Essays on Poverty and Health in Indonesia

A thesis submitted to The University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Humanities

2016

Wulung Anggara Hanandita

School of Social Sciences

Contents

Word count: 43,301

List of Tables

List of Figures

List of Abbreviations

ADB	Asian Development Bank
AF	Alkire-Foster method
AIC	Akaike Information Criterion
API	Annual Parasite Incidence
BMI	Body Mass Index
BPS	*Badan Pusat Statistik* (Statistics Indonesia, Indonesian Statistical Bureau)
CAR	Conditional Autoregressive
CDC	Centers for Disease Control and Prevention
CES-D	Center for Epidemiologic Studies Depression scale
CHOPIT	Compound Hierarchical Ordered Probit model
CI	Confidence Interval, Credible Interval
cm	centimetre
DHS	Demographic and Health Survey
DIC	Deviance Information Criterion
DIF	Differential Item Functioning
DJF	December-January-February climatological season
DOI	Digital Object Identifier
DWD	*Deutscher Wetterdienst* (German Meteorological Service)
FGT	Foster-Greer-Thorbecke measure of unidimensional poverty

GADM	Global Administrative Areas database
GDP	Gross Domestic Product
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GLS	Generalized Least Squares
GMM	Generalized Method of Moments
GPCC	Global Precipitation and Climatology Centre
HPI	Human Poverty Index
IDR	Indonesian Rupiah
IFLS	Indonesia Family Life Survey
IHME	Institute for Health Metrics and Evaluation
IIA	Independence of Irrelevant Alternatives
INLA	Integrated Nested Laplace Approximation
IQR	Interquartile Range
ITN	Insecticide-treated Net
IV	Instrumental Variable
JJA	June-July-August climatological season
JB	Jawa (Java) and Bali island-group
KA	Kalimantan (Borneo) island-group
Kemenkes	*Kementerian Kesehatan* (Ministry of Health)
kg	kilogram
KMO	Kaiser-Meyer-Olkin measure for identity correlation matrix
LATE	Local Average Treatment Effect
LNRI	*Lembaran Negara Republik Indonesia* (State Gazette, Statute Book)
LPM	Linear Probability Model
MAM	March-April-May climatological season

MAP	Malaria Atlas Project
MCMC	Markov Chain Monte Carlo
MDGs	Millennium Development Goals
ML	Maximum Likelihood
mm	milimetre
MPI	Multidimensional Poverty Index
MRF	Markov random field
NT	Nusa Tenggara (Lesser Sunda islands) island-group
OLS	Ordinary Least Squares
OPHI	Oxford Poverty & Human Development Initiative
OPROBIT	Ordered Probit model
OR	Odds Ratio
PCE	Per capita household Consumption Expenditure
PhD	Doctor of Philosophy
PM	Papua and Maluku (the Moluccas) island-group
Podes	*Potensi Desa* (Village Census)
PP	Parasite Prevalence
PPP	Purchasing Power Parity
RAND	Research and Development Corporation
RE	Random Effects
Riskesdas	*Riset Kesehatan Dasar* (National Basic Health Research)
RRR	Relative Risk (odds) Ratio
SD	Standard Deviation
SES	Socio-economic Status
SL	Sulawesi (Celebes) island-group
SM	Sumatra island-group

SON   September-October-November climatological season

SRH   Self-rated Health, Self-report Health

SRQ-20   20-item Self-reporting Questionnaire

Susenas   *Survei Sosial Ekonomi Nasional* (National Socio-economic Survey)

UNDP   United Nations Development Programme

UNSCN   United Nations Standing Committee on Nutrition

UNSD   United Nations Statistics Division

US   United States of America

USD   United States Dollar

WHO   World Health Organization

Essays on Poverty and Health in Indonesia

The University of Manchester, Doctor of Philosophy

Wulung Anggara Hanandita

14th November 2016

Abstract: This thesis presents five standalone essays that demonstrate the feasibility and utility of employing advanced analytic techniques to cross-sectional data from Indonesia in order to deal with some technical challenges typically encountered either in the estimation of social gradient in health or in the monitoring and evaluation of well-being as a multidimensional construct. The first essay estimates the causal effect of poverty on mental health by exploiting a natural experiment induced by weather variability across 440 districts in the Indonesian archipelago. The second essay applies parametric anchoring vignette methodology to investigate the extent to which the estimates of demographic and socio-economic inequalities in self-rated health are biased by survey respondents' differential reporting behaviour. The third essay formally assesses the existence and identifies the social determinants of the double burden of malnutrition in Indonesia using a variant of a generalised linear mixed model. The fourth essay maps the social and spatial distributions of malaria in 27 districts in Indonesian Papua using a probabilistic disease mapping technique that is capable of accounting for the complex dependency structure of spatially-correlated multilevel data. The fifth essay examines the extent and patterns of multidimensional poverty in Indonesia over the last decade using a novel poverty measurement method that is sensitive to both the incidence and intensity of multiple deprivations in income, health and education domains. Together, these essays show that although health and social researchers in the developing world have little choice but to conduct cross-sectional studies, new insights can sometimes be gained if one is willing to look at existing data through a new lens. In all five cases presented here, this approach is proved to be useful in shaping practical policy-making.

