC public health. 2019 Jul 8;19(1):900
94. Aronson J, D. B, Banerjee A, Bankhead C, Brassey J, Chalmers I, et al. Catalogue of Bias 2019 [cited 2019 16 December]. Available from: https://catalogofbias.org/.
95. Imhoff M, Lawrence W, Stutzer D, Elvidge C. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. Remote Sensing of Environment. 1997 1997/09/01/;61(3):361-70
96. Mendez MA, Monteiro CA, Popkin BM. Overweight exceeds underweight among women in most developing countries. The American journal of clinical nutrition. 2005 Mar;81(3):714-21
97. Bixby H, Bentham J, Zhou B, Di Cesare M, Paciorek CJ, Bennett JE, et al. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature. 2019 2019/05/01;569(7755):260-4
98. Hoffman DJ, Reynolds RM, Hardy DB. Developmental origins of health and disease: current knowledge and potential mechanisms. Nutrition reviews. 2017 Dec 1;75(12):951-70
99. Ninomiya T, Perkovic V, Turnbull F, Neal B, Barzi F, Cass A, et al. Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials. BMJ (Clinical research ed). 2013 Oct 3;347:f5680
100. Bundy JD, Li C, Stuchlik P, Bu X, Kelly TN, Mills KT, et al. Systolic Blood Pressure Reduction and Risk of Cardiovascular Disease and Mortality: A Systematic Review and Network Meta-analysis. JAMA Cardiology. 2017;2(7):775-81
101. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. The Lancet. 2010 2010/11/20/;376(9754):1775-84
102. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Switzerland: WHO, 2013.

Box 3 Hypothesised mediators of the effect of urbanisation levels on systolic blood pressure

M1 **Socio-demographic factors:** Occupation (agriculture, daily wage earner, unskilled manual worker, self-employed, salaried worker, student and unemployed); shift work interrupting normal sleep pattern (no, yes); education (no education, up to primary education, up to secondary education and beyond); socio-economic status (score); current use of solid fuel for cooking (yes, no); indoor open fire with wood, crop residues or dung as a primary cooking means for more than 6 months of life (no, yes); number of household members (number count)

M2 **Lifestyle and mental health factors:** Dietary energy (kcal/d), saturated fat (g/d), monounsaturated fat (g/d), polyunsaturated fat (g/d), cholesterol (mg/d), percent carbohydrates of total energy (%/d), sugar (g/d), fibre (g/d), fruits and vegetables (g/d), fish (g/d), sodium (mg/d), folate (µg/d), vitamin B6 (mg/d), vitamin B12 (µg/d), potassium (mg/d), ascorbic acid (mg/d), vitamin E (mg alpha-tocopherol equivalents/d), riboflavin (mg/d), beta-carotene (µg/d), selenium (µg/d); 24h physical activity level (PAL); tobacco use (no, yes); alcohol consumption (g/d); sleep duration (h/d); depression (none, any)

M3 **Metabolic factors:** High-density lipoprotein (mmol/l), low-density lipoprotein, fasting plasma glucose (mmol/l), fasting insulin (mmol/l), triglycerides (mmol/l), high sensitivity-C-reactive protein (mg/l)), body mass index (kg/m^2), waist-to-hip ratio

D - day; h – hour(s); M1-3 - mediator groups; PAL – physical activity level

Table 2 Biochemical assays

Analyte	Instrument	Method	Reagent source
High-sensitivity C-reactive protein	Roche c 311	Particle enhanced immuno-turbidimetric assay	ROCHE
Plasma glucose	Spectrophotometer	Enzymatic glucose oxidase /peroxidase -4-aminophenazone-phenol (GOD-PAP) method	RANDOX
Insulin	Roche e 411	Electro chemiluminescence immune assay 'ECLIA'	ROCHE
Triglycerides	Roche c 311	Enzymatic glycerol phosphate oxidase-peroxidase-4-aminophenazone-phenol (GPO-PAP)	ROCHE
Total cholesterol	Roche c 311	Enzymatic cholesterol oxidase- peroxidase-4-aminophenazone-phenol (CHOD-PAP) method	ROCHE
High-density lipoprotein (HDL)	Roche c 311	Direct method, polyethylene glycol (PEG) modified enzymes (cholesterol esterase and cholesterol oxidase) and dextran sulfate	ROCHE
Low-density lipoprotein (LDL)		Derived indirectly, the Fridewald-Fredrickson formula: LDL=total cholesterol–HDL–triglycerides/5 (with all measures in mg/dl)	

HDL - High-density lipoprotein, LDL - Low-density lipoprotein

Table 3 Participant characteristics

Socio-demographic characteristics	Urbanisation level, men (n= 3106)						Urbanisation level, women (n= 2734)					
	n	Low	n	Medium	n	High	n	Low	n	Medium	n	High
n (%)		874 (53.3)		1108 (53.5)		1124 (52.8)		766 (46.7)		965 (46.5)		1003 (47.2)
Night-time light intensity (unitless), median (IQR)	874	138 (129, 149)	1108	202 (185, 272)	1124	492 (395, 752)	766	138 (120, 149)	965	202 (185, 272)	1003	493 (485, 752)
Age, (y) median (IQR)		28.4 (23.3, 50.6)		28.7 (23.5, 50.4)		27.4 (22.9, 49.1)		40.0 (24.4, 46.0)		40.1 (24.6, 45.9)		38.0 (24.4, 45.1)
Age group (y), n (%)	874		1108		1124		766		965		1003	
<30		479 (54.8)		586 (52.9)		626 (55.7)		299 (39.0)		372 (38.6)		416 (41.5)
30/39		55 (6.3)		71 (6.4)		88 (7.8)		85 (11.1)		104 (10.8)		133 (13.3)
40/49		91 (10.4)		147 (13.3)		143 (12.7)		258 (33.7)		346 (35.9)		344 (34.3)
50+		249 (28.5)		304 (27.4)		267 (23.8)		124 (16.2)		143 (14.8)		110 (11.0)
Caste, n (%)	874		985		1124		765		853		1003	
General caste		49 (5.6)		89 (9.0)		60 (5.3)		50 (6.5)		77 (9.0)		76 (7.6)
Schedule caste		327 (37.4)		418 (42.4)		329 (29.3)		276 (36.1)		373 (43.7)		291 (29.0)
Schedule tribe		10 (1.1)		3 (0.3)		3 (0.3)		9 (1.2)		3 (0.4)		5 (0.5)
Other backward		472 (54.0)		445 (45.2)		724 (64.4)		421 (55.0)		371 (43.5)		624 (62.2)
Other		16 (1.8)		30 (3.1)		8 (0.7)		9 (1.2)		29 (3.4)		7 (0.7)
Religion, n (%)	765		853		1003		874		985		1124	
Muslim		15 (1.7)		41 (4.2)		62 (5.5)		18 (2.4)		36 (4.2)		71 (7.1)
Hindu		852 (97.5)		915 (92.9)		1053 (93.7)		737 (96.3)		780 (91.4)		914 (91.1)
Christian		7 (0.8)		29 (2.9)		8 (0.7)		9 (1.2)		37 (4.3)		16 (1.6)
Other		0 (0)		0 (0)		1 (0.1)		1 (0.1)		0 (0)		2 (0.2)
Marital status, n (%)	874		1107		1124		766		964		1003	
Married		498 (57.0)		649 (58.6)		640 (56.9)		570 (74.4)		726 (75.3)		774 (77.2)
Survey season	874		1108		1124		766		965		1003	
Summer (March-June)		222 (25.4)		350 (31.6)		241 (21.4)		205 (26.8)		321 (33.3)		242 (24.1)

Socio-demographic characteristics	Urbanisation level, men (n= 3106)						Urbanisation level, women (n= 2734)					
	n	Low	n	Medium	n	High	n	Low	n	Medium	n	High
Southwest monsoon (July-September)		225 (25.7)		395 (35.7)		830 (73.8)		211 (27.6)		343 (35.5)		715 (71.3)
Winter (October-February)		427 (48.9)		363 (32.8)		53 (4.7)		350 (45.7)		301 (31.2)		46 (4.6)
Birth season, n (%)	874		1108		1124		766		965		1003	
Summer (March-June)		99 (11.3)		137 (12.4)		132 (11.7)		53 (6.9)		47 (4.9)		42 (4.2)
Southwest monsoon (July-September)		706 (80.8)		874 (78.9)		924 (82.2)		686 (89.6)		875 (90.7)		924 (92.1)
Winter (October-February)		69 (7.9)		97 (8.8)		68 (6.1)		27 (3.5)		43 (4.5)		37 (3.7)
Mother's history of obesity, n (%)	873		1107		1124		766		963		1002	
No		840 (96.2)		1072 (96.8)		1083 (96.4)		740 (96.6)		931 (96.7)		957 (95.5)
Yes		18 (2.1)		15 (1.4)		14 (1.3)		14 (1.8)		19 (2.0)		27 (2.7)
Don`t know		15 (1.7)		20 (1.8)		27 (2.4)		12 (1.6)		13 (1.4)		18 (1.8)
Mother's history of diabetes, n (%)	873		1107		1124		766		963		1002	
No		816 (93.5)		1040 (94.0)		1038 (92.4)		718 (93.7)		918 (95.3)		926 (92.4)
Yes		36 (4.1)		46 (4.2)		43 (3.8)		35 (4.6)		31 (3.2)		57 (5.7)
Don`t know		21 (2.4)		21 (1.9)		43 (3.8)		13 (1.7)		14 (1.5)		19 (1.9)
Mother's history of hypertension, n (%)	873		1107		1124		766		963		1002	
No		749 (85.8)		980 (88.5)		931 (82.8)		645 (84.2)		848 (88.1)		833 (83.1)
Yes		101 (11.6)		106 (9.6)		153 (13.6)		108 (14.1)		101 (10.5)		149 (14.9)
Don`t know		23 (2.6)		21 (1.9)		40 (3.6)		13 (1.7)		14 (1.7)		20 (1.5)
Mother's history of stroke, n (%)	873		1107		1124		766		963		1002	
No		840 (96.2)		1057 (95.5)		1075 (95.6)		728 (95.0)		930 (96.6)		961 (95.9)
Yes		17 (2.0)		30 (2.7)		16 (1.4)		26 (3.4)		19 (2.0)		21 (2.1)
Don`t know		16 (1.8)		20 (1.8)		33 (2.9)		12 (1.6)		14 (1.5)		20 (2.0)
Mother's history of CHD, n (%)	873		1107		1124		766		963		1002	

Socio-demographic characteristics	Urbanisation level, men (n= 3106)						Urbanisation level, women (n= 2734)					
	n	Low	n	Medium	n	High	n	Low	n	Medium	n	High
No		845 (96.8)		1066 (96.3)		1077 (95.8)		738 (96.3)		929 (96.5)		968 (96.6)
Yes		12 (1.4)		20 (1.8)		14 (1.3)		15 (2.0)		19 (2.0)		15 (1.5)
Don`t know		16 (1.8)		21 (1.9)		33 (2.9)		13 (1.7)		15 (1.6)		19 (1.9)
Father's history of obesity, n (%)	873		1107		1124		766		963		1002	
No		834 (95.5)		1043 (94.2)		1080 (96.1)		717 (93.6)		925 (96.1)		935 (93.3)
Yes		23 (2.6)		22 (2.0)		8 (0.7)		22 (2.9)		15 (1.6)		35 (3.5)
Don`t know		16 (1.8)		42 (3.8)		36 (3.2)		27 (3.5)		23 (2.4)		32 (3.2)
Father's history of diabetes, n (%)	873		1107		1124		766		963		1002	
No		793 (90.8)		1002 (90.5)		1007 (89.6)		690 (90.1)		881 (91.5)		909 (90.7)
Yes		46 (5.3)		55 (5.0)		52 (4.6)		39 (5.1)		53 (5.5)		54 (5.4)
Don`t know		34 (3.9)		50 (4.5)		65 (5.8)		37 (4.8)		29 (3.0)		39 (3.9)
Father's history of hypertension, n (%)	873		1107		1124		766		963		1002	
No		747 (85.6)		963 (87.0)		938 (83.5)		647 (84.5)		863 (89.6)		856 (85.4)
Yes		97 (11.1)		97 (8.8)		126 (11.2)		84 (11.0)		71 (7.4)		105 (10.5)
Don`t know		29 (3.3)		47 (4.3)		60 (5.3)		35 (4.6)		29 (3.0)		41 (4.1)
Father's history of stroke, n (%)	873		1107		1124		766		963		1002	
No		815 (93.4)		1026 (92.7)		1037 (92.3)		709 (92.6)		901 (93.6)		943 (94.1)
Yes		35 (4.0)		35 (3.2)		40 (3.6)		27 (3.5)		37 (3.8)		26 (2.6)
don`t know		23 (2.6)		46 (4.2)		47 (4.2)		30 (3.9)		25 (2.6)		33 (3.3)
Father's history of CHD, n (%)	873		1107		1124		766		963		1002	
No		822 (94.2)		1033 (93.3)		1050 (93.4)		712 (93.0)		910 (94.5)		937 (93.5)
Yes		28 (3.2)		28 (2.5)		31 (2.8)		24 (3.1)		27 (2.8)		32 (3.2)
Don`t know		23 (2.6)		46 (4.2)		43 (3.8)		30 (3.9)		26 (2.7)		33 (3.3)

Socio-demographic characteristics	Urbanisation level, men (n= 3106)						Urbanisation level, women (n= 2734)					
	n	Low	n	Medium	n	High	n	Low	n	Medium	n	High
Hyderabad Nutrition Trial, n (%)	226		275		284		129		147		173	
Control		103 (45.6)		130 (47.3)		154 (54.2)		54 (41.9)		82 (55.8)		94 (54.3)
Intervention		123 (54.4)		145 (52.7)		130 (45.8)		75 (58.1)		65 (44.2)		79 (45.7)
Room temperature (ºC), mean (sd)	502	29.8 (3.2)	498	28.3 (3.5)	651	29.8 (3.2)	507	28.7 (3.4)	453	30.2 (3.4)	628	30.1 (2.6)

CHD - coronary heart disease; CI - confidence interval; IQR – interquartile range; n - sample size; y - years; sd - standard deviation

Table 4 Distribution of mediators across low, medium and high urbanisation levels (n=5840)

	Urbanisation level*, men (n= 3106)						Urbanisation level*, women (n= 2734)					
	n	Low	n	Medium	n	High	n	Low	n	Medium	n	High
Socio-demographic mediators												
SES, n (%)												
Low		287 (34.7)		343 (35.7)		275 (24.6)		261 (36.5)		329 (40.1)		315 (31.4)
Medium		306 (37.0)		305 (31.7)		336 (30.0)		252 (35.2)		257 (31.3)		307 (30.6)
High		235 (28.4)		313 (32.6)		509 (45.5)		203 (28.4)		234 (28.5)		380 (37.9)
Education, n (%)												
No education		292 (33.4)		422 (38.1)		394 (35.1)		496 (64.8)		643 (66.7)		632 (63.0)
up to primary education		150 (17.2)		167 (15.1)		173 (15.4)		76 (9.9)		89 (9.2)		104 (10.4)
up to secondary education and beyond		432 (49.4)		518 (46.8)		557 (49.6)		194 (25.3)		232 (24.1)		267 (26.6)
Occupation, n (%)												
Agriculture		148 (16.9)		253 (22.9)		191 (17.0)		116 (15.1)		164 (17.0)		136 (13.6)
Daily wage earner		246 (28.2)		280 (25.3)		270 (24.0)		336 (43.9)		381 (39.5)		354 (35.3)
Unskilled manual worker		47 (5.4)		38 (3.4)		49 (4.4)		186 (24.3)		296 (30.7)		368 (36.7)
Self-employed		216 (24.7)		252 (22.8)		316 (28.1)		41 (5.4)		38 (3.9)		42 (4.2)

	Urbanisation level*, men (n= 3106)						Urbanisation level*, women (n= 2734)					
	n	**Low**	**n**	**Medium**	**n**	**High**	**n**	**Low**	**n**	**Medium**	**n**	**High**
Salaried worker		72 (8.2)		102 (9.2)		90 (8.0)		16 (2.1)		20 (2.1)		25 (2.5)
Student or unemployed		145 (16.6)		182 (16.4)		208 (18.5)		71 (9.3)		65 (6.7)		78 (7.8)
Shift work, n (%)	873	37 (4.2)	1107	53 (4.8)	1124	61 (5.4)	766	6 (0.8)	962	3 (0.3)	1002	1 (0.1)
Indoor open fire (6 months during lifetime), n (%)	874	843 (96.5)	1107	1062 (95.9)	1124	1051 (93.5)	766	733 (95.7)	963	941 (97.7)	1003	923 (92.0)
Solid fuel for cooking, n (%)	832	534 (64.2)	962	561 (58.3)	1124	481 (42.8)	717	459 (64.0)	824	509 (61.8)	1003	496 (49.5)
Number of household members, mean (sd)	874	4.3 (1.5)	1108	4.5 (1.6)	1124	4.3 (1.5)	766	4.7 (1.6)	965	4.5 (1.5)	1003	4.9 (1.8)
Lifestyle and mental health mediators												
Mild to severe depression, n (%)	869	10 (1.2)	1106	17 (1.5)	1123	11 (1.0)	766	19 (2.5)	962	14 (1.5)	1002	25 (2.5)
Physical activity level (PAL), median (IQR)	842	1.6 (1.4, 1.7)	1007	1.5 (1.4, 1.7)	1114	1.5 (1.4, 1.7)	724	1.6 (1.5, 1.9)	866	1.6 (1.5, 1.8)	986	1.6 (1.5, 1.8)
Sleep (h/d), mean (sd)	873	7.7 (1.1)	1107	7.8 (1.0)	1124	7.7 (1.1)	766	8.1 (1.1)	962	8.1 (1.0)	1002	8.2 (1.1)
Use tobacco, n (%)	874	306 (35.0)	1107	374 (33.8)	1124	391 (34.8)	766	94 (12.3)	963	153 (15.9)	1003	107 (10.7)
Alcohol (g/d), median (IQR)	872	61.1 (24.7, 132.2)	1107	56.3 (24.4, 131.4)	1123	46.3 (20.0, 105.7)	766	12.9 (5.0, 29.1)	962	13.6 (3.9, 37.1)	1001	10.2 (4.0, 29.1)
Energy (kcal/d), mean (sd)	848	2560 (1040.0)	1089	2666.2 (1096.2)	1089	2561.4 (1040.1)	748	1907 (677.4)	957	1839.1 (653.9)	978	1887.8 (653.2)
Saturated fat (g/d), median (IQR)	847	12.4 (9.1, 17.7)	1089	12.0 (8.6, 17.4)	1089	12.5 (8.9, 17.7)	748	9.9 (7.2, 14.2)	957	10.0 (7.0, 14.0)	978	10.4 (7.4, 14.5)
Monounsaturated fat (g/d), median (IQR)	847	14.0 (10.1, 19.8)	1089	12.7 (9.4, 18.0)	1089	13.5 (9.7, 18.1)	748	11.0 (7.9, 14.9)	957	10.4 (7.3, 14.1)	978	10.7 (7.7, 15.0)
Polyunsaturated fat (g/d), median (IQR)	848	10.7 (7.3, 16.7)	1089	11.7 (7.6, 18.5)	1089	12.6 (7.8, 19.8)	748	9.1 (6.2, 13.5)	957	9.0 (5.9, 14.0)	978	9.9 (6.2, 15.1)
Cholesterol (mg/d), median (IQR)	846	153.3 (103.5, 215.4)	1089	152.1 (104.6, 210.9)	1089	145.2 (93.3, 209.3)	748	120.8 (72.9, 166.7)	957	115.4 (71.3, 160.1)	978	110.0 (64.9, 163.8)
Percent carbohydrates of total energy (%/d), median (IQR)	872	69.6 (63.2, 74.0)	1107	69.7 (63.1, 74.2)	1123	68.6 (62.4, 73.4)	766	71.1 (66.8, 74.8)	962	70.3 (65.6, 74.2)	1001	70.1 (65.6, 74.1)
Sugar (g/d), median (IQR)	872	18.8 (12.6, 28.2)	1107	17.4 (10.4, 27.2)	1123	19.0 (12.1, 31.2)	766	17.5 (12.1, 25.2)	962	16.6 (10.9, 24.4)	1001	18.1 (12.6, 28.9)
Fibre (g/d), median (IQR)	848	7.9 (5.6, 11.3)	1089	7.2 (5.3, 10.3)	1089	7.7 (5.6, 11.0)	748	6.3 (4.7, 9.5)	957	6.0 (4.4, 8.8)	978	6.2 (4.4, 8.8)
Fruits and vegetables (g/d), median (IQR)	872	189.3 (122.5, 305.8)	1107	167.6 (106.6, 275.1)	1123	191.9 (122.5, 310.1)	766	159.5 (98.5, 247.0)	962	142.4 (92.8, 235.1)	1001	153.6 (98.2, 252.3)

	Urbanisation level*, men (n= 3106)						Urbanisation level*, women (n= 2734)					
	n	**Low**	**n**	**Medium**	**n**	**High**	**n**	**Low**	**n**	**Medium**	**n**	**High**
Fish (g/d), median (IQR)	872	1.1 (0.0, 4.6)	1107	1.1 (0.0, 4.6)	1123	1.5 (0.0, 6.5)	766	0.4 (0.0, 3.0)	962	0.4 (0.0, 3.0)	1001	0.8 (0.0, 5.9)
Salt (g/d), median (IQR)	872	6.2 (4.6, 8.6)	1107	6.1 (4.6, 8.5)	1123	6.1 (4.6, 8.3)	766	5.4 (4.0, 7.0)	962	5.4 (4.0, 7.0)	1001	5.4 (4.0, 7.2)
Folate (µg/d), median (IQR)	848	209.2 (164.3, 281.9)	1089	205.1 (159.0, 266.7)	1089	205.4 (159.4, 265.8)	748	176.4 (132.8, 224.7)	957	169.7 (130.1, 221.9)	978	169.4 (130.6, 226.0)
Vitamin B6 (µg/d), median (IQR)	848	1.6 (1.2, 2.1)	1089	1.4 (1.1, 2 .0)	1089	1.5 (1.2, 2.0)	748	1.2 (0.9, 1.5)	957	1.1 (0.9, 1.5)	978	1.2 (0.9, 1.5)
Vitamin B12 (µg/d), median (IQR)	847	1.4 (0.8, 4.9)	1088	1.3 (0.8, 2.8)	1089	1.7 (1.0, 4.3)	748	1.1 (0.7, 2.3)	957	1.1 (0.7, 2.3)	978	1.3 (0.8, 3.4)
Potassium (mg/d), median (IQR)	848	1673.0 (1270.6, 2227.7)	1089	1568.0 (1206.0, 2182.4)	1089	1709.2 (1282.0, 2344.5)	748	1346.5 (1049.7, 1748.6)	957	1284.8 (988.3, 1661.3)	978	1390.5 (1040.6, 1803.0)
Ascorbic acid (mg/d), median (IQR)	848	92.5 (55.7, 159.7)	1089	75.8 (46.7, 142.5)	1089	93.0 (56.9, 143.6)	748	77.0 (44.3, 136.4)	957	68.3 (39.0, 131.6)	978	73.1 (42.5, 131.1)
Vitamin E (mg alpha-tocopherol equivalents/d), median (IQR)	848	7.8 (5.4, 12.7)	1089	9.3 (5.7, 14.4)	1089	9.9 (6.0, 16.0)	748	6.7 (4.7, 10.7)	957	7.2 (4.9, 11.2)	978	7.5 (4.8, 12.5)
Riboflavin (mg/d), median (IQR)	848	1.2 (0.9, 1.6)	1089	1.1 (0.8, 1.5)	1089	1.1 (0.9, 1.5)	748	1.0 (0.7, 1.3)	957	0.9 (0.7, 1.3)	978	0.9 (0.7, 1.3)
Beta-carotene (µg/d), median (IQR)	848	1396.4 (974.8, 1991.9)	1089	1375.2 (1008.6, 1888.3)	1089	1360.0 (969.3, 1884.4)	748	1263.4 (876.0, 1709.2)	957	1205.6 (887.7, 1706.0)	978	1211.4 (859.5, 1626.0)
Selenium (µg/d), median (IQR)	848	92.8 (69.9, 124.8)	1089	93.7 (68.9, 126.5)	1089	92.2 (69.2, 122.8)	748	74.2 (55.5, 93.5)	957	72.7 (55.7, 93.5)	978	74.8 (55.5, 97.5)
Metabolic mediators												
Body mass index (kg/m^2), mean (sd)	873	21.0 (4.0)	1108	20.1 (3.2)	1121	21.0 (3.6)	764	20.6 (3.7)	961	21.3 (4.0)	1001	21.4 (4.2)
High-density lipoprotein (mmol/l), mean (sd)	855	1.1 (0.3)	1076	1.1 (0.4)	1108	1.1 (0.3)	735	1.2 (0.3)	921	1.2 (0.3)	967	1.1 (0.3)
Low-density lipoprotein (mmol/l), mean (sd)	789	2.5 (0.9)	1012	2.4 (0.8)	997	2.5 (0.9)	700	2.5 (0.7)	885	2.6 (0.8)	903	2.6 (0.8)
C-reactive protein (mg/l), median (IQR)	855	0.9 (0.4, 2.4)	1073	1.0 (0.4, 2.4)	1096	1.0 (0.4, 2.5)	735	1.1 (0.5, 2.6)	920	1.2 (0.5, 2.9)	965	1.2 (0.5, 3.2)
Waist-to-hip ratio, mean (sd)	873	0.9 (0.1)	1107	0.9 (0.1)	1121	0.9 (0.1)	764	0.8 (0.1)	960	0.8 (0.1)	1001	0.8 (0.1)
Fasting plasma glucose (mmol/l), mean (sd)	790	5.2 (1.1)	1019	5.2 (1.0)	1016	5.2 (1.3)	703	5.2 (1.2)	885	5.1 (1.1)	913	5.2 (1.3)

	Urbanisation level*, men (n= 3106)						Urbanisation level*, women (n= 2734)					
	n	**Low**	**n**	**Medium**	**n**	**High**	**n**	**Low**	**n**	**Medium**	**n**	**High**
Serum fasting insulin (mmol/l), median (IQR)	790	4.2 (2.6, 6.9)	1019	4.9 (2.6, 8.4)	1012	5.0 (2.6, 8.0)	700	4.8 (3.2, 7.6)	887	5.2 (3.2, 7.7)	909	6.0 (3.6, 9.0)
Triglycerides (mmol/l), median (IQR)	854	1.2 (0.9, 1.7)	1073	1.3 (0.9, 1.8)	1090	1.3 (0.9, 1.8)	735	1.0 (0.8, 1.4)	919	1.1 (0.8, 1.4)	963	1.1 (0.8, 1.5)

CI - confidence interval; IQR - interquartile range; n – sample size; sd - standard deviation

*Urbanisation levels were defined as thirds of ranked village-level NTLI

Table 5 Adjusted associations of mediators with systolic blood pressure (n=5840)

	Fully adjusted mean SBP change with increased mediator level*					
	Men (n= 3106)			Women (n= 2734)		
	n (%)**	β (95% CI)	p-value	n (%)**	β (95% CI)	p-value
Socio-demographic mediators						
SES, n (%)						
Low	968 (31.2)	Reference		979 (35.8)	Reference	
Medium	1045 (33.6)	0.2 (-1.2, 1.6)	0.81	908 (33.2)	1.7 (0.5, 3.0)	0.01
High	1093 (35.2)	1.8 (0.3, 3.3)	0.02	847 (31.0)	1.9 (0.5, 3.2)	0.01
Education, n (%)						
No education	1108 (35.7)	Reference		1772 (64.8)	Reference	
Up to primary education	490 (15.8)	1.9 (0.2, 3.6)	0.03	269 (9.8)	-0.4 (-2.2, 1.5)	0.69
Up to secondary education and beyond	1508 (48.6)	2.5 (0.7, 4.2)	0.01	693 (25.4)	-1.9 (-3.8, -0.1)	0.04
Occupation, n (%)						
Agriculture	592 (19.1)	Reference		416 (15.2)	Reference	
Daily wage earner	796 (25.6)	1.9 (0.2, 3.5)	0.02	1071 (39.2)	0.1 (-1.4, 1.6)	0.87
Unskilled manual worker	134 (4.3)	5.2 (2.4, 8.0)	<0.001	850 (31.1)	2.3 (0.7, 3.9)	0.004
Self-employed	785 (25.3)	5.9 (4.1, 7.7)	<0.001	121 (4.4)	0.04 (-2.7, 2.7)	0.98
Salaried worker	264 (8.5)	6.7 (4.4, 9.0)	<0.001	61 (2.2)	-0.4 (-4.0, 3.1)	0.82
Student or unemployed	535 (17.2)	4.0 (2.0, 6.0)	<0.001	215 (7.9)	0.4 (-2.1, 2.9)	0.76
Shift work, n (%)	151 (4.9)	-0.3 (-2.7, 2.2)	0.84	10 (0.4)	1.9 (-6.1, 9.9)	0.65
Indoor open fire (6 months during lifetime), n (%)	2957 (95.2)	-1.3 (-3.7, 1.2)	0.32	2599 (95.1)	-1.9 (-4.2, 0.4)	0.10
Solid fuel for cooking, n (%)	1678 (54.0)	-1.4 (-2.5, -0.2)	0.02	1567 (57.3)	-0.8 (-1.9, 0.3)	0.17
Number of household members, mean (sd)	3106 (100)	-0.4 (-0.8, 0.004)	0.05	2357 (100)	0.1 (-0.2, 0.5)	0.50
Lifestyle and mental health mediators						

	Fully adjusted mean SBP change with increased mediator level*					
	Men (n= 3106)			Women (n= 2734)		
	n (%)**	β (95% CI)	p-value	n (%)**	β (95% CI)	p-value
Mild to severe depression, n (%)	39 (1.3)	-1.2 (-6.0, 3.6)	0.63	58 (2.1)	-1.3 (-4.7, 2.0)	0.44
Physical activity level (PAL), median (IQR)	3106 (100)	-7.9 (-10.6, -5.3)	<0.001	2357 (100)	-3.0 (-5.7, -0.3)	0.03
Sleep (h/d), mean (sd)	1071 (34.5)	0.4 (-0.8, 1.6)	0.54	354 (13.0)	0.7 (-0.9, 2.3)	0.39
Use tobacco, n (%)	3106 (100)	0.001 (-0.002, 0.003)	0.58	2357 (100)	-0.001 (-0.01, 0.01)	0.72
Alcohol (g/d), median (IQR)	3107 (100)	0.7 (0.2, 1.2)	0.01	2358 (100)	0.6 (0.1, 1.0)	0.02
Energy (kcal/d), mean (sd)	3108 (100)	0.0002 (-0.0003, 0.001)	0.37	2359 (100)	-0.0002 (-0.001, 0.001)	0.54
Saturated fat (g/d), median (IQR)	3109 (100)	0.01 (-0.1, 0.1)	0.83	2360 (100)	0.02 (-0.1, 0.1)	0.63
Monounsaturated fat (g/d), median (IQR)	3110 (100)	0.01 (-0.1, 0.1)	0.87	2361 (100)	0.02 (-0.1, 0.1)	0.64
Polyunsaturated fat (g/d), median (IQR)	3111 (100)	0.04 (-0.02, 0.1)	0.19	2362 (100)	0.01 (-0.1, 0.1)	0.80
Cholesterol (mg/d), median (IQR)	3112 (100)	0.01 (0.001, 0.01)	0.02	2363 (100)	0.004 (-0.001, 0.01)	0.13
Percent carbohydrates of total energy (%/d), median (IQR)	3113 (100)	-0.05 (-0.1, 0.01)	0.07	2364 (100)	-0.1 (-0.1, -0.004)	0.04
Sugar (g/d), median (IQR)	3114 (100)	-0.01 (-0.03, 0.02)	0.73	2365 (100)	-0.003 (-0.03, 0.02)	0.78
Fibre (g/d), median (IQR)	3115 (100)	-0.04 (-0.1, 0.1)	0.52	2366 (100)	0.03 (-0.1, 0.1)	0.63
Fruits and vegetables (g/d), median (IQR)	3116 (100)	0.001 (-0.002, 0.004)	0.62	2367 (100)	0.001 (-0.003, 0.004)	0.63
Fish (g/d), median (IQR)	3117 (100)	0.02 (-0.01, 0.1)	0.21	2368 (100)	0.01 (-0.03, 0.1)	0.55
Salt (g/d), median (IQR)	3118 (100)	0.01 (-0.1, 0.2)	0.88	2369 (100)	0.02 (-0.2, 0.2)	0.82
Folate (µg/d), median (IQR)	3119 (100)	0.002 (-0.004, 0.01)	0.54	2370 (100)	0.002 (-0.004, 0.01)	0.49
Vitamin B6 (µg/d), median (IQR)	3120 (100)	0.1 (-0.6, 0.8)	0.82	2371 (100)	-0.3 (-1.2, 0.6)	0.53
Vitamin B12 (µg/d), median (IQR)	3121 (100)	0.1 (-0.04, 0.1)	0.27	2372 (100)	0.1 (-0.03, 0.1)	0.24
Potassium (mg/d), median (IQR)	3122 (100)	0.0002 (-0.0004, 0.001)	0.48	2373 (100)	0.0001 (-0.001, 0.001)	0.77
Ascorbic acid (mg/d), median (IQR)	3123 (100)	-0.002 (-0.01, 0.002)	0.37	2374 (100)	0.002 (-0.003, 0.01)	0.53
Vitamin E (mg alpha-tocopherol equivalents/d), median (IQR)	3124 (100)	0.04 (-0.03, 0.1)	0.25	2375 (100)	-0.002 (-0.1, 0.1)	0.97
Riboflavin (mg/d), median (IQR)	3125 (100)	0.1 (-0.7, 1.0)	0.79	2376 (100)	0.5 (-0.5, 1.5)	0.33

	Fully adjusted mean SBP change with increased mediator level*					
	Men (n= 3106)			Women (n= 2734)		
	n (%)**	β (95% CI)	p-value	n (%)**	β (95% CI)	p-value
Beta-carotene (µg/d), median (IQR)	3126 (100)	0.0004 (-0.0001, 0.001)	0.14	2377 (100)	0.001 (-0.0001, 0.001)	0.13
Selenium (µg/d), median (IQR)	3127 (100)	0.004 (-0.01, 0.02)	0.43	2378 (100)	0.003 (-0.01, 0.02)	0.69
Metabolic mediators						
Body mass index (kg/m2), mean (sd)	3126 (100)	1.4 (1.2, 1.5)	<0.001	2378 (100)	0.9 (0.8, 1.1)	<0.001
High-density lipoprotein (mmol/l), mean (sd)	3127 (100)	1.4 (-0.2, 3)	0.09	2379 (100)	-0.7 (-2.5, 1.0)	0.40
Low-density lipoprotein (mmol/l), mean (sd)	3128 (100)	2.1 (1.3, 2.8)	<0.001	2380 (100)	2.2 (1.5, 2.9)	<0.001
C-reactive protein (mg/l), median (IQR)	3129 (100)	-0.01 (-0.1, 0.1)	0.65	2381 (100)	0.1 (0.01, 0.2)	0.03
Waist-to-hip ratio, mean (sd)	3130 (100)	69.8 (60.6, 79.0)	<0.001	2382 (100)	37.2 (29.4, 45.0)	<0.001
Fasting plasma glucose (mmol/l), mean (sd)	3131 (100)	1.3 (0.8, 1.8)	<0.001	2383 (100)	0.9 (0.4, 1.3)	<0.001
Serum fasting insulin (mmol/l), median (IQR)	3132 (100)	0.2 (0.1, 0.3)	<0.001	2384 (100)	0.3 (0.2, 0.4)	<0.001
Triglycerides (mmol/l), median (IQR)	3133 (100)	3.6 (2.9, 4.3)	<0.001	2385 (100)	3.6 (2.8, 4.4)	<0.001

CI - confidence interval; IQR - interquartile range; n - sample size; sd - standard deviation; SBP - systolic blood pressure; β - beta-coefficient

Model predicted β-coefficients (95% CIs) and p-values were obtained from mixed-effects linear regression models with clustering by household and village and adjusted for age, marital status, caste, religion, season of survey, season of birth; parental history of obesity, diabetes, hypertension, stroke and coronary heart disease; and room temperature at measurement of blood pressure

*Mean SBP change per one unit (integer) increase in continuous measures or mean SBP change from base level to comparator level for binary and categorical variables

**Sample sizes for categorical variables were reported for multiple imputations dataset 1

Table 6 Total, direct and indirect effects of night-time light intensity on systolic blood pressure among total sample (n=5840) and participants ≥40 years (n=2526)

	Fully adjusted mean SBP change with increasing NTLI* β (95% CI)			
	Men		**Women**	
	Total	**≥40 years**	**Total**	**≥40 years**
n (%)	3106 (100)	1201 (38.7)	2734 (100)	1325 (48.5)
Total effect	2.0 (0.4, 3.5)	4.2 (1.8, 6.7)	1.3 (0.01, 2.6)	2.8 (0.9, 4.6)
Direct effect	-3.4 (-7.5, 0.7)	-5.1 (-12.4, 2.2)	-0.8 (-2.9, 1.4)	0.1 (-3.2, 3.4)
Indirect effect via socio-demographic factors	0.2 (-0.2, 0.5)	0.1 (-0.6, 0.8)	1.0 (-0.8, 2.7)	1.5 (-1.0, 4.1)
Indirect effect via lifestyle and mental health factors	0.7 (0.1, 1.3)	1.0 (-0.3, 2.2)	0.4 (-0.2, 1.0)	0.7 (-0.4, 1.8)
Indirect effect via metabolic factors	4.6 (2.0, 7.3)	8.3 (2.5, 14.1)	0.7 (0.1, 1.3)	0.4 (-0.5, 1.3)

CI - confidence interval; n - sample size; NTLI – night-time light intensity; SBP – systolic blood pressure; β - beta-coefficient

The direct effect is the effect acting around mediators on SBP; the indirect effect via socio-demographic factors may also act via lifestyle and mental health factors, and metabolic factors; the indirect effect via lifestyle and mental health factors may also act via metabolic factors

Model predicted β-coefficients (95% CIs) were obtained from mixed-effects linear regression models with clustering by household and village and adjusted for age, caste, religion, marital status, season of birth, season of survey, room temperature at measurement of blood pressure and parental history of obesity, diabetes, hypertension, stroke and coronary heart disease and relevant effect modification terms for age and NTLI

*Mean SBP change per one unit (integer) increase in log-transformed NTLI (range 4.1, 7.0)

Figures

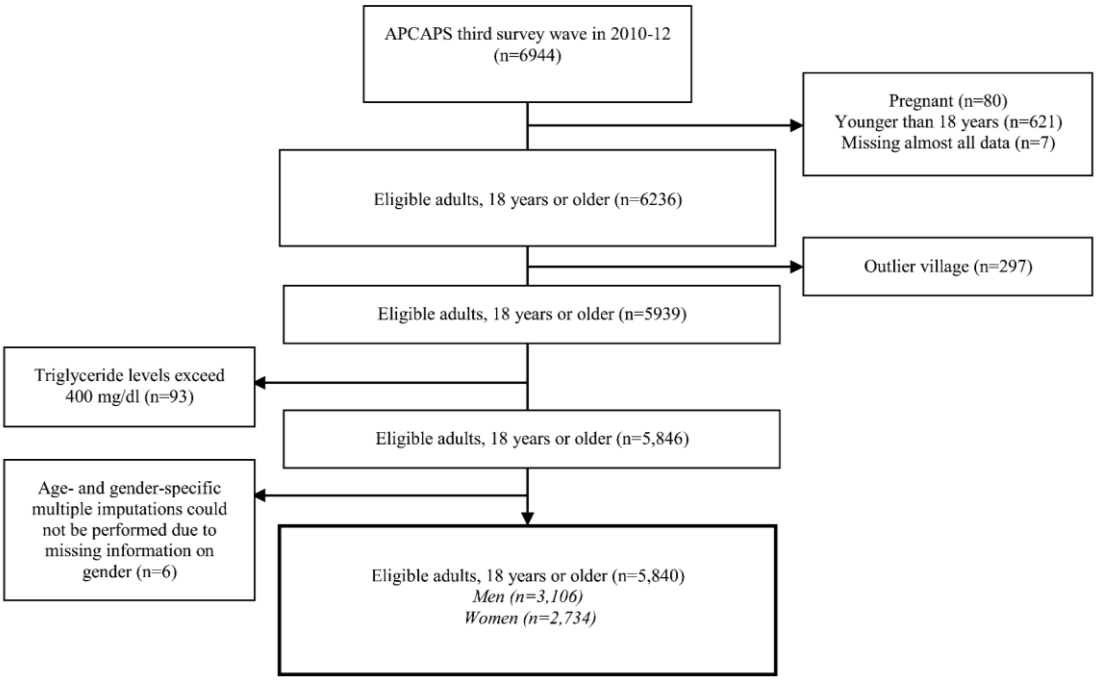


Figure 1 Selection of study participants

APCAPS – Andhra Pradesh Children and Parents Study

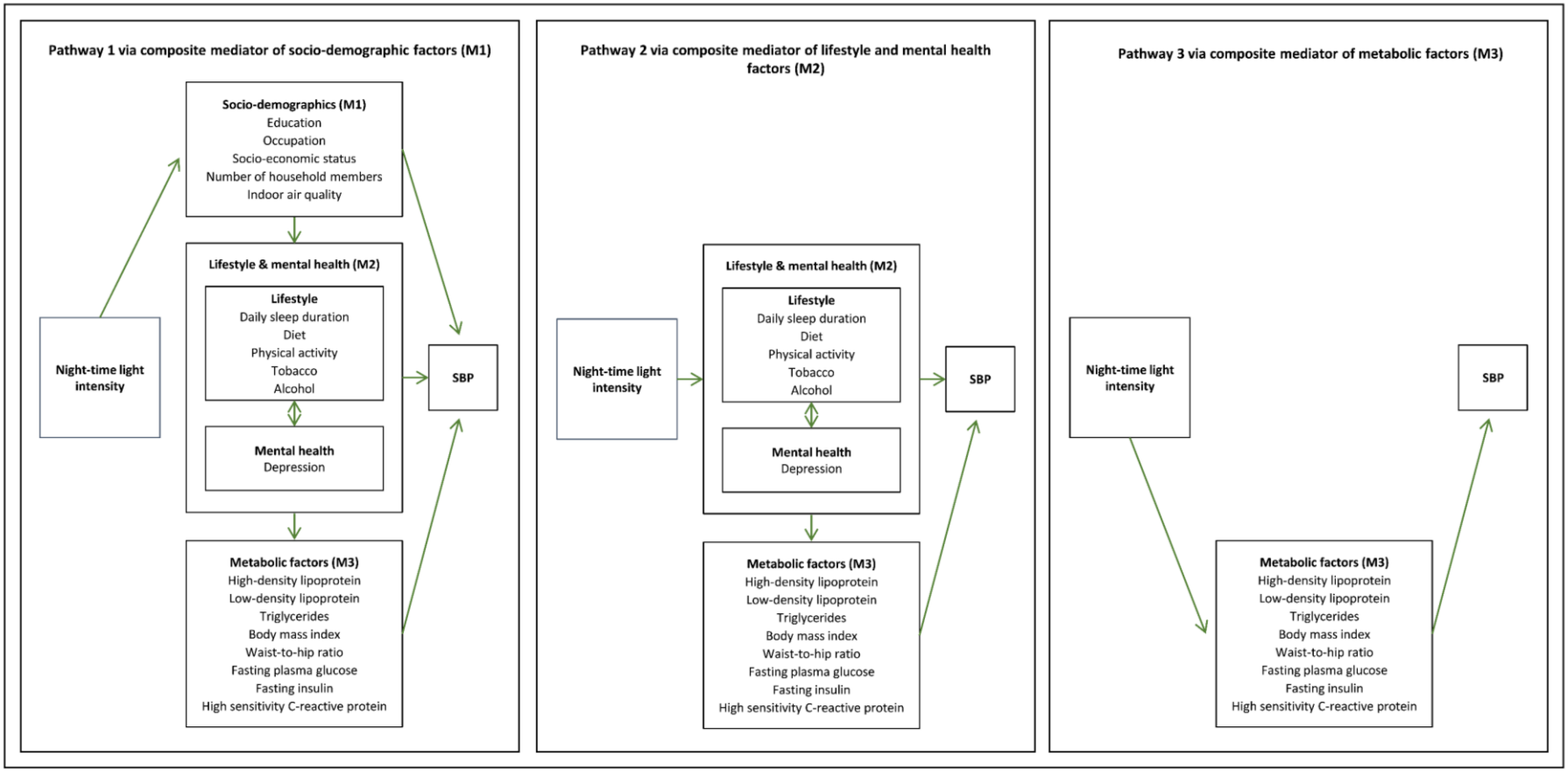


Figure 2 Estimated indirect effects of levels of urbanisation (measured by log-transformed continuous night-time light intensity)