Keywords: poverty, health inequality, mental health, self-rated health, nutrition, malaria, Indonesia

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

All empirical chapters of this thesis have been published as original research articles in peer-reviewed journals:

- Chapter 2:
  Hanandita, W., Tampubolon, G., 2014. Does poverty reduce mental health? An instrumental variable analysis. *Social Science & Medicine* 113, 59–67, doi: `10.1016/j.socscimed.2014.05.005`.

- Chapter 3:
  Hanandita, W., Tampubolon, G., 2016. Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. *Quality of Life Research* 25 (5), 1137–1149,
  doi: `10.1007/s11136-015-1152-y`.

- Chapter 4:
  Hanandita, W., Tampubolon, G., 2015. The double burden of malnutrition in Indonesia: Social determinants and geographical variations. *SSM - Population Health* 1, 16–25, doi: `10.1016/j.ssmph.2015.10.002`.

- Chapter 5:
  Hanandita, W., Tampubolon, G., 2016. Geography and social distribution of malaria in Indonesian Papua: A cross-sectional study. *International Journal of Health Geographics* 15, 13, doi: `10.1186/s12942-016-0043-y`.

- Chapter 6:
  Hanandita, W., Tampubolon, G., 2016. Multidimensional poverty in Indonesia: Trend over the last decade (2003–2013). *Social Indicators Research* 128 (2), 559–587, doi: `10.1007/s11205-015-1044-0`.

These articles are jointly authored with the main supervisor of this thesis, Gindo Tampubolon. For each paper, the student, Wulung Anggara Hanandita, conceived the study, analysed and interpreted the data, and wrote and edited the manuscript. The supervisor contributed to the interpretation of the results and editing of the manuscript. Both student and supervisor declare that there is no conflict of interest.

Acknowledgements

During my study in Manchester, I have been fortunate enough to be surrounded by wonderful people who treat me and my family like their own family. Hanging around with Zaenal Muttaqin, Miranda Zahra, Munasprianto Ramli, Media Yohana Putri, Lidwin Saputra, Kurniati Mustika Ningrum, and Firmandika and Widyastuti Harda keeps my sanity in check; beyond that, I am grateful for their support in the difficult times that followed the birth of my daughter Karina.

My biggest debt is to my family in Indonesia whose prayers, encouragement and financial support have made it possible for me to come this far. My mother Rini, my father Sudiharjo, and my brother Koko are the source of my motivation. I am also grateful to my mother-in-law Fatriani, my father-in-law Isro', and brother-in-law Luthfi for their blessings.

Finally, and most importantly, I owe everything that I achieved during this PhD to my beloved wife Raisa. Her unconditional love is what has kept me standing throughout this journey.

<div align="right">

WULUNG ANGGARA HANANDITA
Manchester, UK
November 14, 2016

</div>

The Author Wulung Anggara Hanandita holds an MSc in Social Research Methods and Statistics (2013) from the University of Manchester. His main research interest is the application of realistically complex statistical models to large-scale health and social survey data so as to maximise the depth of policy-relevant insights that can be derived from them. He is broadly interested in the issues of poverty, inequality and population health. Prior to embarking on social statistics training in Manchester, Wulung graduated from Waseda University (Japan) with a BA in Liberal Arts (2010) and an MA in International Relations (2012).

The Examiners (1) Dr Arief Gusnanto, Department of Statistics, University of Leeds.
(2) Dr Johan Koskinen, Social Statistics Discipline Area, University of Manchester.

To my mother Rini,
my wife Raisa,
and my daughter Karina.