SBP –systolic blood pressure; M1-3 - composite mediators

Arrows represent hypothesised causal effects

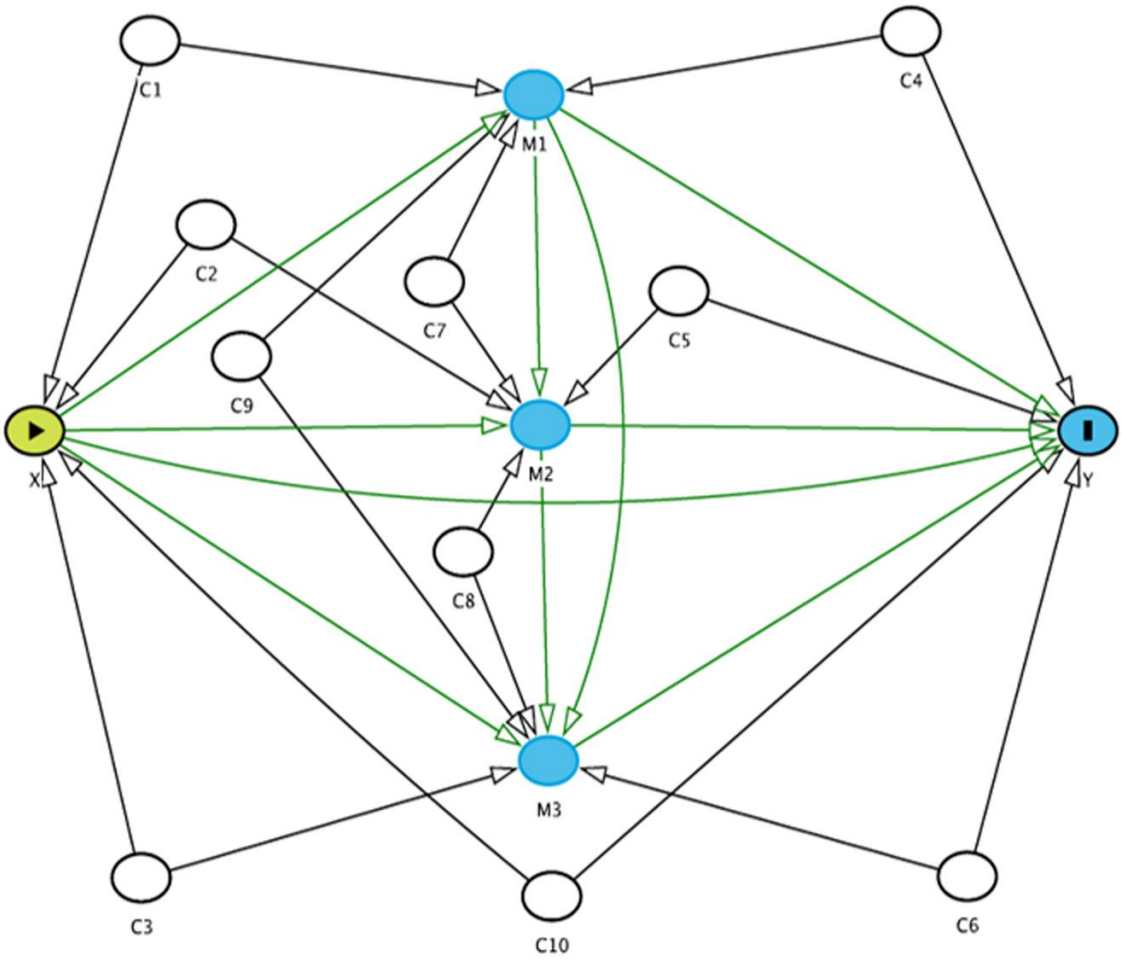


Figure 3 Directed Acyclic Graph of the causal effect of urbanisation level on systolic blood pressure (SBP)

The DAG illustrates the causal effects of urbanisation level (X) on SBP (Y), directly and indirectly via three composite mediators summarising information on socio-demographic factors (M1); lifestyle and mental health factors (M2); and metabolic factors (M3) independent of confounders (C) that may influence relationships of exposure and each of the mediator summaries; relationships between mediator summaries and outcome and relationships between mediator summaries*. The green arrows illustrate causal pathways and black arrows illustrate potential confounding

C - set of confounders; M – composite mediators, X - urbanisation level (measured by continuous night-time light intensity), Y - SBP

*All sets of confounders (C1-10) included age, marital status, caste, religion, survey season, birth season, room temperature; parental history of obesity, diabetes, hypertension, stroke and coronary heart disease. In models for the pathway via M2, sets of confounders (C2, C5, C6, C8, C10) additionally included M1 mediators. In models for the pathway via M3, sets of confounders (C3, C6) additionally included M1 and M2 mediators.

Appendix I Theoretical framework

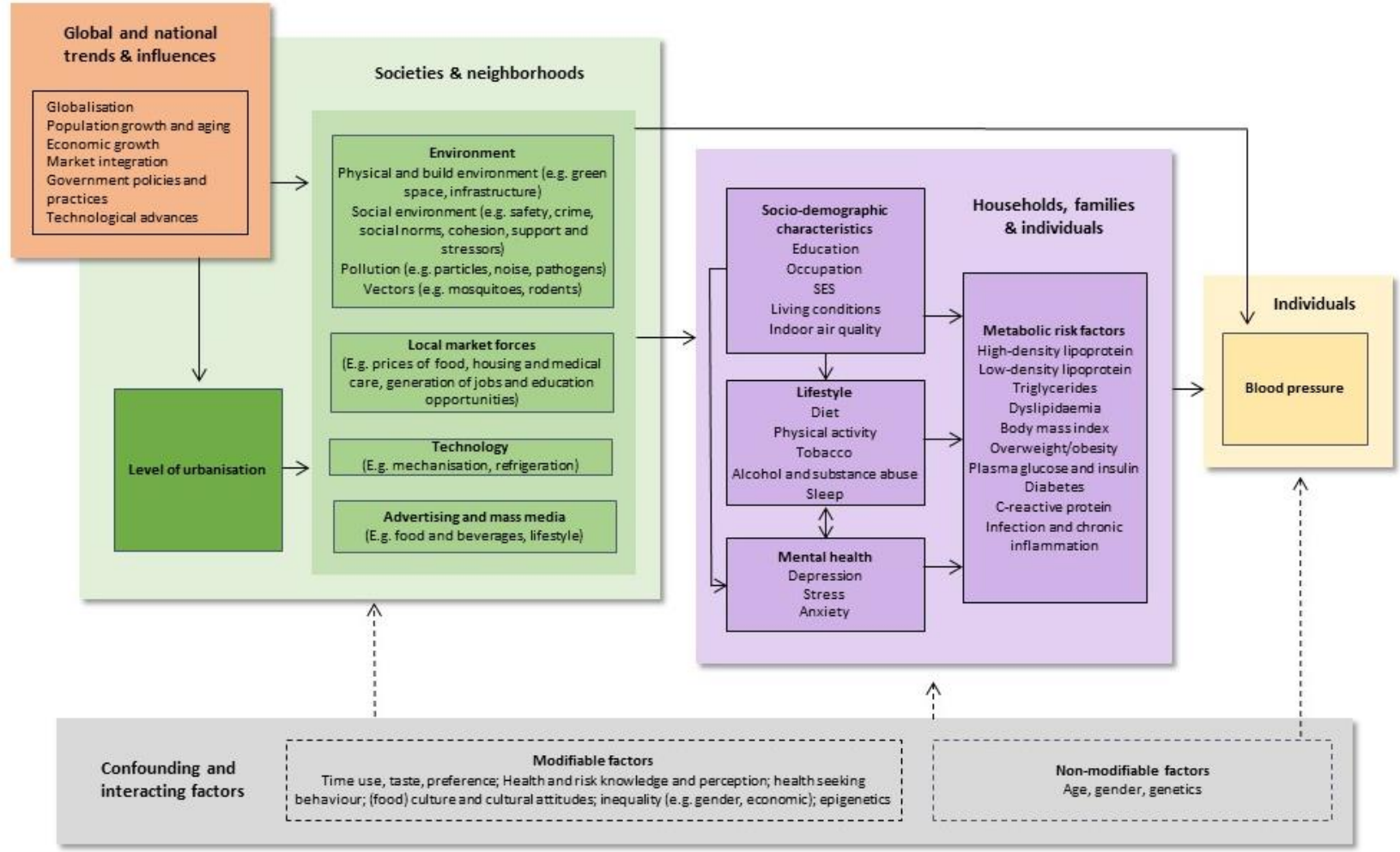


Figure A1 Theoretical framework for associations of urbanisation levels with blood pressure[1-26]

References

1. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
2. Ezzati M, Webster CJ, Doyle YG, Rashid S, Owusu G, Leung GM. Cities for global health. 2018;363:k3794
3. World Health Organization. Urbanization and health 2019 [cited 2019 16 July]. Available from: https://www.who.int/bulletin/volumes/88/4/10-010410/en/.
4. Vlahov D, Freudenberg N, Proietti F, Ompad D, Galea S. A conceptual framework for organizing determinants of urban health. Thematic paper 1 for Knowledge Network on Urban Settings WHO Centre for Health Development2008.
5. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012 Jan;70(1):3-21
6. Estache A, Garsous G. The scope for an impact of infrastructure investments on jobs in developing countries International Finance Corporation, 2012.
7. Eckert S, Kohler S. Urbanization and health in developing countries: a systematic review. World health & population. 2014;15(1):7-20
8. Zhao B, Zheng H, Wang S, Smith KR, Lu X, Aunan K, et al. Change in household fuels dominates the decrease in PM2.5 exposure and premature mortality in China in 2005-2015. Proceedings of the National Academy of Sciences of the United States of America. 2018 Dec 4;115(49):12401-6
9. Yamamoto SS, Phalkey R, Malik AA. A systematic review of air pollution as a risk factor for cardiovascular disease in South Asia: limited evidence from India and Pakistan. International journal of hygiene and environmental health. 2014 Mar;217(2-3):133-44
10. Uzoigwe JC, Prum T, Bresnahan E, Garelnabi M. The emerging role of outdoor and indoor air pollution in cardiovascular disease. N Am J Med Sci. 2013 Aug;5(8):445-53
11. Moore M, Gould P, Keary BS. Global urbanization and impact on health. International journal of hygiene and environmental health. 2003 2003/01/01/;206(4):269-78
12. Vogel CF, Sciullo E, Wong P, Kuzmicky P, Kado N, Matsumura F. Induction of proinflammatory cytokines and C-reactive protein in human macrophage cell line U937 exposed to air pollution particulates. Environmental health perspectives. 2005 Nov;113(11):1536-41
13. Cyril S, Oldroyd JC, Renzaho A. Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales. BMC public health. 2013 May 28;13:513
14. Meo SA, Suraya F. Effect of environmental air pollution on cardiovascular diseases. European review for medical and pharmacological sciences. 2015 Dec;19(24):4890-7
15. India State-level Disease Burden Initiative Collaborators. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet. 2017 Nov 13
16. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008 Nov;85(6):938-51
17. Goryakin Y, Suhrcke M. Economic development, urbanization, technological change and overweight: what do we learn from 244 Demographic and Health Surveys? Economics and human biology. 2014 Jul;14:109-27
18. Lane KJ, Stokes EC, Seto KC, Thanikachalam S, Thanikachalam M, Bell ML. Associations between Greenness, Impervious Surface Area, and Nighttime Lights on Biomarkers of Vascular Aging in Chennai, India. Environmental health perspectives. 2017 Aug 2;125(8):087003
19. Sullivan R, Kinra S, Ekelund U, Bharathi AV, Vaz M, Kurpad A, et al. Socio-demographic patterning of physical activity across migrant groups in India: results from the Indian Migration Study. PloS one. 2011;6(10):e24898
20. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March. Report No.
21. UNDP/UNFPA, UNICEF, WFP. Agenda Item 1: Population Growth and Rapid Urbanization: Food Insecurity on the Rise in Urban Settings. UNDP/UNFPA, UNICEF, WFP, 2009.
22. Popkin BM, Horton S, Kim S, Mahal A, Shuigao J. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutrition reviews. 2001 Dec;59(12):379-90

23. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. The American journal of clinical nutrition. 2006;84(2):289-98
24. Godfrey R, Julien M. Urbanisation and health. Clin Med (Lond). 2005 Mar-Apr;5(2):137-41
25. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.
26. Ibrahim MM, Damasceno A. Hypertension in developing countries. Lancet. 2012 Aug 11;380(9841):611-9

Appendix II Mathematical notes

Web Appendix for 'Quantifying the influence of location of residence on blood pressure in urbanising South India: A path analysis with multiple mediators'

1 Mediation analysis

The analysis was based on a mediation analysis under linear models for 1 outcome Y and 3 mediators M_1, M_2, M_3. Although one can in principle appeal to linear structural equation models, the concern is that identification of certain pathways requires very stringent assumptions (about the correlation between counterfactuals for the same individual corresponding to different levels of the mediator) that are almost guaranteed to be misspecified (Daniel et al., 2015). We therefore chose to limit the mediation analysis to the pathways that can be identified under less restrictive assumptions (VanderWeele and Vansteelandt, 2013).

First, we fitted model

$$Y = \beta_0 X + \beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3 + \beta_4 C + b_h + b_v + \epsilon,$$

where C is a vector of confounders (including 1 for the intercept), b_h and b_v are mutually independent random effects to capture the effects of household and village, respectively, and ϵ is a residual error term. Then β_0 captures the direct effect of X on Y via none of the mediators M_1, M_2 and M_3. Next, we defined

$$M_3^* = \beta_3 M_3$$

and fitted model

$$M_3^* = \gamma_0 X + \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 C + a_h + a_v + \nu,$$

where a_h and a_v are mutually independent random effects to capture the effects of household and village, respectively, and ν is a residual error term. Here, γ_0 captures the indirect effect of X on Y via M_3 (but not M_1 and M_2). That is, the effect along the pathway $X \to M_3 \to Y$.

Next, we dropped M_3 from all analyses and repeated the above strategy. With a slight abuse of notation, we thus fitted model

$$Y = \beta_0' X + \beta_1' M_1 + \beta_2' M_2 + \beta_3' C + b_h + b_v + \epsilon,$$

defined

$$M_2^* = \beta_2' M_2$$

and subsequently fitted model

$$M_2^* = \gamma_0' X + \gamma_1' M_1 + \gamma_2' C + a_h + a_v + \nu.$$

Here, γ_0' captures the indirect effect of X on Y via M_2 (but not M_1). That is, the combined effect along the pathways $X \to M_2 \to Y$ and $X \to M_2 \to M_3 \to Y$.

Finally, we also leave out M_2 from all analyses and repeat the above strategy. With a slight abuse of notation, we thus fitted model

$$Y = \beta_0'' X + \beta_1'' M_1 + \beta_2'' C + b_h + b_v + \epsilon,$$

defined

$$M_1^* = \beta_1'' M_1$$

and subsequently fitted model

$$M_1^* = \gamma_0'' X + \gamma_1'' C + a_h + a_v + \nu.$$

Here, γ_0'' captures the indirect effect of X on Y via M_1. That is, the combined effect along the pathways $X \to M_1 \to Y$, $X \to M_1 \to M_2 \to Y$, $X \to M_1 \to M_3 \to Y$ and $X \to M_1 \to M_2 \to M_3 \to Y$.

For all analyses, the key assumptions are that C is sufficient to adjust for confounding of the exposure-outcome association and the exposure-mediator association (for each of the mediators separately), and that C along with the exposure and the 'previous' mediators is sufficient to adjust for confounding of the association between the considered mediator M_k and outcome ($k = 1, 2, 3$). Our analysis also invokes the more technical assumption that the causal diagram that we consider represents a nonparametric structural equations model with independent errors. This assumption would be violated if some of the confounders of the mediator-outcome relation are affected by the exposure (as may be the case when there are other mediators that were not considered, and affect either M_1, M_2 or M_3).

2 Calculation of standard errors

Standard error for the direct effect were readily obtained from the standard Stata output. The calculation of standard errors on the indirect effects was more involved, because the standard errors calculated by the software when fitting the models for M_1^*, M_2^* and M_3^* do not acknowledge that these variables were constructed using estimated regression coefficients from a model for the outcome, which have their own uncertainty. To reflect this additional uncertainty, we therefore added noise to M_1^*, M_2^* and M_3^* by perturbing the coefficients from the outcome model with a random normal measurement with mean zero in all components and covariance matrix given by the estimated covariance matrix of the outcome model coefficients, which is readily available from the Stata output. Based on the resulting new coefficient values, we recalculated M_1^*, M_2^* and M_3^* as before and proceeded with the rest of the analysis exactly as before. We thus repeated the analysis 300 times, corresponding with 300 perturbations of M_1^*, M_2^* and M_3^*.

We then calculated the variance of the indirect effect as the sum of the sample variance of the 300 obtained estimates of the indirect effect, plus the mean of the 300 estimated variances of the indirect effect, as reported by the standard software, which ignore the uncertainty in the coefficients β_1'', β_2' and β_3.

Because of missing data, the analyses were repeated on 10 imputed data sets and Rubin's rules were used to adjust the variance of the direct and indirect effects for the additional uncertainty due to the missing data.

References

Daniel, R.M., De Stavola, B.L., Cousens, S.N. and Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. Biometrics, 71, 1-14.

VanderWeele, T. and Vansteelandt, S. (2013). Mediation Analysis with Multiple Mediators. Epidemiologic Methods, 2, 95-115.

Appendix III Missing data

Table A1 Missing data

Variable	n missing observations	n complete observations	Percentage missing
Gender	6	5,840	0.1
Caste	242	5,604	4.1
Married	8	5,838	0.1
Birth season	6	5,840	0.1
Room temperature	2,606	3,240	44.6
Index child	4,611	1,235	78.9
Mother obese	11	5,835	0.2
Mother diabetes	11	5,835	0.2
Mother hypertension	11	5,835	0.2
Mother had a stroke	11	5,835	0.2
Mother had coronary heart disease	11	5,835	0.2
Father obese	11	5,835	0.2
Father diabetes	11	5,835	0.2
Father hypertension	11	5,835	0.2
Father had a stroke	11	5,835	0.2
Father had coronary heart disease	11	5,835	0.2
Socio-economic status	399	5,447	6.8
Education	8	5,838	0.1
Occupation	8	5,838	0.1
Shift work	12	5,834	0.2
Open fire	9	5,837	0.2
Solid fuel	384	5,462	6.6
Depression group	18	5,828	0.3
Physical activity level	307	5,539	5.3
Any tobacco	9	5,837	0.2
Alcohol	15	5,831	0.3
Hours of sleep	12	5,834	0.2
Energy	137	5,709	2.3
Saturated fat	138	5,708	2.4
Monounsaturated fat	138	5,708	2.4
Polyunsaturated fat	137	5,709	2.3
Cholesterol	139	5,707	2.4
Percent carbohydrates	15	5,831	0.3
Sugar/sweets	15	5,831	0.3
Fibre	137	5,709	2.3
Fruit and vegetables	15	5,831	0.3
Fish	15	5,831	0.3
Salt	15	5,831	0.3
Folate	137	5,709	2.3
Vitaminb6	137	5,709	2.3

Variable	n missing observations	n complete observations	Percentage missing
Vitaminb12	139	5,707	2.4
Potassium	137	5,709	2.3
Vitamin c	137	5,709	2.3
Vitamin e	137	5,709	2.3
Riboflavin	137	5,709	2.3
Carotene	137	5,709	2.3
Selenium	137	5,709	2.3
Body mass index	14	5,832	0.2
Systolic blood pressure	4	5,842	0.1
High-density lipoprotein	181	5,665	3.1
Low-density lipoprotein	558	5,288	9.5
C-Reactive Protein	199	5,647	3.4
Waist-to hip ratio	16	5,830	0.3
Fasting plasma glucose	518	5,328	8.9
Fasting insulin	529	5,317	9.0
Triglycerides	209	5,637	3.6

n – sample size

Appendix IV Stata code (for analysis of men)

```
*******************************************************************************
*PATH ANALYSIS STEP 1
*Generating mediator groups M1-M3 and estimating path-specific effects
*******************************************************************************
* Note: load flong separated data

*Model 0: modelling the total effect of NTLI on SBP, adjusted for confounders
mi estimate, post: mixed SBP lognli ///
i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev
*saving estimate of the total effect (pooled with Rubin's rule) in each MI dataset
mi xeq: gen totaleffect =_b[lognli]

/*MODEL 1: Modelling the (pooled) direct effect of NTLI on SBP, not through any of
the mediators, adjusted for confounders*/
mi estimate, post: mixed SBP lognli  ///
i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp ///
i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
energy satfat monounsaturatedfat polyunsaturatedfat  ///
cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
hdl_mol BMI crp LDL whr glucose fastinginsulin trigly_mol ///
c.hdl_mol#i.agegp c.BMI#i.agegp c.fastinginsulin#i.agegp c.trigly_mol#i.agegp ///
c.crp#c.lognli c.glucose#c.lognli  ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev
*saving the direct effect (pooled with Rubin's rule) in each MI dataset
mi xeq: gen direct =_b[lognli]

***** Estimating the effect via M3* metabolic factors)*****

*In 10 MI datasets: Generating individual-level predictions for mediators belonging to M3
mi xeq: mixed SBP lognli ///
i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp ///
i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
energy satfat monounsaturatedfat polyunsaturatedfat  ///
cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
hdl_mol BMI crp LDL  whr glucose fastinginsulin trigly_mol ///
c.hdl_mol#i.agegp c.BMI#i.agegp c.fastinginsulin#i.agegp c.trigly_mol#i.agegp ///
c.crp#c.lognli c.glucose#c.lognli  ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev; ///
matrix result=r(table); matrix list r(table); ///
gen hdl_star=result[1,83] * hdl_mol; ///
gen BMI_star=result[1,84] * BMI; ///
gen crp_star=result[1,85] * crp; ///
gen LDL_star=result[1,86] * LDL; ///
gen whr_star=result[1,87] * whr; ///
gen glucose_star=result[1,88] * glucose; ///
gen fastinginsulin_star=result[1,89] * fastinginsulin; ///
gen trigly_star=result[1,90] * trigly_mol; ///
gen hdlage0_star=result[1,91] * 0.agegp#c.hdl_mol; ///
gen hdlage1_star=result[1,92] * 1.agegp#c.hdl_mol; ///
gen hdlage2_star=result[1,93] * 2.agegp#c.hdl_mol; ///
gen hdlage3_star=result[1,94] * 3.agegp#c.hdl_mol; ///
gen bmiage0_star=result[1,95] * 0.agegp#c.BMI; ///
gen bmiage1_star=result[1,96] * 1.agegp#c.BMI; ///
gen bmiage2_star=result[1,97] * 2.agegp#c.BMI; ///
gen bmiage3_star=result[1,98] * 3.agegp#c.BMI; ///
```

```
gen fastinginsulinage0_star=result[1,99] * 0.agegp#c.fastinginsulin; ///
gen fastinginsulinage1_star=result[1,100] * 1.agegp#c.fastinginsulin; ///
gen fastinginsulinage2_star=result[1,101] * 2.agegp#c.fastinginsulin; ///
gen fastinginsulinage3_star=result[1,102] * 3.agegp#c.fastinginsulin; ///
gen trigly_molage0_star=result[1,103] * 0.agegp#c.trigly_mol; ///
gen trigly_molage1_star=result[1,104] * 1.agegp#c.trigly_mol; ///
gen trigly_molage2_star=result[1,105] * 2.agegp#c.trigly_mol; ///
gen trigly_molage3_star=result[1,106] * 3.agegp#c.trigly_mol; ///
gen crpntli_star=result[1,107] * c.crp#c.lognli; ///
gen glucosentli_star=result[1,108] * c.glucose#c.lognli

        /*In 10 MI datasets: generating mediator group M3* by summing predictions for
        individual metabolic mediators (n=8) */
        mi xeq: egen M3_star= rowtotal(hdl_star BMI_star crp_star LDL_star whr_star ///
        glucose_star fastinginsulin_star trigly_star hdlage0_star hdlage1_star hdlage2_star ///
        hdlage3_star bmiage0_star bmiage1_star bmiage2_star bmiage3_star fastinginsulinage0_star ///
        fastinginsulinage1_star fastinginsulinage2_star fastinginsulinage3_star ///
        trigly_molage0_star trigly_molage1_star trigly_molage2_star trigly_molage3_star ///
        crpntli_star glucosentli_star)


*MODEL 2: The indirect (pooled) effect of NTLI on SBP via M3 only, adjusted for confounders
mi estimate, post : mixed M3_star lognli ///
i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp  ///
i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
energy satfat monounsaturatedfat polyunsaturatedfat  ///
cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev
*Saving the indirect effect via M3* (pooled with Rubin's rule) in each MI dataset
mi xeq: gen M3effect=_b[lognli]

        ***** Estimating the effect via M2* (lifestyle and depression) and M3*****
                        *****leaving out M3 mediators *****

        /*In 10 MI datasets: Generating individual-level predictions for mediators belonging to M2*/
        mi xeq: mixed SBP lognli ///
        i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp ///
        i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
        energy satfat monounsaturatedfat polyunsaturatedfat ///
        cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
        vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
        c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
        || village: || famid:, reml stddev; ///
        matrix result=r(table); matrix list r(table); ///
        gen tobacco_star=result[1,43+anytobacco]; ///
        gen depression_star=result[1,45+depressiongp]; ///
        gen pal_star=result[1,47] *pal; ///
        gen sleep_star=result[1,48]  *sleep_hours_day; ///
        gen alc_star=result[1,49] *beveragesalcohol; ///
        gen energy_star=result[1,50] *energy; ///
        gen satfat_star=result[1,51] *satfat; ///
        gen mufa_star=result[1,52] *monounsaturatedfat; ///
        gen pufa_star=result[1,53] *polyunsaturatedfat; ///
        gen cholesterol_star=result[1,54] *cholesterol; ///
        gen percarb_star=result[1,55] *percarb; ///
        gen sugar_star=result[1,56] *sugarsweets; ///
        gen fibre_star=result[1,57] *fibertotaldietary; ///
        gen fruitveg_star=result[1,58] *fruitveg; ///
        gen fish_star=result[1,59] *fish; ///
        gen salt_star=result[1,60] *salt; ///
        gen folate_star=result[1,61] *folate; ///
        gen vitaminb6_star=result[1,62] *vitaminb6; ///
        gen vitaminb12_star=result[1,63] *vitaminb12; ///
```

```
gen potassium_star=result[1,64] *potassiumk; ///
gen vitc_star=result[1,65] *vitamincascorbicacid; ///
gen vitamine_star=result[1,66] *vitamine; ///
gen riboflavin_star=result[1,67] *riboflavin; ///
gen carotene_star=result[1,68] *carotene; ///
gen seleniumse_star=result[1,69] *seleniumse; ///
gen sleepage0_star=result[1, 70] *0.agegp#c.sleep_hours_day; ///
gen sleepage1_star=result[1, 71] *1.agegp#c.sleep_hours_day; ///
gen sleepage2_star=result[1, 72] *2.agegp#c.sleep_hours_day; ///
gen sleepage3_star=result[1, 73] *3.agegp#c.sleep_hours_day; ///
gen alcntli_star=result[1,74] * c.beveragesalcohol#c.lognli; ///
gen percarbage0_star=result[1, 75] *0.agegp#c.percarb; ///
gen percarbage1_star=result[1, 76] *1.agegp#c.percarb; ///
gen percarbage2_star=result[1, 77] *2.agegp#c.percarb; ///
gen percarbage3_star=result[1, 78] *3.agegp#c.percarb; ///
gen vitaminb6age0_star=result[1,79] *0.agegp#c.vitaminb6; ///
gen vitaminb6age1_star=result[1, 80] *1.agegp#c.vitaminb6; ///
gen vitaminb6age2_star=result[1, 81] *2.agegp#c.vitaminb6; ///
gen vitaminb6age3_star=result[1, 82] *3.agegp#c.vitaminb6

        /* In 10 MI datasets: Generating mediator group M2* (Lifestyle and depression)
        by summing predicted individual level lifestyle and depression factors (n=25) */
        mi xeq : egen M2_star = rowtotal(tobacco_star depression_star pal_star ///
        sleep_star alc_star energy_star satfat_star mufa_star pufa_star cholesterol_star ///
        percarb_star sugar_star fibre_star fruitveg_star fish_star salt_star folate_star ///
        vitaminb6_star vitaminb12_star potassium_star vitc_star vitamine_star ///
        riboflavin_star carotene_star seleniumse_star sleepage0_star sleepage1_star ///
        sleepage2_star sleepage3_star alcntli_star percarbage0_star percarbage1_star ///
        percarbage2_star percarbage3_star vitaminb6age0_star vitaminb6age1_star ///
        vitaminb6age2_star vitaminb6age3_star)

/*MODEL 3: The indirect (pooled) effect via M2* and M3*, adjusted for confounders */
mi estimate, post: mixed M2_star lognli  ///
i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp  ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev
*saving indirect effect via M2 (pooled with Rubin's rule) in each MI dataset
mi xeq: gen M2effect=_b[lognli]

        ***** Estimating the effect via M1* (metabolic factors), M2 and M3 *****
                *****leaving out M2 and M3 mediators *****

        /*In each of the 10 MI datasets: Generating individual-level predictions
        for mediators belonging to M1*/
        mi xeq: mixed SBP lognli  ///
        i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel ///
        i.occu_path#i.agegp  ///
        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
        || village: || famid: , reml stddev; ///
        matrix result=r(table); matrix list r(table); ///
        gen edu_star=result[1,2+edu]; ///
        gen ses_star=result[1,5]*zpc2; ///
        gen n_hh_star=result[1,6]*n_hh; ///
        gen occu_star=result[1,7+occu_path]; ///
        gen shiftwork_star =result[1,13+shiftwork]; ///
        gen openfire_star=result[1,15+openfire]; ///
        gen solidfuel_star=result[1,17+solidfuel]; ///
        gen occuage00_star=result[1,19] * 0.occu_path#0.agegp; ///
        gen occuage01_star=result[1,20] * 0.occu_path#1.agegp; ///
        gen occuage02_star=result[1,21] * 0.occu_path#2.agegp; ///
        gen occuage03_star=result[1,22] * 0.occu_path#3.agegp; ///
        gen occuage10_star=result[1,23] * 1.occu_path#0.agegp; ///
        gen occuage11_star=result[1,24] * 1.occu_path#1.agegp; ///
        gen occuage12_star=result[1,25] * 1.occu_path#2.agegp; ///
        gen occuage13_star=result[1,26] * 1.occu_path#3.agegp; ///
        gen occuage20_star=result[1,27] * 2.occu_path#0.agegp; ///
        gen occuage21_star=result[1,28] * 2.occu_path#1.agegp; ///
```

```
gen occuage22_star=result[1,29] * 2.occu_path#2.agegp; ///
gen occuage23_star=result[1,30] * 2.occu_path#3.agegp; ///
gen occuage30_star=result[1,31] * 3.occu_path#0.agegp; ///
gen occuage31_star=result[1,32] * 3.occu_path#1.agegp; ///
gen occuage32_star=result[1,33] * 3.occu_path#2.agegp; ///
gen occuage33_star=result[1,34] * 3.occu_path#3.agegp; ///
gen occuage40_star=result[1,35] * 4.occu_path#0.agegp; ///
gen occuage41_star=result[1,36] * 4.occu_path#1.agegp; ///
gen occuage42_star=result[1,37] * 4.occu_path#2.agegp; ///
gen occuage43_star=result[1,38] * 4.occu_path#3.agegp; ///
gen occuage50_star=result[1,39] * 5.occu_path#0.agegp; ///
gen occuage51_star=result[1,40] * 5.occu_path#1.agegp; ///
gen occuage52_star=result[1,41] * 5.occu_path#2.agegp; ///
gen occuage53_star=result[1,42] * 5.occu_path#3.agegp

        /*In 10 MI datasets: generating M1* by summing predicted
        socio-demo factors (n=6)*/
        mi xeq: egen M1_star = rowtotal(edu_star ses_star n_hh_star ///
        occu_star shiftwork_star openfire_star solidfuel_star ///
        occuage00_star occuage01_star occuage02_star occuage03_star ///
        occuage10_star occuage11_star occuage12_star occuage13_star ///
        occuage20_star occuage21_star occuage22_star occuage23_star ///
        occuage30_star occuage31_star occuage32_star occuage33_star ///
        occuage40_star occuage41_star occuage42_star occuage43_star ///
        occuage50_star occuage51_star occuage52_star occuage53_star)

/*MODEL 4: The (pooled) effect of NTLI on SBP via M1, M2 and M3,
adjusted for confounders */
mi estimate, post: mixed M1_star lognli   ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid: , reml stddev
*saving effect via M1 (pooled with Rubin's rule) in each MI dataset
mi xeq: gen M1effect=_b[lognli]

****DERIVING BETAS FOR EACH MODEL FOR CALCULATION OF STANDARD ERRORS IN LATER STEPS *****

*In 10 MI datasets: Deriving beta from Model 2 (via M3)
forvalues mi =1/10 {
        mi xeq `mi': mixed M3_star lognli ///
        i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.occu_path#i.agegp  ///
        i.anytobacco i.depressiongp        pal sleep_hours_day beveragesalcohol ///
        energy satfat monounsaturatedfat polyunsaturatedfat  ///
        cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
        vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
        c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
        || village: || famid:, reml stddev
        mi passive: gen M3effect_`mi' =_b[lognli]
        mi xeq: sum M3effect_`mi'
}
                **In 10 MI datasets: Deriving beta from Model 3 (via M2, M3)
                forvalues mi =1/10 {
                        mi xeq `mi': mixed M2_star lognli  ///
                        i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.occu_path#i.agegp  ///
                        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
                        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
                        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
                        || village: || famid:, reml stddev
                                mi passive: gen M2effect_`mi' =_b[lognli]
                                mi xeq: sum M2effect_`mi'
                }
                                **In 10 MI datasets: Deriving beta from Model 4 (via M1, M2, M3)
                                forvalues mi =1/10 {
                                        mi xeq `mi': mixed M1_star lognli   ///
                                        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
                                        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
                                        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
```

```
                                        || village: || famid: , reml stddev
                                        mi passive: gen M1effect_`mi'=_b[lognli]
                                        mi xeq: sum M1effect_`mi'
                                }

*******************************************************************************
/* Generating 'noise' for calculating bootstrap-based CIs (1000 random draws per MI dataset)*/
*******************************************************************************

******generating noise for dataset 0 (complete data -no imputations)*****

*[load dataset]

/*saving covariance matrix of Model 1 in MI dataset 0*/
mi xeq 0: mixed SBP lognli  ///
i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp  ///
i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
energy satfat monounsaturatedfat polyunsaturatedfat  ///
cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
hdl_mol BMI crp LDL whr glucose fastinginsulin trigly_mol ///
c.hdl_mol#i.agegp c.BMI#i.agegp c.fastinginsulin#i.agegp c.trigly_mol#i.agegp ///
c.crp#c.lognli c.glucose#c.lognli  ///
i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
|| village: || famid:, reml stddev
estat vce
matrix covariance_matx_SBP_men0 = r(V)
matrix list covariance_matx_SBP_men0

/*Generating dataset with 'noise' variables based on covariance matrix of
MODEL 1 (the overall model) in MI datasets 0 (complete data -no imputations)*/
forvalues x = 0/0 {
        clear
        set seed 2389
        drawnorm x1 x3 x4 x5 x6 x8 x9 x10 x11 x12 x14 x16 x18 x24 x25 x26 x28 x29 ///
        x30 x32 x33 x34 x36 x37 x38 x41 x42 x44 x46 x47 x48 x49 x50 x51 x52 x53 ///
        x54 x55 x56 x57 x58 x59 x60 x61 x62 x63 x64 x65 x66 x67 x68 x69 x71 x72 x73 ///
        x74 x76 x77 x78 x80 x81 x82 x83 x84 x85 x86 x87 x88 x89 x90 x92 x93 x94 x96 ///
        x97 x98 x100 x101 x102 x104 x105 x106 x107 x108 x110 x111 x112 x114 x116 x117 ///
        x119 x120 x122 x123 x124 x125 x127 x128 x130 x132 x133 x135 x136 x138 ///
        x139 x141 x142 x144 x145 x147 x148 x150 x151 x153 x154 x156 x157 x159 x160 ///
        x161 x162 x163 x164, n(1000) cov(covariance_matx_SBP_men0)
        *generating variables with values 0 for the baselevels of categorical variables
        gen x2=0
        gen x7=0
        gen x13=0
        gen x15=0
        gen x17=0
        gen x19=0
        gen x20=0
        gen x21=0
        gen x22=0
        gen x23=0
        gen x27=0
        gen x31=0
        gen x35=0
        gen x39=0
        gen x43=0
        gen x45=0
        gen x70=0
        gen x75=0
        gen x79=0
        gen x91=0
        gen x95=0
        gen x99=0
        gen x103=0
        gen x109=0
        gen x113=0
```

```
		gen x115=0
		gen x118=0
		gen x121=0
		gen x126=0
		gen x131=0
		gen x134=0
		gen x137=0
		gen x140=0
		gen x143=0
		gen x146=0
		gen x149=0
		gen x152=0
		gen x155=0
		gen x158=0
		*generating x variables for omitted categories to avoid conformability error
		gen x40=0
		gen x129=0
		*Ordering noise variables by number, inspecting them and saving data
		order *, sequential
		sum *
		*[save data]
}

**********generating noise for MI dataset 1-10 **********
clear all
set matsize 11000

*[load dataset]

*saving covariance matrix of Model 1 (overall model) in MI datasets 1-10
forvalues x = 1/10 {
		mi xeq `x' :  mixed SBP lognli  ///
		i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp  ///
		i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
		energy satfat monounsaturatedfat polyunsaturatedfat  ///
		cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
		vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
		c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
		hdl_mol BMI crp LDL whr glucose fastinginsulin trigly_mol ///
		c.hdl_mol#i.agegp c.BMI#i.agegp c.fastinginsulin#i.agegp c.trigly_mol#i.agegp ///
		c.crp#c.lognli c.glucose#c.lognli  ///
		i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
		i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
		i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
		|| village: || famid:, reml stddev
		estat vce
		matrix covariance_matx_SBP_men`x' = r(V)
		matrix list covariance_matx_SBP_men`x'
}

**********
/* Generating dataset with 'noise' variables based on covariance matrices of MODEL 1
in each of the 10 MI datasets */
forvalues x = 1/10 {
		clear
		set seed 2389
		drawnorm x1 x3 x4 x5 x6 x8 x9 x10 x11 x12 x14 x16 x18 x24 x25 x26 x28 x29 ///
		x30 x32 x33 x34 x36 x37 x38 x40 x41 x42 x44 x46 x47 x48 x49 x50 x51 x52 x53 ///
		x54 x55 x56 x57 x58 x59 x60 x61 x62 x63 x64 x65 x66 x67 x68 x69 x71 x72 x73 ///
		x74 x76 x77 x78 x80 x81 x82 x83 x84 x85 x86 x87 x88 x89 x90 x92 x93 x94 x96 ///
		x97 x98 x100 x101 x102 x104 x105 x106 x107 x108 x110 x111 x112 x114 x116 x117 ///
		x119 x120 x122 x123 x124 x125 x127 x128 x129 x130 x132 x133 x135 x136 x138 ///
		x139 x141 x142 x144 x145 x147 x148 x150 x151 x153 x154 x156 x157 x159 x160 ///
		x161 x162 x163 x164, n(1000) cov(covariance_matx_SBP_men`x')
		*generating variables for the baselevels of categorical variables
		gen x2=0
		gen x7=0
		gen x13=0
		gen x15=0
		gen x17=0
		gen x19=0
```

```
		gen x20=0
		gen x21=0
		gen x22=0
		gen x23=0
		gen x27=0
		gen x31=0
		gen x35=0
		gen x39=0
		gen x43=0
		gen x45=0
		gen x70=0
		gen x75=0
		gen x79=0
		gen x91=0
		gen x95=0
		gen x99=0
		gen x103=0
		gen x109=0
		gen x113=0
		gen x115=0
		gen x118=0
		gen x121=0
		gen x126=0
		gen x131=0
		gen x134=0
		gen x137=0
		gen x140=0
		gen x143=0
		gen x146=0
		gen x149=0
		gen x152=0
		gen x155=0
		gen x158=0
		*Ordering noise variables by number, inspecting them and saving data
		order *, sequential
		sum *
		*[save 'noise dataset x']
}

/* Creating matrices from 'noise datasets' */
preserve
set matsize  11000
		forvalues x = 0/10 {
			*use ['noise dataset x' (1-10)], clear
			mkmat * , matrix(noise_men`x')
			matrix list noise_men`x'
		}
restore

		**************************************************
		/*Generating the variables (empty) for use in following steps*/

		*MI dataset 0
		*[load dataset]
		forvalues mi =0/10 {
			forvalues x=1/3{
				forvalues i=1/300 {
					gen int M`x'_star_effect_`mi'_`i'=.
					gen int M`x'_star_SE_`mi'_`i'=.
				}
			}
		}
		*updating all mi datasets to include effect and SE_0_`i' variables
		mi update

*******************************************************************************
*PATH ANALYSIS STEP 2 -Adding noise to coefficients and SEs in each of the 10 MI datasets
*******************************************************************************

/*[The following steps will to be performed in each flongsep datasets to avoid
slowing down the process from programming. Below is an example for one dataset]*/
```

```
*****MI dataset 0
*[load data]

/*running a loop creating variables holding coefficients + noise for each of the
300 random draws of noise (each row in noise datasets corresponds to a draw of noise) */
 forvalues i=1/300 {
		gen hdl_star_noise = hdl_star + (noise_men0[`i',83] * hdl_mol)
		gen BMI_star_noise = BMI_star + (noise_men0[`i',84] * BMI)
		gen crp_star_noise = crp_star + (noise_men0[`i',85] * crp)
		gen LDL_star_noise = LDL_star + (noise_men0[`i',86] * LDL)
		gen whr_star_noise = whr_star + (noise_men0[`i',87] * whr)
		gen glucose_star_noise = glucose_star + (noise_men0[`i',88] * glucose)
		gen fastinginsulin_star_noise = fastinginsulin_star  + (noise_men0[`i',89] * fastinginsulin)
		gen trigly_star_noise = trigly_star + (noise_men0[`i',90] * trigly_mol)
		gen hdlage0_star_noise = hdlage0_star + (noise_men0[`i',91] * 0.agegp#c.hdl_mol)
		gen hdlage1_star_noise= hdlage1_star + (noise_men0[`i',92] * 1.agegp#c.hdl_mol)
		gen hdlage2_star_noise=hdlage2_star + (noise_men0[`i',93] * 2.agegp#c.hdl_mol)
		gen hdlage3_star_noise=hdlage3_star + (noise_men0[`i',94] * 3.agegp#c.hdl_mol)
		gen bmiage0_star_noise=bmiage0_star + (noise_men0[`i',95] * 0.agegp#c.BMI)
		gen bmiage1_star_noise=bmiage1_star + (noise_men0[`i',96] * 1.agegp#c.BMI)
		gen bmiage2_star_noise=bmiage2_star + (noise_men0[`i',97] * 2.agegp#c.BMI)
		gen bmiage3_star_noise=bmiage3_star + (noise_men0[`i',98] * 3.agegp#c.BMI)
		gen fastinginsulinage0_star_noise= fastinginsulinage0_star + (noise_men0[`i',99] * 0.agegp#c.fastinginsulin)
		gen fastinginsulinage1_star_noise=fastinginsulinage1_star + (noise_men0[`i',100] * 1.agegp#c.fastinginsulin)