Only those who will risk going too far
can possibly find out how far one can go.
T.S. ELIOT

Statisticians do not in general exactly agree on how to analyze anything
but the simplest of problems. The fact that statistical inference uses
mathematics does not imply that there is only one reasonable or useful
way to conduct an analysis. Engineering uses math as well, but there are
many ways to build a bridge.
RICHARD MCELREATH

Chapter 1

Introduction

1.1 Motivation for the PhD

Observations of socio-economic inequalities in health are among the most pervasive and enduring in history. Since the mid-17th century, epidemiologists have documented that individuals of lower social class who are generally subjected to poor housing conditions, as well as harsh labour and physical environments, tend to have worse health status and die earlier than their richer counterparts (Graunt, 1662; Villermé, 1830; Chadwick, 1842; Virchow, 1848). Today, more than 350 years since the first pioneering epidemiological observation was published (Morabia, 2013), and despite modern advances in medicine, this phenomenon persists. Research continually shows that socio-economic disparities in health are observed across a wide spectrum of health dimensions spanning from physical to mental health, and from infectious to chronic diseases (Marmot, 2005). Evidence suggests that this variation exists across and within countries, regardless of a nation's income level (Marmot, 2005, 2013, 2015). In fact, the Whitehall study (Marmot et al., 1984, 1991) reveals that instead of obeying a threshold-model, the relationship between wealth and health is best characterised by a gradient model, implying the existence of a health continuum that still holds true even for those who are well above the absolute poverty line (Adler and Ostrove, 1999).

Two major pathways are hypothesised to link socio-economic status, or poverty for that matter, to health: the material pathway and the psychosocial pathway (Benzeval et al., 2014). In resource-poor settings, where material deprivation is severe, a social gradient in health could predominantly arise from degrees of absolute deprivation

Figure 1.1: Hypothesised pathways linking socio-economic status to health
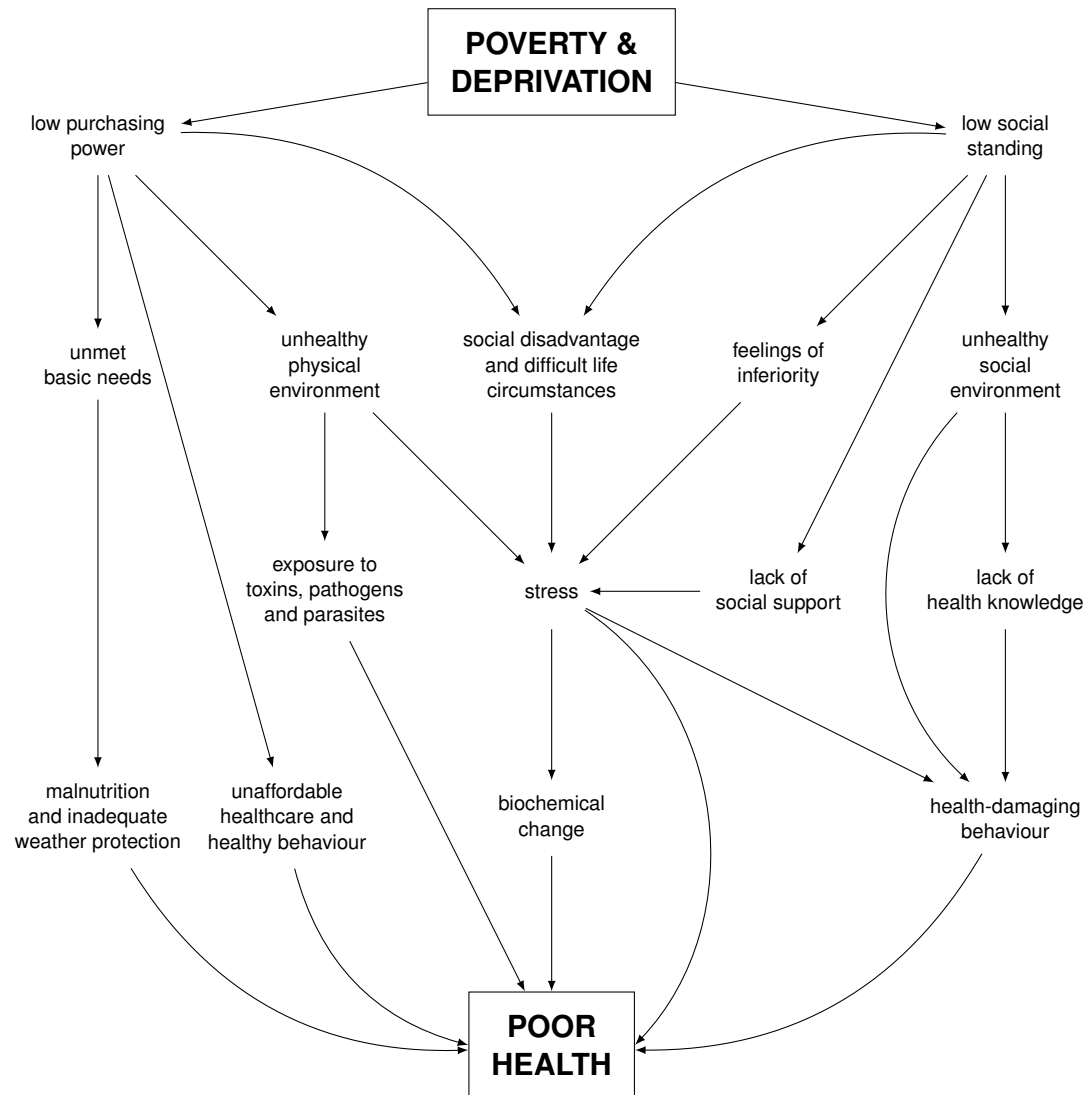
and the resulting behavioural adaptations to difficult life circumstances (the left arm of Figure 1.1; Marmot, 2005). In high-income countries with little or no absolute poverty, however, psychosocial harms originating from relative deprivation in social functionings may be more dominant (the right arm of the same figure; Marmot, 2005). Either way, a growing body of literature shows that stress emanating from both forms of deprivation is capable of triggering physiological alterations that may eventually affect health status via allostatic load, weathering, biological imprinting, and hormonal or inflammatory processes (Kubzansky et al., 2014). The life course literature further suggests that social gradient in health could be 'a result of inequalities in the accumulation of social, psychological and biological advantages and disadvantages over time' (Skalická et al., 2009: 1273), highlighting the importance of both the intensity and timing of social exposure in influencing health (Glymour et al., 2014).