		gen fastinginsulinage2_star_noise= fastinginsulinage2_star+ (noise_men0[`i',101] * 2.agegp#c.fastinginsulin)
		gen fastinginsulinage3_star_noise=fastinginsulinage2_star + (noise_men0[`i',102] * 3.agegp#c.fastinginsulin)
		gen trigly_molage0_star_noise=trigly_molage0_star + (noise_men0[`i',103] * 0.agegp#c.trigly_mol)
		gen trigly_molage1_star_noise=trigly_molage1_star + (noise_men0[`i',104] * 1.agegp#c.trigly_mol)
		gen trigly_molage2_star_noise=trigly_molage2_star + (noise_men0[`i',105] * 2.agegp#c.trigly_mol)
		gen trigly_molage3_star_noise= trigly_molage3_star + (noise_men0[`i',106] * 3.agegp#c.trigly_mol)
		gen crpntli_star_noise=crpntli_star + (noise_men0[`i',107] * c.crp#c.lognli)
		gen glucosentli_star_noise=glucosentli_star + (noise_men0[`i',108] * c.glucose#c.lognli)
			egen M3_star_noise = rowtotal(hdl_star_noise BMI_star_noise ///
			crp_star_noise LDL_star_noise whr_star_noise glucose_star_noise ///
			fastinginsulin_star_noise trigly_star_noise ///
			hdlage0_star_noise hdlage1_star_noise hdlage2_star_noise hdlage3_star_noise ///
			bmiage0_star_noise bmiage1_star_noise bmiage2_star_noise bmiage3_star_noise ///
			fastinginsulinage0_star_noise fastinginsulinage1_star_noise ///
			fastinginsulinage2_star_noise fastinginsulinage3_star_noise ///
			trigly_molage0_star_noise trigly_molage1_star_noise ///
			trigly_molage2_star_noise trigly_molage3_star_noise ///
			crpntli_star_noise glucosentli_star_noise)
		mixed M3_star_noise lognli ///
		i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp  ///
		i.anytobacco i.depressiongp pal sleep_hours_day beveragesalcohol ///
		energy satfat monounsaturatedfat polyunsaturatedfat  ///
		cholesterol percarb sugarsweets fibertotaldietary fruitveg fish salt folate ///
		vitaminb6 vitaminb12 potassiumk vitamincascorbicacid vitamine riboflavin carotene seleniumse ///
		c.sleep_hours_day#i.agegp  c.beveragesalcohol#c.lognli c.percarb#i.agegp c.vitaminb6#i.agegp ///
		i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
		i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
		i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
		|| village: || famid:, reml stddev  iterate(20)
			replace M3_star_effect_0_`i'=_b[lognli]
			replace M3_star_SE_0_`i'=_se[lognli]
		drop hdl_star_noise BMI_star_noise crp_star_noise LDL_star_noise ///
		whr_star_noise glucose_star_noise fastinginsulin_star_noise ///
		trigly_star_noise M3_star_noise ///
		hdlage0_star_noise hdlage1_star_noise hdlage2_star_noise hdlage3_star_noise ///
		bmiage0_star_noise bmiage1_star_noise bmiage2_star_noise bmiage3_star_noise ///
		fastinginsulinage0_star_noise fastinginsulinage1_star_noise ///
		fastinginsulinage2_star_noise fastinginsulinage3_star_noise ///
		trigly_molage0_star_noise trigly_molage1_star_noise ///
		trigly_molage2_star_noise trigly_molage3_star_noise ///
		crpntli_star_noise  glucosentli_star_noise
		di "M3 dataset 0 loop `i'"
}
*[save data]

*[repeat in dataset MI 1-10]
```

```
*******************************************************************************

********* In 10 MI datasets, generating M2* with noise, leaving out M3 mediators *****

/*[The following steps will to be performed in each flongsep datasets to avoid
slowing down the process from programming. Below is an example for one dataset]*/

****MI dataset 0
*[load data]
/*running a loop creating variables holding coefficients + noise for each of
the 300 random draws (each row of noise datasets corresponds to a draw of noise) */
forvalues i=1/300 {
        gen tobacco_star_noise = tobacco_star + (noise_men0[`i',43+anytobacco])
        gen depression_star_noise  = depression_star + (noise_men0[`i',45+depressiongp])
        gen pal_star_noise = pal_star + (noise_men0[`i',47] *pal)
        gen sleep_star_noise = sleep_star + (noise_men0[`i',48]  *sleep_hours_day)
        gen alc_star_noise = alc_star + (noise_men0[`i',49] *beveragesalcohol)
        gen energy_star_noise = energy_star + (noise_men0[`i',50] *energy)
        gen satfat_star_noise = satfat_star + (noise_men0[`i',51] *satfat)
        gen mufa_star_noise = mufa_star + (noise_men0[`i',52] *monounsaturatedfat)
        gen pufa_star_noise = pufa_star + (noise_men0[`i',53] *polyunsaturatedfat)
        gen cholesterol_star_noise = cholesterol_star  + (noise_men0[`i',54] *cholesterol)
        gen percarb_star_noise = percarb_star  + (noise_men0[`i',55] *percarb)
        gen sugar_star_noise = sugar_star  + (noise_men0[`i',56] *sugarsweets)
        gen fibre_star_noise = fibre_star + (noise_men0[`i',57] *fibertotaldietary)
        gen fruitveg_star_noise = fruitveg_star  + (noise_men0[`i',58] *fruitveg)
        gen fish_star_noise = fish_star  + (noise_men0[`i',59] *fish)
        gen salt_star_noise = salt_star  + (noise_men0[`i',60] *salt)
        gen folate_star_noise = folate_star + (noise_men0[`i',61] *folate)
        gen vitaminb6_star_noise = vitaminb6_star  + (noise_men0[`i',62] *vitaminb6)
        gen vitaminb12_star_noise = vitaminb12_star  + (noise_men0[`i',63] *vitaminb12)
        gen potassium_star_noise = potassium_star + (noise_men0[`i',64] *potassiumk)
        gen vitc_star_noise = vitc_star + (noise_men0[`i',65] *vitamincascorbicacid)
        gen vitamine_star_noise = vitamine_star  + (noise_men0[`i',66] *vitamine)
        gen riboflavin_star_noise = riboflavin_star  + (noise_men0[`i',67] *riboflavin)
        gen carotene_star_noise = carotene_star + (noise_men0[`i',68] *carotene)
        gen seleniumse_star_noise = seleniumse_star + (noise_men0[`i',69] *seleniumse)
        gen sleepage0_star_noise=sleepage0_star + (noise_men0[`i', 70] *0.agegp#c.sleep_hours_day)
        gen sleepage1_star_noise=sleepage1_star + (noise_men0[`i', 71] *1.agegp#c.sleep_hours_day)
        gen sleepage2_star_noise=sleepage2_star + (noise_men0[`i', 72] *2.agegp#c.sleep_hours_day)
        gen sleepage3_star_noise=sleepage3_star + (noise_men0[`i', 73] *3.agegp#c.sleep_hours_day)
        gen alcntli_star_noise=alcntli_star + (noise_men0[`i',74] * c.beveragesalcohol#c.lognli)
        gen percarbage0_star_noise=percarbage0_star + (noise_men0[`i', 75] *0.agegp#c.percarb)
        gen percarbage1_star_noise=percarbage1_star + (noise_men0[`i', 76] *1.agegp#c.percarb)
        gen percarbage2_star_noise=percarbage2_star + (noise_men0[`i', 77] *2.agegp#c.percarb)
        gen percarbage3_star_noise=percarbage3_star + (noise_men0[`i', 78] *3.agegp#c.percarb)
        gen vitaminb6age0_star_noise=vitaminb6age0_star + (noise_men0[`i',79] *0.agegp#c.vitaminb6)
        gen vitaminb6age1_star_noise=vitaminb6age1_star + (noise_men0[`i', 80] *1.agegp#c.vitaminb6)
        gen vitaminb6age2_star_noise=vitaminb6age2_star + (noise_men0[`i', 81] *2.agegp#c.vitaminb6)
        gen vitaminb6age3_star_noise=vitaminb6age3_star + (noise_men0[`i', 82] *3.agegp#c.vitaminb6)
                egen M2_star_noise = rowtotal(tobacco_star_noise depression_star_noise ///
                pal_star_noise sleep_star_noise alc_star_noise energy_star_noise satfat_star_noise ///
                mufa_star_noise pufa_star_noise cholesterol_star_noise percarb_star_noise ///
                sugar_star_noise fibre_star_noise fruitveg_star_noise fish_star_noise ///
                salt_star_noise folate_star_noise vitaminb6_star_noise vitaminb12_star_noise ///
                potassium_star_noise vitc_star_noise vitamine_star_noise riboflavin_star_noise ///
                carotene_star_noise seleniumse_star_noise ///
                sleepage0_star_noise sleepage1_star_noise ///
                sleepage2_star_noise sleepage3_star_noise ///
                alcntli_star_noise percarbage0_star_noise percarbage1_star_noise ///
                percarbage2_star_noise percarbage3_star_noise ///
                vitaminb6age0_star_noise vitaminb6age1_star_noise ///
                vitaminb6age2_star_noise vitaminb6age3_star_noise)
        mixed M2_star_noise lognli ///
        i.edu zpc2 n_hh i.occu_path i.shiftwork i.openfire i.solidfuel i.occu_path#i.agegp ///
        i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
        i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
        i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
        || village: || famid:, reml stddev  iterate(20)
```

```
                replace M2_star_effect_0_`i'=_b[lognli]
                replace M2_star_SE_0_`i'=_se[lognli]
        drop tobacco_star_noise depression_star_noise ///
        pal_star_noise sleep_star_noise alc_star_noise energy_star_noise satfat_star_noise ///
        mufa_star_noise pufa_star_noise cholesterol_star_noise percarb_star_noise ///
        sugar_star_noise fibre_star_noise fruitveg_star_noise fish_star_noise ///
        salt_star_noise folate_star_noise vitaminb6_star_noise vitaminb12_star_noise ///
        potassium_star_noise vitc_star_noise vitamine_star_noise riboflavin_star_noise ///
        carotene_star_noise seleniumse_star_noise sleepage0_star_noise sleepage1_star_noise ///
        sleepage2_star_noise sleepage3_star_noise alcntli_star_noise percarbage0_star_noise ///
        percarbage1_star_noise percarbage2_star_noise percarbage3_star_noise ///
        vitaminb6age0_star_noise vitaminb6age1_star_noise vitaminb6age2_star_noise ///
        vitaminb6age3_star_noise M2_star_noise
                di "M2 dataset 0 loop `i'"
}
*[save data]

*[repeat in dataset MI 1-10]

*******************************************************************************

********* In 10 MI datasets, generating M1* with noise, leaving out M2 and M3 mediators *****
/*[The following steps will to be performed in each flongsep datasets to avoid
slowing down the process from programming. Below is an example for one dataset]*/

*****MI dataset 0
*[load data]

/*running a loop creating variables holding coefficients + noise for each
of the 300 random draws (each row in noise datasets corresponds to a draw of noise) */
forvalues i=1/300 {
        gen edu_star_noise = edu_star + (noise_men0[`i',2+edu])
        gen ses_star_noise = ses_star + (noise_men0[`i', 5] *zpc2)
        gen n_hh_star_noise = n_hh_star + (noise_men0[`i',6] *n_hh)
        gen occu_star_noise = occu_star + (noise_men0[`i',7+occu_path])
        gen shiftwork_star_noise =shiftwork_star + (noise_men0[`i',13+shiftwork])
        gen openfire_star_noise = openfire_star + (noise_men0[`i',15+openfire])
        gen solidfuel_star_noise= solidfuel_star + (noise_men0[`i',17+solidfuel])
        gen occuage00_star_noise=occuage00_star + (noise_men0[`i',19] * 0.occu_path#0.agegp)
        gen occuage01_star_noise=occuage01_star + (noise_men0[`i',20] * 0.occu_path#1.ageg)
        gen occuage02_star_noise=occuage02_star + (noise_men0[`i',21] * 0.occu_path#2.agegp)
        gen occuage03_star_noise=occuage03_star + (noise_men0[`i',22] * 0.occu_path#3.agegp)
        gen occuage10_star_noise=occuage10_star + (noise_men0[`i',23] * 1.occu_path#0.agegp)
        gen occuage11_star_noise=occuage11_star + (noise_men0[`i',24] * 1.occu_path#1.agegp)
        gen occuage12_star_noise=occuage12_star + (noise_men0[`i',25] * 1.occu_path#2.agegp)
        gen occuage13_star_noise=occuage13_star + (noise_men0[`i',26] * 1.occu_path#3.agegp)
        gen occuage20_star_noise=occuage20_star + (noise_men0[`i',27] * 2.occu_path#0.agegp)
        gen occuage21_star_noise=occuage21_star + (noise_men0[`i',28] * 2.occu_path#1.agegp)
        gen occuage22_star_noise=occuage22_star + (noise_men0[`i',29] * 2.occu_path#2.agegp)
        gen occuage23_star_noise=occuage23_star + (noise_men0[`i',30] * 2.occu_path#3.agegp)
        gen occuage30_star_noise=occuage30_star + (noise_men0[`i',31] * 3.occu_path#0.agegp)
        gen occuage31_star_noise=occuage31_star + (noise_men0[`i',32] * 3.occu_path#1.agegp)
        gen occuage32_star_noise=occuage32_star + (noise_men0[`i',33] * 3.occu_path#2.agegp)
        gen occuage33_star_noise=occuage33_star + (noise_men0[`i',34] * 3.occu_path#3.agegp)
        gen occuage40_star_noise=occuage40_star + (noise_men0[`i',35] * 4.occu_path#0.agegp)
        gen occuage41_star_noise=occuage41_star + (noise_men0[`i',36] * 4.occu_path#1.agegp)
        gen occuage42_star_noise=occuage42_star + (noise_men0[`i',37] * 4.occu_path#2.agegp)
        gen occuage43_star_noise=occuage43_star + (noise_men0[`i',38] * 4.occu_path#3.agegp)
        gen occuage50_star_noise=occuage50_star + (noise_men0[`i',39] * 5.occu_path#0.agegp)
        gen occuage51_star_noise=occuage51_star + (noise_men0[`i',40] * 5.occu_path#1.agegp)
        gen occuage52_star_noise=occuage52_star + (noise_men0[`i',41] * 5.occu_path#2.agegp)
        gen occuage53_star_noise=occuage53_star + (noise_men0[`i',42] * 5.occu_path#3.agegp)
                egen M1_star_noise = rowtotal(edu_star_noise ses_star_noise n_hh_star_noise ///
                occu_star_noise shiftwork_star_noise openfire_star_noise solidfuel_star_noise ///
                occuage00_star_noise occuage01_star_noise occuage02_star_noise occuage03_star_noise ///
                occuage10_star_noise occuage11_star_noise occuage12_star_noise occuage13_star_noise ///
                occuage20_star_noise occuage21_star_noise occuage22_star_noise occuage23_star_noise ///
                occuage30_star_noise occuage31_star_noise occuage32_star_noise occuage33_star_noise ///
                occuage40_star_noise occuage41_star_noise occuage42_star_noise occuage43_star_noise ///
                occuage50_star_noise occuage51_star_noise occuage52_star_noise occuage53_star_noise)
        mixed M1_star_noise lognli ///
```

```
		i.agegp i.married i.season i.birthseason i.caste i.religion bproomtemp ///
		i.moth_diab i.moth_bp i.moth_stroke i.chd_mum i.moth_obese ///
		i.fath_diab i.fath_bp i.fath_stroke i.fath_chd i.fath_obese ///
		|| village: || famid:, reml stddev  iterate(20)
			replace M1_star_effect_0_`i'=_b[lognli]
			replace M1_star_SE_0_`i'=_se[lognli]
		drop edu_star_noise ses_star_noise n_hh_star_noise ///
		occu_star_noise shiftwork_star_noise openfire_star_noise solidfuel_star_noise ///
		occuage00_star_noise occuage01_star_noise occuage02_star_noise occuage03_star_noise ///
		occuage10_star_noise occuage11_star_noise occuage12_star_noise occuage13_star_noise ///
		occuage20_star_noise occuage21_star_noise occuage22_star_noise occuage23_star_noise ///
		occuage30_star_noise occuage31_star_noise occuage32_star_noise occuage33_star_noise ///
		occuage40_star_noise occuage41_star_noise occuage42_star_noise occuage43_star_noise ///
		occuage50_star_noise occuage51_star_noise occuage52_star_noise occuage53_star_noise M1_star_noise
			di "M1 dataset 0 loop `i'"
}
*[save data]


*******************************************************************************
*Data management pre Step 3
*******************************************************************************
/*As in step 2 we need to generate all mi dataset-specific variables in all MI
datasets to avoid saved estimates being dropped when converting format */

*MI dataset 0
*[load data]
forvalues mi =0/10 {
		forvalues x=1/3{
			gen mean_M`x'_star_effect_`mi'=.
			gen sum_diff_sq_M`x'_`mi'=.
			gen samp_variance_M`x'_`mi'=.
			gen avr_variance_M`x'_star_`mi'=.
			gen path_M`x'_variance_`mi'=.
			gen path_M`x'_se_`mi'=.
				forvalues i=1/300 {
					gen diff_sq_M`x'_`mi'_`i'=.
					gen variance_M`x'_star_`mi'_`i'=.
					gen diff_M`x'_star_effect_`mi'_`i'=.
				}
		}
*[save data], replace
}

*updating all MI datasets to include the dataset 0-specific variables
mi update

*******************************************************************************
*PATH ANALYSIS STEP 3
/*In each of 10 MI dataset and for each of the paths: calculating sample
variance and variance of the 300 noise beta-coefficients (_b)*/
*******************************************************************************

/*[The following steps will to be performed in each flongsep datasets to avoid
slowing down the process from programming. Below is an example for one dataset]*/

*******MI DATASET 0

*[loading MI dataset 0]

*** SAMPLE VARIANCE (variance of noise _b estimates (n=300))

*Generating variable holding the sample variance
forvalues x = 1/3{
		*Ordering variables in alphabetical order for programming reasons
		order *, sequential
		*generating variable holding the average of the 300 estimates of _b for M1-3*
		drop mean_M`x'_star_effect_0
		egen  mean_M`x'_star_effect_0 = rowmean(M`x'_star_effect_0_1-M`x'_star_effect_0_300)
		/*generating the difference from the mean: subtracting each of the
		300 noise _bs from the average of the noise _bs  */
```

```
		forvalues i=1/300 {
			replace diff_M`x'_star_effect_0_`i' = M`x'_star_effect_0_`i' - mean_M`x'_star_effect_0
			*generate variable holding 'the difference from the mean, squared'
			replace diff_sq_M`x'_0_`i' = diff_M`x'_star_effect_0_`i' ^2
			di "Difference from the mean, mi dataset 0, M`x', noise loop `i'"
		}
}

*Summing the 'difference from the mean, squared' and dividing by n-1
forvalues x = 1/3{
		*Ordering variables in alphabetical order for programming reasons
		order *, sequential
		*generating variable holding the sum of derived differences from the mean squared
		drop sum_diff_sq_M`x'_0
		egen  sum_diff_sq_M`x'_0 = rowtotal(diff_sq_M`x'_0_1-diff_sq_M`x'_0_300)
		*generating sample variance (sum of diff sq/(n-1)
		replace samp_variance_M`x'_0 =sum_diff_sq_M`x'_0 / (300-1)
		di "Sample variance mi dataset 0, M`x'"
}

*** VARIANCE OF ESTIMATES (sum of SQUAREED SEs)

*Generating variable holding the variance (SE  squared) of each of the 300 noise _bs
forvalues x = 1/3{
		forvalues i=1/300{
			replace variance_M`x'_star_0_`i' = M`x'_star_SE_0_`i' ^2
			di "Variance of beta estimates (SEs^2), mi dataset 0, M`x', noise loop `i'"
		}
}
*generating variable holding the average variance across the 300 estimates
forvalues x = 1/3{
		order *, sequential
		drop avr_variance_M`x'_star_0
		egen  avr_variance_M`x'_star_0 = rowmean(variance_M`x'_star_0_1-variance_M`x'_star_0_300)
		di "Average variance from 300 additions of noise, mi dataset 0 M`x'"
}

*[save dataset 0]
save Step3_MIpmm_SBPm_workcopy_EM, replace


*******************************************************************************
*DATA MANAGEMENT pre step 4 to enable pooling estimates across the 10 MI datasets:
*******************************************************************************
*[load data]
*convert data to mlong format for max speed in next steps where mi commands are needed

/*I generated mi dataset specific variables in all mi datasets, however
the estimates were only saved in the mi dataset (identified by _mi_m code)
in which they were generated. To ease programming, I will carry forward
(fill missing with the appropriate estimate) to all _mi_m codes the mi dataset-specific
estimates I need to calculate the final SEs (variable names will
identify the mi dataset in which they were originally calculated)*/

*installing carryforward (by David Kantor)
ssc install carryforward

forvalues mi =0/10 {
		forvalues x=1/3{
				sort path_M`x'_variance_`mi'
				carryforward path_M`x'_variance_`mi', gen (path_M`x'_variance_`mi'Q)
				replace path_M`x'_variance_`mi' = path_M`x'_variance_`mi'Q
				di "Path variance carried forward from dataset `mi', M`x'"
		}
}
*[save data]

*******************************************************************************
*PATH ANALYSIS STEP 4
*******************************************************************************
/* Pooling MI dataset-specific sample variance and variance of estimates to produce
```

```
overall path specific SEs */
*******************************************************************************

*[load mlong file (to enable pooling estimates across MI datasets)]

*** POOLED SAMPLE VARIANCE (variance of MI dataset-specific and path-specific _bs)
*Generating variable holding the sample variance (variance of _b estimates) across 10 MI datasets
forvalues x = 1/3{
	*Ordering variables in alphabetical order for programming reasons
	order *, sequential
	*generating variable holding the average of the 10 mi dataset-specific _b for M1-3*
	mi passive: egen mean_M`x'effect_mispec = rowmean(M`x'effect_1-M`x'effect_10)
	di "Mean effect across 10 mi datasets, M`x'"
}

/*generating the difference from the mean: subtracting each of the 10 mi
dataset-specific _b from the average of the 10 _bs*/
forvalues mi =1/10 {
	forvalues x = 1/3{
		mi passive: gen diff_mean_M`x'effect_mispec_`mi' = ///
		M`x'effect_`mi' - mean_M`x'effect_mispec
		*display the mi dataset # to be able to follow the process
		di "Difference from the mean effect (across 10 mi datasets), mi dataset `mi', M`x'"
	}
}

*generate variable holding 'the difference from the mean, squared'
forvalues mi =1/10 {
	forvalues x = 1/3{
		mi passive: gen diff_m_sq_M`x'_b`mi' = diff_mean_M`x'effect_mispec_`mi' ^2
		*display the mi dataset # to be able to follow the process
		di "Difference from the mean effect (across 10 mi datasets) squared, mi dataset `mi', M`x'"
	}
}

*Summing the 'difference from the mean, squared' and dividing by n-1
forvalues x = 1/3{
	order *, sequential
	mi passive: egen sum_diff_m_sq_M`x'_mispec = ///
	rowtotal(diff_m_sq_M`x'_b1-diff_m_sq_M`x'_b10)
	*generating the sample variance of the 10 mi dataset spec betas
	mi passive: gen samp_variance_M`x'_mispec = sum_diff_m_sq_M`x'_mispec /(1-10)
	*Penalising estimate due to low number of repetitions (n=10)
	mi passive: gen p_samp_variance_M`x'_mispec = ///
	samp_variance_M`x'_mispec + (samp_variance_M`x'_mispec /10)
	di "Sum of diff from the mean effect (across 10 mi datasets)  squared, M`x'"
}


*** POOLED VARIANCE OF MI DATASET SEPCIFIC ESTIMATES ***
*ordering variables and generating the average variance across the 10 MI datasets (noise is included)
forvalues x = 1/3{
	order *, sequential
	mi passive: egen avr_variance_M`x'_SE300 = ///
	rowmean(path_M`x'_variance_1-path_M`x'_variance_10)
}


*******************************************************************************
*PATH ANALYSIS STEP 5

/*calculating FINAL SEs and 95% confidence intervals of final estimates,
pooled across MI dataset estimates*/
*******************************************************************************
**** FINAL POOLED SE OF THE PATHWAYS VIA M1-M3 ***

/*generating final variance by adding the average variance and the penalised
sample variance (incl. noise and mi variation)*/
forvalues x = 1/3{
	mi passive: egen final_variance_M`x'= ///
	rowtotal(avr_variance_M`x'_SE300 p_samp_variance_M`x'_mispec)
```

```
		*generating final SE = sqrt of the final variance
		mi passive: gen SE_M`x' = sqrt(final_variance_M`x')
}

***CALCULATING 95% CIs FOR PATHWAYS    M1-M3

forvalues x = 1/3{
*recalling the indirect direct effects via M1-M3
sum M`x'effect
*generating lower and upper confidence boundaries for paths
mi passive: gen M`x'_lower = M`x'effect - (1.96*SE_M`x')
mi passive: gen M`x'_upper = M`x'effect + (1.96*SE_M`x')
*displaying estimated effect of the path via Mx and 95% CI
list M`x'effect M`x'_lower M`x'_upper in 2/4
}
```

Appendix V Age-specific associations of mediators with SBP

Table A2 Age-specific associations of mediators with SBP (where test for interaction was significant at the 5% level), men

	Fully adjusted model predicted mean SBP change with increase in level of mediators*											
	age 18-29			age 30-39			age 40-59			age 50 +		
	n**	β (95% CI)	p-value	n**	β (95% CI)	p-value	n**	β (95% CI)	p-value	n**	β (95% CI)	p-value
Socio demographic factors (mediators)												
Occupation, n (%)												
Agriculture	148 (8.8)	reference		35 (16.4)	reference		129 (33.9)	reference		280 (34.2)	reference	
Daily wage earner	207 (12.2)	0.4 (2.7, 3.5)	0.78	46 (21.5)	0.1 (6.3, 6.5)	0.97	165 (43.3)	0.9 (2.5, 4.3)	0.6	378 (46.1)	3.1 (0.8, 5.4)	0.007
Unskilled manual worker	71 (4.2)	2 (2.2, 6.1)	0.35	9 (4.2)	-0.4 (10.9, 10.1)	0.35	15 (3.9)	1.4 (6.3, 9.1)	0.73	39 (4.8)	10.5 (5.6, 15.3)	<0.001
Self-employed	588 (34.8)	3.1 (0.4, 5.7)	0.03	94 (43.9)	3.4 (2.2, 9.1)	0.24	50 (13.1)	7.5 (2.8, 12.3)	0.002	53 (6.5)	11.0s (6.8, 15.3)	<0.001
Salaried worker	200 (11.8)	4.3 (1.2, 7.5)	0.007	26 (12.2)	3.8 (3.5, 11.2)	0.31	14 (3.7)	15.0 (7, 23.1)	<0.001	24 (2.9)	2.5 (3.5, 8.5)	0.42
student or unemployed	477 (28.2)	0.6 (2.2, 3.4)	0.66	4 (1.9)	-2.2 (17.2, 12.7)	0.77	8 (2.1)	12.2 (1.8, 22.5)	0.02	46 (5.6)	11.2 (6.7, 15.8)	<0.001
Lifestyle and mental health factors (mediators)												
Sleep (h/d)	1691 (54.4)	-0.1 (-0.8, 0.5)	0.689	214 (6.9)	0.6 (-1.4, 2.6)	0.564	381 (12.3)	1.2 (-0.2, 2.6)	0.102	820 (26.4)	2.0 (1.1, 2.8)	<0.001
Percent carbohydrates of total energy	1692 (54.4)	-0.03 (-0.1, 0.1)	0.53	215 (6.9)	-0.1 (-0.3, 0.1)	0.34	382 (12.3)	-0.1 (-0.2, -0.02)	0.02	821 (26.4)	-0.01 (-0.1, 0.07)	0.741
Vitamin b6	1693 (54.4)	0.7 (-0.1, 1.6)	0.09	216 (6.9)	0.6 (-1.9, 3.1)	0.64	383 (12.3)	-0.1 (-2.8, 2.6)	0.97	822 (26.4)	-3.7 (-5.7, -1.8)	<0.001
Metabolic factors (mediators)												
Body mass index (kg/m^2)	1695 (54.4)	1.4 (1.2, 1.6)	<0.001	218 (6.9)	0.8 (0.3, 1.3)	0.001	385 (12.3)	1.8 (1.4, 2.2)	<0.001	824 (26.4)	1.3 (1.1, 1.6)	<0.001
High-density lipoprotein (mmol/l)	1694 (54.4)	0.8 (-1.5, 3.2)	0.49	217 (6.9)	1.2 (-4.5, 6.9)	0.69	384 (12.3)	-4 (-7.8, -0.1)	0.05	823 (26.4)	4.6 (2.0, 7.2)	0.001
Serum fasting insulin mmol/l	1696 (54.4)	0.1 (0.1, 0.2)	0.001	219 (6.9)	0.4 (0.1, 0.7)	0.02	386 (12.3)	0.6 (0.3, 0.8)	<0.001	825 (26.4)	0.4 (0.1, 0.6)	0.004
Triglycerides mmol/l	1697 (54.4)	3.7 (2.7, 4.8)	<0.001	220 (6.9)	3.5 (1.2, 5.7)	0.003	387 (12.3)	6.5 (4.6, 8.3)	<0.001	826 (26.4)	2.1 (0.8, 3.4)	0.002

β - beta-coefficient; CI - confidence interval; n - sample size; SBP - systolic blood pressure; sd - standard deviation (continued next page)

Model predicted β-coefficients (95% CIs) and p-values were obtained from mixed-effects linear regression models with clustering by household and village and adjusted for age, marital status, caste, religion, season of survey, season of birth; parental history of obesity, diabetes, hypertension, stroke and coronary heart disease; and room temperature at measurement of blood pressure

*Mean SBP change per one unit (integer) increase in continuous mediators or mean SBP change from base level to comparator level for binary and categorical mediators

**Sample sizes for categorical variables were reported for multiple imputations dataset 1

Table A3 Age-specific associations of mediators with SBP (where test for interaction was significant at the 5% level), women (n=2734)

	Fully adjusted model predicted mean SBP change with increase in level of mediators*											
	age 18-29			**age 30-39**			**age 40-59**			**age 50 +**		
	n*	**β (95% CI)**	**p-value**	**n****	**β (95% CI)**	**p-value**	**n****	**β (95% CI)**	**p-value**	**n****	**β (95% CI)**	**p-value**
Socio demographic factors (mediators)												
SES	1087 (39.8)	0.5 (-0.2, 1.3)	0.17	322 (11.8)	1.2 (-0.2, 2.7)	0.10	948 (34.7)	1.1 (0.2, 2)	0.02	377 (13.8)	-0.7 (-2.1, 0.7)	0.33
Low	297 (27.3)	reference		117 (36.3)	reference		319 (33.7)	reference		180 (47.8)	reference	
Medium	329 (30.3)	1 (-1, 3)	0.35	111 (34.5)	0.7 (-2.8, 4.3)	0.68	351 (37)	3.5 (1.5, 5.6)	0.001	121 (32.1)	0.1 (-3.2, 3.5)	0.95
High	461 (42.4)	1.4 (-0.5, 3.4)	0.14	94 (29.2)	1.8 (-1.9, 5.4)	0.34	278 (29.3)	2.9 (0.8, 5.1)	0.01	76 (20.2)	0.5 (-3.3, 4.4)	0.79
Occupation												
Agriculture	85 (7.8)	reference		49 (15.2)	reference		215 (22.7)	reference		67 (17.8)	reference	
Daily wage earner	196 (18)	-0.5 (-3.7, 2.7)	0.75	171 (53.1)	0.1 (-3.9, 4.2)	0.95	509 (53.7)	0.1 (-1.9, 2.2)	0.89	195 (51.7)	1.2 (-2.4, 4.7)	0.53
Unskilled manual worker	478 (44)	0 (-2.9, 2.9)	0.98	84 (26.1)	1.3 (-3.2, 5.8)	0.57	190 (20)	4.1 (1.6, 6.6)	0.00	98 (26)	3.8 (-0.2, 7.8)	0.06
Self-employed	76 (7)	-1.4 (-5.3, 2.6)	0.50	12 (3.7)	2.2 (-5.7, 10.1)	0.58	26 (2.7)	-1.9 (-7.1, 3.3)	0.46	7 (1.9)	0.2 (-10.1, 10.4)	0.98
Salaried worker	203 (18.7)	-2.3 (-6.8, 2.2)	0.31	1 (0.3)	-1.8 (-13.3, 9.7)	0.76	3 (0.3)	-0.7 (-12, 10.7)	0.91	8 (2.1)	0.7 (-17.1, 18.4)	0.94
student or unemployed	1087 (100)	-2.5 (-5.9, 0.8)	0.14	322 (100)	-18.1 (-42.2, 6)	0.14	948 (100)	6.3 (-8, 20.5)	0.39	377 (100)	29.3 (19.5, 39)	<0.001
Do shift work												

	Fully adjusted model predicted mean SBP change with increase in level of mediators*											
		age 18-29			age 30-39			age 40-59			age 50 +	
	n*	β (95% CI)	p-value	n**	β (95% CI)	p-value	n**	β (95% CI)	p-value	n**	β (95% CI)	p-value
Indoor open fire (6 months during lifetime)	1010 (92.9)	-1.4 (-4.4, 1.5)	0.34	301 (93.5)	-3.5 (-9.2, 2.2)	0.23	917 (96.7)	-5.3 (-9.9, -0.6)	0.03	371 (98.4)	14.1 (3.6, 24.5)	0.01
Use solid fuel for cooking												
Number of household members	1087 (39.8)	-0.1 (-0.6, 0.5)	0.78	322 (11.8)	-0.1 (-1, 0.7)	0.78	948 (34.7)	-0.3 (-0.9, 0.3)	0.35	377 (13.8)	1.3 (0.6, 2)	<0.001
Lifestyle and mental health factors (mediators)												
Use tobacco	4 (0.4)	-9.3 (-21.8, 3.1)	0.14	17 (5.3)	8.9 (2.5, 15.3)	0.006	180 (19)	1.6 (-0.5, 3.7)	0.125	153 (40.6)	-1.8 (-4.4, 0.9)	0.187
Sleep (h/d)	1087 (39.8)	-0.1 (-0.8, 0.6)	0.77	322 (11.8)	-0.2 (-1.6, 1.2)	0.78	948 (34.7)	1 (0.2, 1.9)	0.01	377 (13.8)	1.9 (0.8, 3.1)	0.001

β - beta-coefficient; CI - confidence interval; n - sample size; SBP - systolic blood pressure; sd - standard deviation

Model predicted β-coefficients (95% CIs) and p-values were obtained from mixed-effects linear regression models with clustering by household and village and adjusted for age, marital status, caste, religion, season of survey, season of birth; parental history of obesity, diabetes, hypertension, stroke and coronary heart disease; and room temperature at measurement of blood pressure

*Mean SBP change per unit (integer) increase in continuous mediators or mean SBP change from base level to comparator level for binary and categorical mediators

**Sample sizes for categorical variables were reported for multiple imputations dataset 1

CHAPTER 6 DISCUSSION

6.1 Summary of PhD findings and unanswered questions

This PhD examined the links between place of residence and risk factors for CVD in South India. India has undergone rapid urbanisation and shifts in labour-force out of traditional economic activities, such as agriculture, and concurrently seen steep rises in premature deaths and disability from CVDs, particularly ischemic heart disease (IHD) and stroke.[1-4] Incidence and prevalence of CVDs vary considerably between India's 29 states with the highest estimated burdens in the South.[1, 5] Ischemic heart disease, India's number one killer, has made the largest contribution to the rise in age-standardised disease prevalence since 1990 in India and the prevalence of hypertension has increased 30-fold during the same period. [2, 6] Large variations in the prevalence of hypertension and CVD are reported by level of urbanisation from less than 10% in rural to >30% in urban India.[14, 15] Concerns have been expressed that low-and middle-income countries (LMICs), including India, will be unable to provide the growing urban population, and especially rural migrants, with 'decent' and 'fair' employment, with potential adverse consequences for cardiovascular health.[7] In order to curb expected rises in CVD burdens across India with continued urbanisation, it is becoming increasingly important to understand the association of place of residence with CVDs as well as identify and quantify the most important pathways.

This thesis addressed three objectives, repeated below, in three connected research papers.

1. **S**ystematically review the available published evidence on the associations of engaging in agriculture compared with types of non-agricultural employment (across the urbanisation continuum) with CVD incidence, prevalence and CVD risk factors in LMICs
2. Explore associations of night-time light intensity (NTLI), a novel, globally comparable, continuous proxy for level of urbanisation, with mean systolic blood pressure (SBP), body mass index (BMI), low-density lipoprotein (LDL) and fasting plasma glucose (FPG) among adults at early stages of urbanisation in Telangana, South India
3. Decompose the total effect of level of urbanisation on mean SBP into direct and indirect effects via hypothesised pathways, among adults at early stages of urbanisation in Telangana, South India

6.1.1 Addressing objective 1

The first research paper of this thesis (see Chapter 3) systematically reviewed the available published evidence on the associations of engaging in agriculture compared to other types of employment with CVD incidence prevalence and CVD risk factors in LMICs. This objective was achieved through a systematic literature review conducted in four electronic databases and supplemented by bibliography, google and google scholar searches. Narrative analysis was

provided for two studies from India and Vietnam that reported the primary outcome of interest, atherosclerotic CVDs. Thirteen further research articles from five LMICs were included that reported secondary outcomes of interest namely atherosclerotic CVD risk factor. High quality evidence was lacking, particularly on the primary outcomes, CVD events. However, the available evidence suggested that, compared with non-agricultural workers mainly living in urban areas, agricultural workers living in rural areas had a lower prevalence of some important CVD risk factors (hypertension and BMI) but higher prevalence of others (smoking and low BMI). Most studies sampled different types of labour from different locations (e.g. rural agricultural workers and urban government workers) and associations of type of labour could not be separated from those of location of residence.

Seven of the 13 included studies were from India; five were set in rural areas and two included participants from rural and urban areas. Four studies reported on overweight or obesity, four on underweight, four on hypertension and one on tobacco use. Both male and female agricultural workers (analysed together and separately) were less likely to be overweight and obese than all other types of rural and urban workers they were compared to (for example house workers, business and sedentary workers), except for male manual workers in one study[8] (Figure 6.1). Agricultural workers were generally more likely to be underweight than types of non-agricultural rural and urban workers (e.g. business, government and non-government worker), however two studies did not observe differences between agriculturalists and self-employed women[9]; a group of non-agricultural men and woman[10]; or rural students, unemployed and retired sedentary men and women[9]. The four studies that reported on prevalence of hypertension observed few differences between agricultural and different types of rural and urban non-agricultural workers (e.g. daily wage earners, business, unemployed and house-wives). Only one study, which included both rural and urban participants, reported a higher likelihood of hypertension among retired sedentary workers and a lower likelihood among manual than agricultural workers. No difference was observed for the prevalence of tobacco use in the single study that compared agricultural to business workers in rural India. In summary, the evidence from these seven studies suggested that in India, agricultural workers were less likely to be overweight and obese than a range of different types of non-agricultural workers in both rural and urban settings. Agricultural workers were additionally less likely to be underweight than most other types of workers in rural and urban areas. The evidence on hypertension was inconclusive and studies on tobacco use were lacking from India.

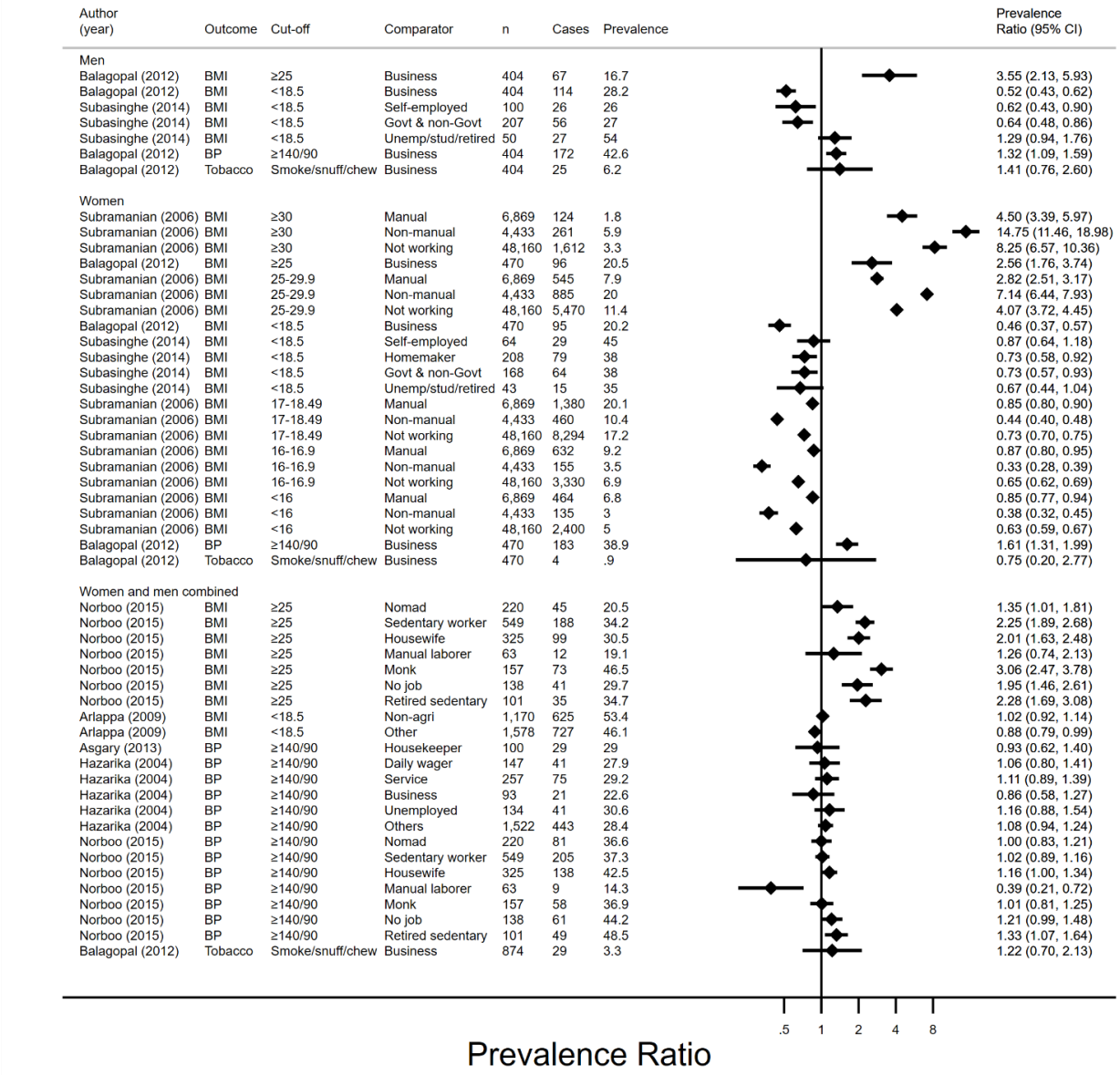


Figure 6.1 Prevalence ratios and 95% confidence intervals of overweight, underweight, hypertension and tobacco use among non-agricultural compared to agricultural workers in seven Indian studies

Agri – agriculture, BMI – body mass index, BP – Blood pressure, CI – confidence interval, Govt – government, n – sample size, PR – prevalence ratio, stud – student

Prevalence ratios were derived from comparing each non-agricultural group (coded 1) to the agricultural group (coded 0)

Only one eligible publication[11] was identified that assessed associations of shifts away from agriculture and into other types of (unspecified) employment (and locations) with CVD risk factors. In this Chinese study assessing the prevalence of smoking over a 10 year period, the prevalence declined among men, who shifted out of agriculture and remained unchanged among

men who remained in agriculture, factory or office work. The prevalence of smoking did not differ between the four groups at the 10-year follow-up.

Substantial methodological shortcomings were identified in all reviewed publications from risk of bias assessment in seven domains (confounding, selection of study participants, measurement of exposures/interventions, departures from intended interventions, missing data, measurement of outcomes and selection of the reported results). The overall quality of evidence rating across the seven domains indicated that risk of bias was 'poorly addressed' in all 13 included studies. Main concerns included that i) none of the studies described exposure, comparator(s) or outcome(s) in sufficient detail to rate the risk of bias relating to these domains; ii) selection of participants was poorly addressed in all included studies; and iii) only one study adjusted appropriately for confounders, i.e. adjusted for confounders without also including mediators in non-causal models. The lack of adjustment for important confounders, such as age and sex (as a minimum), education, season of survey, etc. made it difficult to draw any conclusions from the analyses, particularly where the characteristics of exposure and comparator group likely differed substantially, for example as a result of sampling the exposure and comparator groups from widely different settings.

With continued facilitation of labour-force shifts away from agriculture and into types of non-agricultural employment during urbanisation and projected rises in CVD burdens in LMICs, it is becoming increasingly important to understand the links between occupation patterns, labour-force transitions and NCDs. My exploration of the associations of 'current labour' with CVDs is only a part of the puzzle and we cannot extrapolate the results to infer impacts of labour-force shifts with CVD risk factors. That is, shifting away from (rural) agriculture and into other types of (urban) employment does not necessarily translate into greater likelihood of hypertension and overweight and reduced likelihood of tobacco use and underweight in LMICs.

Hypothesised pathways linking current engagement in agriculture vs. more urbanised employment with CVD risk factors are usually concerned with differences in the characteristics of particular types of labour (e.g. wages, exposure to environmental hazards or physical workload). For instance, a described pathway linking occupation (agriculture vs. more urbanised employment) with CVD risk factors is via productivity and income characteristics of different occupations. That is, non-agricultural employment in more urbanised areas, compared to engaging in agriculture, is associated with higher productivity and wages.[4] Greater wealth combined with exposure to urban environments is associated with improved access to a number of goods and services that can be both protective and deleterious to cardiovascular health, e.g. fresh or processed food, alcohol, tobacco, motor vehicle ownership and health care.[12-16] Another example is the contribution of reduced reliance on labour- and energy-intensive agriculture and

an increase in urban sedentary work (e.g. assembly line or office work) to observed large reductions in physical activity across LMICs.[4, 15, 17, 18] Although it should be noted that the development and increased use of labour saving technologies in agriculture e.g. mechanisation of agriculture, may also reduce work-related physical activity in rural areas.[15, 17, 18]

There may be added dimensions of vulnerability affecting people who shift out of agriculture, as opposed to people who are currently in various types of non-agricultural employment, and migrants to urban areas who leave behind their home and social network may be particularly affected.[7, 19] The increasing demand for non-agricultural labour in urban settings may attract (or retain) healthy, often male, workers from rural areas who pursue improved incomes or work regarded as more 'secure' or desirable.[20-25] However the reality they meet in more urbanised areas may be challenging. High housing prices, competition for low-skilled jobs, and issues with social integration may increase economic and social pressures on migrants, particularly the poor, who are highly depend on regular cash incomes to meet their basic needs, such as food, shelter, water, sanitation and health care.[4, 17] Rural migrant workers are most often employed in the informal sector in low-skill, low-productivity and low-pay work.[7, 19] Informal workers are often described as more likely than the formally employed to have physically demanding, high workload employment with little opportunities for skills development and greater exposure to toxic chemicals, excessive noise, violence and sexual assault.[4, 7] Employment in the informal sector is associated with poverty and living in sub-standard housing with insufficient access to basic infrastructure and services, e.g. sanitation, waste management and health care.[7] Substandard living conditions and slum dwelling is associated with high dual-burdens of infectious and chronic diseases, including CVD.[4] These added vulnerabilities may identify people who transition out of agriculture and perhaps migrate for work as a high-risk group that may need particular attention in strategies to curb expected increases in CVD burdens with continued urbanisation in India and other LMICs. The evidence base on associations of types of employment, and particularly employment transitions, with chronic diseases is, however, fragmented. Employment-related studies in the health sciences typically assess the association of exposures or interventions with outcomes within a single type of employment. For example, several systematic reviews asses the evidence from LMICs on the effectiveness of improving or intensifying agriculture to improve nutritional status and chronic disease indicators in agricultural communities.[26-31] The employment sector on the other hand typically focuses on shorter-term outcomes relating to health and safety regulations, e.g. ergonomics, injury risk, and exposure to noise and biological or toxic agents, that are often grounded in experiences of NGOs, employers, social movements and communities.[7] This disconnect between the employment and health sectors and sciences makes the respective data hard to integrate in

epidemiological studies to support evidence-based policy. Potential solutions will be discussed in in later sections.

Despite contrasts between India's and China's urban trajectory and labour-force trends[xii], there may be important lessons to learn from China's recent urbanisation and associated management of (or initial failure to manage) transitions out of agriculture that were conducive to health.[32-34] During initial phases of China's recent urbanisation, peasants were poorly integrated into urban areas and their inability to enforce their rights to employment and equal pay typically resulted in employment in low grade, physically demanding, high workload and poorly paid jobs. Failure to enforce the peasant migrants' rights to housing and public services in urban China and their consistent social exclusion were also major issues that typically forced them into substandard living conditions.[33] There are several reports of the environmental and social implications of the early stages of China's recent urbanisation, however studies are yet to assess the related health effects. It would be interesting to explore in further evaluations of China's urbanisation process the extent to which labour-force transitions contribute to intra-urban health inequalities, for example, as a result of expansions of informal settlements.

6.1.2 Addressing objective 2

Initial inspection of the three survey waves of the APCAPS (described in Chapter 2) suggested that I could perform longitudinal analysis of employment shifts out of agriculture and into specific types of non-agricultural occupations. Unfortunately, on closer analysis it became clear that there were not sufficient data overlap among adults across the survey waves for this type of analysis (See Chapter 2, Table 2.2). I considered it unlikely that changes in employment-related CVD risk factors would be detectable between the second (2009-2010) and third (2010-2012) survey waves. As a result, I amended the focus of my PhD research to investigate whether and how place of residence, defined by levels of urbanisation, is associated with CVD risk factors.

The second research paper (see Chapter 4) aimed to explore the associations of place of residence with mean levels of SBP, BMI, LDL and FPG among adults at various stages of urbanisation in Telangana, South India. This objective was achieved through a cross-sectional analysis of data from 6236 adults participating in the third survey wave of the APCAPS set in 27 urbanising villages in Telangana. Remote sensing data on NTLI was used as a proxy for the urbanisation level continuum across participating villages, ranging from a rural settlement to a medium sized town. The study found that even moderate increases in urbanisation levels across the continuum was associated with greater mean SBP and BMI, but not LDL and FPG, after

[xii] In contrast to India, the size of China's urban populations were strictly controlled until the late 70s and Chinese farmers could not freely take up residence and work in urban areas.[32]

adjusting for age, gender and other confounders. The similar trends of these findings and those of the existing urbanisation literature suggested that NTLI is a potentially important new proxy indicator of urbanisation levels particularly in LMIC where objective data on urbanisation factors are not readily available.

In the backdrop of SDG 11 on sustainable cities, considerable attention and investments have been made into making cities more environmentally friendly, green and conducive to health, e.g. through urban parks and infrastructure developments to mitigate pollution and promote physical activity. However, issues of sustainable development and health inequality extend beyond cities. Even more so than cities, towns in LMICs experience increasing pressure on already marginal planning capacities and stretched budgets to deliver basic infrastructure and services to support health of their growing populations.[32] My research indicated that even moderate increases in urbanisation levels across a continuum from a rural settlement to a small town are associated with increasing SBP and BMI among both men and women. Prioritising CVD prevention efforts, such as the population-wide strategies outlined in the WHO Global Action Plan for the Prevention and Control of NCDs,[35] at all stages of urbanisation in India, might help curb expected continued increases in CVD risk factors during continued urbanisation. A proactive approach could potentially help avert steep health costs and losses of productivity from risk factors manifesting into CVDs at early ages in South India.

A major question that emerged during the conduct of this PhD was, through which pathways are urbanisation levels acting on SBP and BMI? A number of studies endeavour to identify variables that mediate the association of urbanisation level with CVD risk factors. A common approach has been to contrast results from one model adjusting for potential confounders to results from one or more models additionally adjusting for potential mediators. Typically these studies, while interesting, fail to represent the full causal web. The strong assumptions that underlie causal inference from this type of analysis may, however, not be met and therefore result in biased estimates.[36] Bias could, for example, arise from bidirectional links between some variables in the analysis (e.g. physical activity and diet) or the presence of unmeasured confounding of exposure-mediator, mediator-mediator or mediator-outcome relationships.[36] Important advances have been made in the field of causal inference during the past decades, lead by the work of Robins and Greenland (1992)[37] and Pearl (2001).[38] Most contributions have focused on methodologies that accommodate a single mediator or sets of mediators considered *en bloc*.[39, 40] More recent work extends these approaches to allow for decomposition of the effects via multiple mediators and other pathways simultaneously (known as direct and indirect effects), including in situations with interaction and non-linearity.[40]

6.1.3 Addressing objective 3

The third research paper of this thesis (see Chapter 5) used path analysis methods first described by Vanderweele and Vansteelandt[40] to address my final PhD objective: To decompose the total effect of increasing level of urbanisation on mean SBP into direct and indirect effects via hypothesised pathways ([i] socio-demographics, [ii] lifestyle and mental health, [iii] metabolic factors) among adults at early stages of urbanisation in Telangana, South India. The findings identified a strong pathway via metabolic factors (i.e. the joint effect of metabolic risk factors hypothesised to mediate the association of urbanisation level with SBP), independent of socio-demographic, lifestyle and mental health factors. The metabolic pathway accounted for a rise in SBP by 13mmHg among men and 2mmHg among women across the urbanisation continuum. Among men, but not women, the pathway via lifestyle and mental health factors, independent of socio-demographic factors, contributed a smaller but considerable rise in SBP by 2mmHg across the urbanisation continuum. The pathway via socio-demographic factors did not affect mean SBP among men or women. The observed urbanisation-related rises in mean SBP potentially already have significant clinical implications in terms of CVD morbidity and mortality in the study population, particularly among men. Two large international systematic reviews and meta-analyses reported large (17-29%) reductions in major CVD events from SBP decreases by 5-10 mmHg,[41, 42] even when blood pressure was in the normal range (i.e. decreased from 130-134mmHg to 120-124mmHg[41]).

Separating the hypothesised pathways into testable components was not straightforward. Ideally, we would want to estimate the individual contribution of each mediator (including the effect via subsequent mediators further downstream the pathway). However, this level of detail could not be achieved by the mediation analysis for reasons I will briefly repeat in the following. The decomposition was largely restricted by methodological limitations to handling complex interrelation of variables in situations with multiple mediators. Uncertainties of the structural dependence (flow of effect) between some variables (e.g. SES and occupation or physical activity, diet and mental health) further limited a more detailed decomposition. Furthermore, the interpretation of the path-specific findings was also not straightforward and largely limited to joint effects of sets of closely related variables summarised in the composite mediators. In order to add more detail to the proposed DAG (Chapter 5 Figure 3), we would need to understand better the structural dependence of all implicated mediators from longitudinal analyses.

Although this study does not seek to generalise it's results to the whole of India, exploring the representativeness of APCAPS to the general Indian population can provide some information on whether the associations reported in this thesis might be more broadly applicable to the

Indian population.[43, 44] **Error! Reference source not found.** provides a comparison of APCAPS participants with a recent representative sample of the Indian poupation[45] across a number of core socio-demographic factors, including place of residence (i.e. rural vs urban or more urbanised areas). There are differences between the two samples in the population distribution by age groups across place of residence. In the Indian sample, people appear equally distributed across rural and urban areas in all age bands up to the age of 60+ years after which rural living became more prevalent. In the APCAPS sample, younger people appear more likely to reside in more urbanised areas and older people appear more likely to reside in rural areas, and there was a greater proportion of the APCAPS population aged 50+ years, than of the Indian sample, living in rural areas. This difference in the distribution of place of residence among different age groups suggests that the findings of the effect of urbanisation on SBP derived from the APCAPS sample may not be fully generalisable to the Indian population as a whole.

More broadly, the APCAPS sample included participants living in settlements ranging from rural communities up to medium-sized towns, while the Indian population sample also includes participants living in major Indian cities. It would therefore certainly also be inappropriate to extrapolate the results reported in this thesis beyond individuals living in medium-sized towns in India. Ideally, the representativeness and generalisability of the results reported here would benefit from standardisation against an Indian or a standard population (for factors with considerably different distribution across place of residence). However, at this initial stage of exploring causal effects of urbanisation levels on SBP, standardisation was not feasible given the complexity of the applied statistical method and the novelty of the urbanisation measure.

Table 6.1 Crude distribution of core socio-demographic factors in India and the Andhra Pradesh Children and Parents Study (APCAPS)

	India			APCAPS		
	Total	**Rural**	**Urban**	**Total**	**Lowest urbanisation level[i]**	**Highest urbanisation level[i]**
Sex ratio (women per 100 men)	99	100	99	90	90	91
Age groups (%)[ii]						
20-29	17.4	16.7	18.8	37.6	38.0	38.9
30-39	13.8	13.1	15.5	8.4	7.8	9.5
40-49	11.4	10.7	12.6	20.5	19.7	20.5
50-59	9.0	8.6	9.6	13.7	15.7	12.0
60+	10.4	10.9	9.5	4.7	5.3	4.1
Caste (%)						
General caste	na	na	na	6.6	6.2	5.5
Scheduled caste	20.6	22.6	16.8	38.7	36.8	35.6
Scheduled tribe	9.2	12.0	3.9	0.7	1.2	0.7
Other backward class	42.2	42.2	42.2	52.4	54.4	57.7
Other	27.2	22.5	36.0	1.6	1.4	0.6
Don't know	0.8	0.6	1.0	na	na	na
Religion (%)						
Hindu	81.4	83.7	77.2	93.4	97.0	92.0
Muslim	12.5	10.6	16.2	4.5	2.0	6.3
Christian	2.7	2.5	3.1	2.0	0.9	1.7
Sikh	1.6	1.7	1.5	a	a	a
Neo-Buddhist	1.0	0.8	1.3	a	a	a
Jain	0.2	0.1	0.5	a	a	a
Other	0.5	0.7	0.3	0.1	0.1	0.1

	India			APCAPS		
	Total	**Rural**	**Urban**	**Total**	**Lowest urbanisation level[i]**	**Highest urbanisation level[i]**
Education level (%)						
No education	15.1	na	na	44.4	44.1	43.3
Up to primary education	na	na	na	15.4	15.6	15.7
Up to secondary education and beyond	na	na	na	40.1	40.3	41.0
Married (aged 15-49 years) (%)[iii]	52.4	49.7	55.1	56.3	55.1	57.8
Currently employed (%)	24.0	25.5	21.3	84.3	87.0	82.1
Agricultural workers (18+ years) (%)	48.3	63.6	11.7	17.3	16.34	15.32
Own agricultural land (%)	38.9	52.6	13.4	54.33	60.72	49.63
Mean number of household members (mean)	4.6	4.7	4.3	4.6	4.6	4.8

APCAPS - Andhra Pradesh Children and Parents Study, DLHS - District Level Household and Facility Survey, NFHS - National Family Health Survey, NTLI - night-time light intensity,

Estimates for India were derived from the NFHS-4 (2015-2016)[45] if not otherwise stated

a Data were included in the 'other' category

[i] Urbanisation levels were defined as thirds of ranked village-level NTLI

[ii] Age band-specific percentages for the adult Indian sample were calculated from information reported by the NFHS-4 (2015-2016)[45]

[iii] Estimates for India were derived from the DLHS-3 (2007-8)[46]

As urbanisation processes and their impacts on cardiovascular health may vary across settings and countries it will be important to perform multi-setting/country analyses or replicate analysis of longitudinal data across settings, countries and regions. Study designs and analyses should take appropriate steps to minimise the risk of measurement and apprehension bias discussed in the second and third research paper of this thesis.

In the following sections, how my findings might inform policy and future research aimed at tackling CVDs in India.

6.2 Implications for research and policy

6.2.1 Existing and new Indicators of urbanisation levels

No single definitions of urbanisation or urbanisation level exist (see Chapter 1, section 1.5), nor is there consensus on how to best measure urbanisation levels. Most studies that explore links between urbanisation levels, CVDs and CVD risk factors compare rural and urban populations and migrant populations. Given the difficulty of these studies, such study designs are parsimonious[xiii] but are not able to disentangle the many inherent differences between these population groups beyond levels of urbanisation (exposure). As a result, residual confounding from a number of unmeasured factors may bias results. In addition, it is challenging to explore the importance of mediators that are hypothesised to be on the causal pathway between residing in an area of low versus high urbanisation level (i.e. rural versus urban), such as education and affluence. The decomposition of effect via different mediators is challenging because the levels of exposure and mediators typically follow similar trends (or correlate) in these types of studies. For example, residents in an area of high urbanisation level (urban) will typically also have higher education levels, higher incomes and better access to goods and services than residents in an area of low urbanisation level (rural) have. At the same time, the levels of these (and other) factors tend to be more similar within population groups at either extreme of the urbanisation continuum, thus mirroring the trend of the binary rural-urban exposure.[47, 48] This limited discordance in levels of covariates, including mediators, at either extreme of the urbanisation continuum typically renders rural-urban comparisons insufficient to adequately control for confounding and explore potential pathways.[16, 48]

Remote sensing data have been widely used by geographers to identify and describe built-up area, green space and urban landscapes.[49, 50] However, the use of remote sensing data in urbanisation-related health research has only recently become more common. One reason might be that calibration methods to correct for systematic differences in satellite orbits and satellite

[xiii] The simplest model invoking least assumptions

instrument lifetimes were previously laborious and associated with risk of bias from user interpretation.[51] In the current PhD, I identified NTLI (derived and calibrated by Dr Robin Wilson) as a potentially important novel continuous proxy of urbanisation levels. Computational efficiency of the data calibration was achieved through application of the 'ridgeline sampling regression' method, thus overcoming previous challenges of laborious calibration. This method further produces consistent global NTLI estimates, reflecting multiple dimensions of urbanisation, such as population density, energy consumption, built up area and settlement size, which have been suggested to be valid proxies for urbanisation levels over time (dating back to 1992), while minimising bias from user interpretation.[49, 51, 52] Using the NTLI data as a proxy for urbanisation level had some limitations as discussed in the second and third research papers of this thesis. A major drawback of using the NTLI data was the technical skill needed to calibrate the data. Another drawback of using the continuous NTLI as the exposure was that it limited my ability to compare my findings directly to existing evidence that used other types of exposure measures, such as rural urban comparisons. I hope that generating evidence using this standardised globally available continuous urbanisation proxy will enable consistent comparisons of research that uses the same urbanisation metric in the future, and contribute to advancing our knowledge on urbanisation and health links.

But it is worth advocating for the use of NTLI as a standalone proxy for a process as complex as urbanisation? And particularly in contexts of recent studies reporting a convergence of some CVD risk factors (BMI) between rural and urban areas?[53] Based on my experience throughout this PhD, I do believe that NTLI is an important tool for learning about urbanisation and health links as there is currently no apparently better, globally available and cost-effective alternative. I will discuss potential future alternatives in section 6.2.2. At the same time I also support the call from other researchers from diverse fields for the establishment of an interdisciplinary 'urbanisation science' that explores "what makes up the most fundamental aspects of urbanization, across space, place, time, and cultures" (Solecki 2013[54], p.14:20). I believe that we will need to consider the whole 'system' of urbanisation (i.e. all the fundamental aspects of urbanisation) and its performance in order to more accurately estimate and predict potential health impacts of urbanisation and identify opportunities for disease prevention. We will further need to explore possible effects from interactions with other changes in other systems including the environment. This approach could perhaps answer a recently posed question of whether the convergence of BMI between rural and urban areas in urbanising settings means that urbanisation does not promote weight gain. And, whether the convergence is perhaps due to interactions of urbanisation with other processes such as climate changes and food systems that feedback on food production and diets across the urbanisation continuum.[34, 54] To this end, I advocate for the establishment of interdisciplinary initiatives that bring together multiple

stakeholders, including politicians, economists, epidemiologists, geographers, public health practitioners, civil society and communities, to identify opportunities for health promoting urbanisation. This approach could further help bridge the disconnect between research on health and employment during urbanisation. To achieve this, it will be crucial to identify a standardised measure of 'types of employment' that is informative for health research as well as other users of employment data. A definition could, for example, build on The Employment Conditions Knowledge Network's theoretical model for linking employment and health inequalities, and guidelines for monitoring national and regional progress towards SDG 8 on decent work and economic growth.[7, 55]

6.2.2 Bridging the data gap

6.2.2.1 An interdisciplinary international data information system

Steps have recently been taken towards establishing a multidisciplinary data information system. In 2008, the European Space Agency (ESA) initiated the Earth Observation for Sustainable Development (EO4SD) initiative in collaboration with international financing institutions (e.g. the World Bank) and their client countries.[56, 57] The EO4SD aims to increase the uptake of Earth Observation (EO)-based information (i.e. drone and satellite derived geo-spatial and remote sensing data) in programs to facilitate achievement of the SDGs[xiv].[58] Initial phases of the EO4SD (2016 to 2018) focused on deriving and mapping key EO geo-information products (e.g. 3D images of buildings and extents and types of land use, infrastructure and air pollution) as well as demonstrating the utility of EO data for supporting countries in their progress towards SDGs relating to three high-priority themes: (i) urban development (ii) agriculture and rural development, (iii) water resources management.[56] So far, The EO4SD has been rolled out in a limited number of cities and rural areas, predominantly in LMICs, and impact evaluations are pending. The EO4SD initiative brings together a wide range of experts across disciplines to produce harmonised, standardised and globally comparable statistical data on a wide range of factors relating to urban and rural development (e.g. land use, crop types, irrigation cover, soil quality, climate, built-up area; type, size and height of buildings and settlements; infrastructure, etc.). When scaled up, I believe this initiative, combined with improved national and regional demographic and disease surveillance, will be an important future resource for studying the development challenges explored in this thesis, as well as

[xiv] As well as urban sustainability indicators e.g. developed by the United Nations (2007), the UN-Habitat (2004), the World Bank (2008) the Asia Development Bank (2001) and the European Commission (2003)[58]

informing a gold standard for defining urbanisation and urbanisation levels across the entire continuum.

6.2.2.2 Collection of nationally representative data in India

National-level data on CVD prevalence, incidence and risk factors are lacking from India and large-scale analyses and projections from studies such as the Global Burden of Disease Study are based on multiple data sources, ranging from national household surveys to smaller-scale epidemiological studies. To address this data gap, the Indian government recently committed to strengthening the health surveillance system and establishing a new disease register by 2020 in their National Health Policy 2017.[59, 60] National and state-level periodical surveys in India (the National Family Health Survey,[45] the District Level Household and facility Survey (DLHS)[61] and the Annual Health Survey (AHS)[62]) collect data on selected risk factors for CVD, such as tobacco use and alcohol, BMI, blood pressure. In more recent surveys, blood-derived measured of HIV and anaemia status have been supplemented by measures of fasting or random plasma glucose. It might be feasible to expand the scope of large periodical surveys further to collect information on a wider range of core CVD indicators and risk factors.[63] For example, any surplus blood could be used for deriving serum cholesterol from dried blood spots on filter paper, which has been recommended for use in Indian studies.[63] It is crucial that data on CVDs and risk factors are collected using standardised definitions, e.g. informed by the WHO STEPwise approach to surveillance (STEPS) of risk factors for NCDs.[63-65]

A number of studies have compared the performance of various CVD risk estimators to predict the risk of CVD in India.[66] The Framingham Risk scores for CVD has emerged as a potentially useful tool for identifying high-risk groups (and individuals) in India, although recalibration of the estimator to local data is needed.[66, 67] The INHEART risk score, which demands a smaller amount of (non-invasive) data, may further be useful for estimating CVD risk in India. A comparison of the estimator's performance in low-, middle, and high-income countries, however, suggests that further calibration is needed to correct for higher risk of CVD at lower levels of risk factors in low-income countries.[68] Using predicted risk of CVDs in place of data on CVD events, will add some uncertainty to statistical models that explore associations of urbanisation on CVDs, and in turn challenge clear interpretations and identification of opportunities for action. It would be interesting to perform further context-specific calibration and validation of the CVD risk scores, to explore their utility to predict CVD events more adequately in settings with limited CVD events data.

The third paper of the current thesis identified a need to understand better the indirect effects of urbanisation level on SBP via upstream determinants (e.g. ambient air pollution and mass media advertising). These data are not readily available from many LMIC settings and their collection

will need greater priority in future surveillance, surveys and epidemiological studies. Major limitations to the collection and analysis of data on diet and physical activity were also highlighted that repeated previous calls for novel metrics to capture diet and physical activity data more efficiently and accurately. There is also a need to identify metrics and collect data on knowledge, attitudes, social norms and preferences that interact with aspects of urbanisation to shape CVD related behaviours, for example, those relating to diet, smoking and alcohol.[69] Depression was the only mental health indicator included in the current PhD research and due to unknown or bidirectional flows of effect between mental health and lifestyle factors (for example bidirectional links between mental health and physical activity), I could not estimate a separate pathway via mental health. It is possible that ongoing research into the links between mental health and lifestyle factors will enable the identification of a separate mental health pathway in future. Going forward, it will be important to collect and include information on other types of psychosocial factors linked to CVD including stress from work, home, finances or major life events.[70] Further studies of urbanisation and CVD risk links would further benefit from including information on epigenetics, early life determinants and comorbidities that may affect CVD risk, such as chronic kidney disease.[71] The current study could for example be replicated with an increased scope, using the freely available data from the Institute for Health Metrics and Evaluation (used in the GBD studies). Future GBDs studies could also include information on NTLI-derived urbanisation levels.

6.2.3 Aligning national strategies for economic growth, sustainable development and the prevention and control of CVDs in India

There is a well-recognised need to align evidence-based strategies and legislation on economic growth, sustainable development and prevention and control of NCDs to meet global goals[xv] for sustainable development and reduction of CVDs.[35, 55]

To stimulate economic growth, the National Institution for Transforming India's three-year action agenda (2017-20) advocates continued urbanisation while addressing current challenges of affordable housing, transport and basic infrastructure services and land use. A second key focus area is concerned with boosting skills development and employment generation in formal sectors outside of agriculture, mainly in the industry and service sectors, to facilitate shifts out of agriculture that parallel urbanisation in India (and elsewhere).[72] At the same time, India's National health policy (NHP) 2017[59] sets out the ambitious objective to reduce premature mortality from cardiovascular diseases, cancer, diabetes and chronic respiratory diseases by 25% by 2025. To attain this objective, the NHP and The National Program for Prevention and

[xv] For example the UN sustainable development goals (SDG)[55] and the WHO global action plan for the prevention and control of noncommunicable diseases[35]

Control of Cancer, Diabetes, CVD and Stroke (NPCDCS)[60] prioritise strengthening the health care system, with predominant focus on improving secondary and tertiary care[xvi]. The NHP and NPCDCS further advocate for 'coordinated action' to address major CVD risk factors, including tobacco, alcohol, substance abuse, diet, physical inactivity, stress, road safety, safe water and sanitation, solid waste management, indoor and outdoor air pollution and health seeking behaviour. The NPCDCS advocates for targeting 'unhealthy lifestyles' through education and health promotion, whereas the NHP does not outline specific strategies for action. The NHP further identifies a need to meet health care needs of vulnerable urban populations such as temporary migrants. Ensuring equitable and quality health care for all and strengthening the health system's capacity for early detection and treatment of CVDs may considerably reduce the rate and case fatality of CVD across the urbanisation continuum.[73] However, in order to avoid considerable health care expenditure for the government and individuals, it would be recommended to scale up prevention measures and set measurable goals for their achievement.[68, 74]

The third research paper of this thesis (Chapter 5) discussed the potentially far reaching effects of small reductions in blood pressure even within the normal range. In brief, two large trials suggested that a reduction of 5 to 10mmHg could reduce major CVD events, CVD mortality and all-cause mortality by between 17% and 29%, including in non-hypertensive people.[41, 42] The results of the third research paper suggested that the effect of urbanisation on SBP via the metabolic pathway, independently of socio-demographic, lifestyle and mental health factors contributed an increase in SBP among both men and women of up to 13mmHg and an additional sizable indirect effect via lifestyle and mental health among men. These large effects suggest that India could potentially reduce major CVD events, CVD mortality and all-cause mortality considerably by acting on the indirect effects observed in my study across socio-demographic strata country-wide. However, as discussed in section 6.1.3 variations in the distribution of population characteristics and differences in the magnitude and process of urbanisation across India's 29 diverse states need careful consideration when planning interventions. Triangulating my results with published evidence on hypothesised social and environmental upstream determinants of the observed effects (discussed in Chapter 5), made a strong case for implementing available population-wide prevention strategies discussed later in this section. Even when considering some dilution of effects, e.g. from measurement error and apprehension bias, the magnitude of the effects in my study far exceeded the 5mmHg that significantly reduced CVD events in previously discussed trials.[41, 42] In the light of expected continued urbanisation and associated rises in CVD risk factors in India suggested by the

[xvi] Attainment of quality and affordable health coverage for all and ensuring early detection and control of NCDs.

research presented in this thesis, it is recommended to implement available, evidence-based, population-wide interventions that target urbanisation-related social and environmental factors across sociodemographic profiles and levels of urbanisation, i.e. including urbanising settlements and medium-sized towns. A number of relevant cost-effective strategies are available from the WHO's best buys for NCD prevention and control. These, for example, include (i) marketing restrictions on alcohol, tobacco and unhealthy foods (ii) taxation of alcohol and food content (such as salt) and (iii) restriction of physical availability of alcohol, tobacco.[71, 75] India has already committed to the WHO Framework Convention on Tobacco Control, however improved enforcement and monitoring of their Cigarettes and Other Tobacco Products Act is needed for implementation to take full effect.[71] The WHO and the Climate and Clean Air Coalition to Reduce Short-Lived Climate Pollutants additionally recommend a number of priority areas and strategies to mitigate negative health effect from ambient air pollution, for example though shifting to low-emission public transport options.[76] The cost-effectiveness of these interventions, however, need further exploration. Evidence from LMICs suggest that population-based prevention strategies, including fiscal measures (e.g. affecting food prices) and multi-intervention strategies (e.g. combining health promotion campaigns, fiscal measures and regulation of food advertising) are more cost-effective for tackling major CVD risk factors (such as obesity) than individual-focused interventions,[71, 73] which currently dominate national priorities in India. Studies from India suggest that fiscal measures targeting ambient and food environments through taxation of petrol, palm oil, sugar sweetened beverages, tobacco and alcohol, will substantially benefit cardiovascular health.[65, 72] A major advantage of fiscal policies is that they typically pay for themselves by reducing healthcare costs.[73]

A new innovative 'Smart City Mission' in India[77] takes the critical first steps to addressing a broad set of environmental factors associated with city living and hypothesised to mediate the association between urbanisation and blood pressure. The Smart City Mission is a five year renewal and retrofitting programme, through which the Indian government has incentivised 100 Indian cities to provide equitable decent quality of life for their citizens by becoming cleaner, greener, and more environmentally sustainable while providing core infrastructure.[77] Cities are for example encouraged to establish pedestrian and cycling infrastructure as well as preserving and developing parks and recreational spaces with the aim of reducing congestion and air pollution and promoting physical activity and mental health. Evaluation of this innovative programme has unique opportunities to inform national intersectoral policies[xvii] addressing CVD risk factors associated with urbanisation and urban living. Expanding best practices in line with existing evidence-based (simple and affordable) intervention options[35, 74] to non-participating

[xvii] working with sectors lying outside the health sector, e.g. addressing poverty, lack of education, sustainable and resilient agriculture and unhealthy environment

cities and ensuring their implementation during future urban development, could help curb urbanisation-related rises in CVD risk factors anticipated in my PhD research. My findings suggest that major CVD risk factors, SBP and BMI, increase with urbanisation levels, even at early stages of urbanisation and before settlements reach 'city status'. Limited resources and planning capacities of towns in India could interact with rising levels of risk factors to accelerate rises in CVD events if action is not taken in these settings. I therefore support the UNPF's call for a pre-emptive approach with greater focus on towns and small cities in urban planning efforts and strategies to prevent CVDs in India.[32] Given the increases in SBP and BMI across the urbanisation continuum from a rural settlement to a medium sized town, I also urge that 'Smart' initiatives and sustainable planning be extended to transitioning areas at early stages of urbanisation in India.

6.3 Conclusion

The findings of this PhD research show that increasing level of urbanisation across a continuum ranging from a rural settlement to a medium sized town is associated with rises is mean SBP and BMI but not LDL or FPG that might have far reaching implications for CVD morbidity and mortality in South India. The rise in SBP across the urbanisation level continuum was mainly due to differences in metabolic factors, and among men, partially due to lifestyle and mental health factors. These findings offer new insight into possible mechanisms underlying the links between urbanisation and CVD risk factors and stresses the need for considering the prevention of increased blood pressure during continued urbanisation in India.

India is already taking important steps towards sustainable development in relation to urbanisation and labour-force transitions. My findings make a case for the need simultaneously to strengthen efforts aimed at preventing negative changes in CVD risk factors that occur with urbanisation. In many developing nations, including India, urbanisation has outpaced the policies and service provision needed to prevent rises in intra-urban health inequalities.[3, 32] Agricultural workers who migrate for work may experience particular patterns of vulnerability that may need particular attention in strategies to curb expected rises in CVDs with continued urbanisation in India. There are, however, many significant challenges to measuring urbanisation levels, types of employment, lifestyle and mental health indicators that need to be address in a new type of interdisciplinary approach, across demography, environment, urbanisation, migration, sustainable development, and health for us to better understand how urbanisation acts on CVDs in LMICs.

References

1. India State-Level Disease Burden Initiative CVD Collaborators. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990-2016. The Lancet Global health. 2018 Dec;6(12):e1339-e51
2. India State-level Disease Burden Initiative Collaborators. Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet. 2017 Nov 13
3. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010 Sep 27;365(1554):2809-20
4. Tacoli C. Urbanization and emerging population issues working paper 7. Urbanization, gender and urban poverty: paid work and unpaid carework in the city. United Nations Population Fund, International Institute for Environment and Development,, 2012 March. Report No.
5. Gupta R, Mohan I, Narula J. Trends in Coronary Heart Disease Epidemiology in India. Annals of global health. 2016 Mar-Apr;82(2):307-15
6. Ibrahim MM, Damasceno A. Hypertension in developing countries. Lancet. 2012 Aug 11;380(9841):611-9
7. Employment Conditions Knowledge Network. Employment Conditions and Health Inequalities: Final Report to the WHO. 2007.
8. Norboo T, Stobdan T, Tsering N, Angchuk N, Tsering P, Ahmed I, et al. Prevalence of hypertension at high altitude: Cross-sectional survey in Ladakh, Northern India 2007-2011. BMJ Open. 2015;5 (4) (no pagination)(e007026)
9. Subasinghe AK, Walker KZ, Evans RG, Srikanth V, Arabshahi S, Kartik K, et al. Association between farming and chronic energy deficiency in rural South India. PLoS ONE [Electronic Resource]. 2014;9(1):e87423
10. Arlappa N, Venkaiah K, Rao KM, Reddy Ch G, Kumar SS, Ravindranath M, et al. Prevalence of chronic energy deficiency in rural-dwelling older Indian adults during a period of severe drought. Journal of Nutrition for the Elderly. 2009 Jul;28(3):301-12
11. Zhou B, Li Y, Stamler J, Tao S, Davis CE, Wu Y, et al. Relation of occupational change to cardiovascular risk factor levels in rural Chinese men: the People's Republic of China-United States Collaborative Study on Cardiovascular and Cardiopulmonary Epidemiology. Am J Public Health. 2003 Dec;93(12):2049-51
12. Ezzati M, Webster CJ, Doyle YG, Rashid S, Owusu G, Leung GM. Cities for global health. 2018;363:k3794
13. World Health Organization. Urbanization and health 2019 [cited 2019 16 July]. Available from: https://www.who.int/bulletin/volumes/88/4/10-010410/en/.
14. Vlahov D, Freudenberg N, Proietti F, Ompad D, Galea S. A conceptual framework for organizing determinants of urban health. Thematic paper 1 for Knowledge Network on Urban Settings WHO Centre for Health Development2008.
15. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition reviews. 2012 Jan;70(1):3-21
16. Allender S, Foster C, Hutchinson L, Arambepola C. Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review. Journal of urban health : bulletin of the New York Academy of Medicine. 2008 Nov;85(6):938-51
17. UNDP/UNFPA, UNICEF, WFP. Agenda Item 1: Population Growth and Rapid Urbanization: Food Insecurity on the Rise in Urban Settings. UNDP/UNFPA, UNICEF, WFP, 2009.
18. Popkin BM, Horton S, Kim S, Mahal A, Shuigao J. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutrition reviews. 2001 Dec;59(12):379-90
19. Dodd W, Humphries S, Patel K, Majowicz S, Little M, Dewey C. Determinants of internal migrant health and the healthy migrant effect in South India: a mixed methods study. BMC Int Health Hum Rights. 2017 Sep 12;17(1):23
20. Allender S, Wickramasinghe K, Goldacre M, Matthews D, Katulanda P. Quantifying urbanization as a risk factor for noncommunicable disease. Journal of urban health : bulletin of the New York Academy of Medicine. 2011 Oct;88(5):906-18
21. Tacoli C. Urbainzation, gender and poverty. New York, USA: United Nations Population Fund, International Institute for Environment and Development, 2012 March. Report No.

22. Goryakin Y, Suhrcke M. Economic development, urbanization, technological change and overweight: what do we learn from 244 Demographic and Health Surveys? Econ Hum Biol. 2014 Jul;14:109-27
23. Bhandari PB. Rural livelihood change? Household capital, community resources and livelihood transition. Journal of rural studies. 2013 Oct 1;32:126-36
24. World Bank. Employment in agriculture (% of total employment) (modelled ILO estimate): World Bank; 2018 [updated 2018; cited 2018 28 September]. Available from: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?contextual=employment-by-sector&locations=XO.
25. Cervantes-Godoy D, Brooks J. Smallholder Adjustment in Middle-Income Countries: Issues and Policy Responses. OECD publishing, 2008.
26. Masset E, Haddad L, Cornelius A, Isaza-Castro J. Effectiveness of agricultural interventions that aim to improve nutritional status of children: systematic review. Bmj. 2012;344:d8222
27. Berti PR, Krasevec J, FitzGerald S. A review of the effectiveness of agriculture interventions in improving nutrition outcomes. Public Health Nutr. 2004 Aug;7(5):599-609
28. Girard AW, Self JL, McAuliffe C, Olude O. The effects of household food production strategies on the health and nutrition outcomes of women and young children: a systematic review. Paediatr Perinat Epidemiol. 2012 Jul;26 Suppl 1:205-22
29. Turner R, Hawkes C, Jeff W, Ferguson E, Haseen F, Homans H, et al. Agriculture for improved nutrition: the current research landscape. Food Nutr Bull. 2013 Dec;34(4):369-77
30. Kadiyala S, Harris J, Headey D, Yosef S, Gillespie S. Agriculture and nutrition in India: mapping evidence to pathways. Annals of the New York Academy of Sciences. 2014 Dec;1331:43-56
31. Hawkes C, Ruel M. The links between agriculture and health: an intersectoral opportunity to improve the health and livelihoods of the poor. Bulletin of the World Health Organization. 2006 Dec;84(12):984-90
32. United Nations Population Fund. state of world population 2007: Unleashing the Potential of Urban Growth. New York: UNPF, 2007.
33. Chen M, Gong Y, Lu D, Ye C. Build a people-oriented urbanization: China's new-type urbanization dream and Anhui model. Land Use Policy. 2019 2019/01/01/;80:1-9
34. Chen M, Liu W, Lu D. Challenges and the way forward in China's new-type urbanization. Land Use Policy. 2016 2016/09/01/;55:334-9
35. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Switzerland: WHO, 2013.
36. VanderWeele T, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface. 2009;2(4):457–68
37. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology (Cambridge, Mass). 1992 Mar;3(2):143-55
38. Pearl J, editor Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence; 2001; San Francisco: Morgan Kaufmann.
39. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015 Mar;71(1):1-14
40. VanderWeele T, Vansteelandt S. Mediation Analysis with Multiple Mediators. Epidemiologic methods. 2014 Jan;2(1):95-115
41. Bundy JD, Li C, Stuchlik P, Bu X, Kelly TN, Mills KT, et al. Systolic Blood Pressure Reduction and Risk of Cardiovascular Disease and Mortality: A Systematic Review and Network Meta-analysis. JAMA Cardiology. 2017;2(7):775-81
42. Ninomiya T, Perkovic V, Turnbull F, Neal B, Barzi F, Cass A, et al. Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials. BMJ (Clinical research ed). 2013 Oct 3;347:f5680
43. Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. International journal of epidemiology. 2013;42(4):1018-22
44. Kukull WA, Ganguli M. Generalizability: the trees, the forest, and the low-hanging fruit. Neurology. 2012;78(23):1886-91
45. International Institute for Population Sciences, ICF. National Family Health Survey (NFHS-4) 2015-16. Mumbai ,India International Institute for Population Sciences, 2017.
46. International Institute for Population Sciences. District Level Household and Facility Survey (DLHS-3), 2007-08. Mumbai India: Minestry of Health and Family Welfare Government of India, 2010.
47. Knowledge Network on Urban Settings. Our cities, our health, our future. Kobe City, Japan: World Health Organization, 2008.

48. Dahly DL, Adair LS. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. Social science & medicine (1982). 2007 12/29;64(7):1407-19
49. Yi K, Tani H, Li Q, Zhang J, Guo M, Bao Y, et al. Mapping and evaluating the urbanization process in Northeast China using DMSP/OLS nighttime light data. Sensors (Basel, Switzerland). 2014;14(2):3207-26
50. Ma T, Zhou C, Pei T, Haynie S, Fan J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. Remote Sensing of Environment. 2012 2012/09/01/;124(Supplement C):99-107
51. Zhang Q, Pandey B, Seto KC. A Robust Method to Generate a Consistent Time Series From DMSP/OLS Nighttime Light Data. IEEE Transactions on geoscience and remote sensisng. 2016
52. Zhang Q, Seto KC. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. Remote Sensing of Environment. 2011 2011/09/15/;115(9):2320-9
53. Bixby H, Bentham J, Zhou B, Di Cesare M, Paciorek CJ, Bennett JE, et al. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature. 2019 2019/05/01;569(7755):260-4
54. Solecki W, Seto KC, Marcotullio PJ. It's Time for an Urbanization Science. Environment: Science and Policy for Sustainable Development. 2013 2013/01/01;55(1):12-7
55. United Nations. Sustainable Development Goals: United Nations; [updated na; cited 2019 18 September]. Available from: https://sustainabledevelopment.un.org/?menu=1300.
56. Earth observation for sustainable development (EO4SD). urban development. Earth observation for sustainable development (EO4SD). na
57. European Space Agency. About the EO4SD initiative 2019 [cited 2019 24 September]. Available from: http://eo4sd.esa.int/about/.
58. European Space Agency. Earth Observation for sustainable development. Urban Development - Service Portolio. na.
59. Government of India. National health policy. 2017.
60. Government of India. National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke (NPCDCS). 2012-13.
61. International institute for population sciences. District level household and facility survey - 4. State fact sheet Telangana 2012-13 Mumbai, India: Governmet of India, 2012-13.
62. Government of India. Annual Health Survey 2019 [cited 2019 19 August]. Available from: http://censusindia.gov.in/2011-Common/AHSurvey.html.
63. Raban MZ, Dandona R, Dandona L. Availability of data for monitoring noncommunicable disease risk factors in India. Bulletin of the World Health Organization. 2012 Jan 1;90(1):20-9
64. World Health Organization. STEPwise approach to surveillance (STEPS): World Health Organization; 2019 [cited 2019 27 August]. Available from: https://www.who.int/ncds/surveillance/steps/en/.
65. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India: Current Epidemiology and Future Directions. Circulation. 2016 Apr 19;133(16):1605-20
66. Garg N, Muduli SK, Kapoor A, Tewari S, Kumar S, Khanna R, et al. Comparison of different cardiovascular risk score calculators for cardiovascular risk prediction and guideline recommended statin uses. Indian heart journal. 2017 2017/07/01/;69(4):458-63
67. Chow CK, Joshi R, Celermajer DS, Patel A, Neal BC. Recalibration of a Framingham risk equation for a rural population in India. Journal of Epidemiology and Community Health. 2009;63(5):379
68. Yusuf S, Rangarajan S, Teo K, Islam S, Li W, Liu L, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. The New England journal of medicine. 2014 Aug 28;371(9):818-27
69. GBD 2017 SDG Collaborators. Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018 Nov 10;392(10159):2091-138
70. Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. Lancet. 2004 Sep 11-17;364(9438):953-62
71. Prabhakaran D, Singh K, Roth GA, Banerjee A, Pagidipati NJ, Huffman MD. Cardiovascular Diseases in India Compared With the United States. Journal of the American College of Cardiology. 2018 Jul 3;72(1):79-95
72. The National Institution for Transforming India. The Three Year Action Agenda 2017-18 to 2019-20. Delhi, India: Government of India, 2017.

73. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. The Lancet. 2010 2010/11/20/;376(9754):1775-84
74. World Health Organization. Preventing chronic diseases a vital investment. Geneva, Switzerland: World Health Organization, 2005.
75. World Health Organization. The updated Appendix 3 of the WHO Global NCD Action Plan 2013-2020. Geneva, Switzerland: World Health Organization, 2017.
76. World Health Organization (WHO), the Climate and Clean Air Coalition to Reduce Short-Lived Climate Pollutants (CCAC). Reducing global health risks through mitigation of short-lived climate pollutants. Scoping report for policymakers. Geneva, Switzerland: World Health Organzation, 2015.
77. Ministry of Urban Development Government of India. Smart Cities. Mission Statement & Guidelines. 2015.