Although such inequalities undoubtedly threaten the economy through reduced worker productivity and inflated healthcare costs (LaVeist et al., 2011), they are not, in general, inevitable (Whitehead, 1992). There is potential to reduce them since it is known that (1) within-society health disparities are largely attributed to preventable differences in access to basic amenities, education and healthcare services (Arcaya et al., 2015), and that (2) notable variation does exist across societies with differing welfare arrangements (Huijts et al., 2010). Curbing health inequalities through the implementation of public policies (Osypuk et al., 2014; Yamey, 2012), therefore, has both moral (Faden and Powers, 2008; Whitehead, 1992) and rational (LaVeist et al., 2011) groundings. In particular, as these inequalities reflect not only the differential in risk exposures but also the unequal or even unfair distribution of resources that are essential for maintaining good health (Link and Phelan, 1995; Phelan et al., 2010; Phelan and Link, 2013), a serious consideration is usually given to resource prioritisation and redistribution policies such as in the case of 'Obamacare' in the United States (Cockerham, 2013; Leonhardt, 2010).

Given their potential overarching impacts on a society, it is indisputable that any policy intervention strategies aiming to reduce health inequalities need to be informed by robust empirical evidence (WHO, 2013). However, to date, few studies have rigorously examined the extent of health inequalities in the context of low- and middle-income countries. Much of what we know about socio-economic inequalities in health has thus far been derived from studies originating from the United States, the United Kingdom or other parts of the Western world (see for an overview Berkman et al., 2014). As Araya et al. (2003) cautioned, academic discourse is all too often based on the implicit assumption that whatever is true in the industrialised world also holds true in resource-poor settings (see also Vathesatogkit et al., 2014 on mortality; Lund et al., 2010 on mental health; Groce et al., 2011, Hosseinpoor et al., 2015 and Simkiss et al., 2011 on disability; Subramanian et al., 2010 on self-rated health; Prince et al., 2008 on dementia; and Worrall et al., 2005 on malaria).

While part of this trend is attributable to 'stakeholders' focus on achieving the health-related Millennium Development Goals (MDGs), which are measured at the aggregate level [instead of at the within-country level]' (Alonge and Peters, 2015: 2), it is undeniable that data deficiency remains a major, if not the main, impediment to research in the developing world (Nuyens, 2007; Howe et al., 2012; Alkire and Santos, 2014). In contrast to the evidence base in high-income countries, empirical evidence in low- and middle-income countries is largely based on simple descriptive and bivariate analyses performed to small facility or community samples that are prone to con-

Figure 1.2: Map of Indonesia

scale approx 1:26,000,000

0   2000   4000 6000 km

Wallace Line

Equator Line

**Island Group
in Indonesia**
 A Sumatra
 B Jawa-Bali
 C Kalimantan
 D Sulawesi
 E Nusa Tenggara
 F Maluku
 G Papua
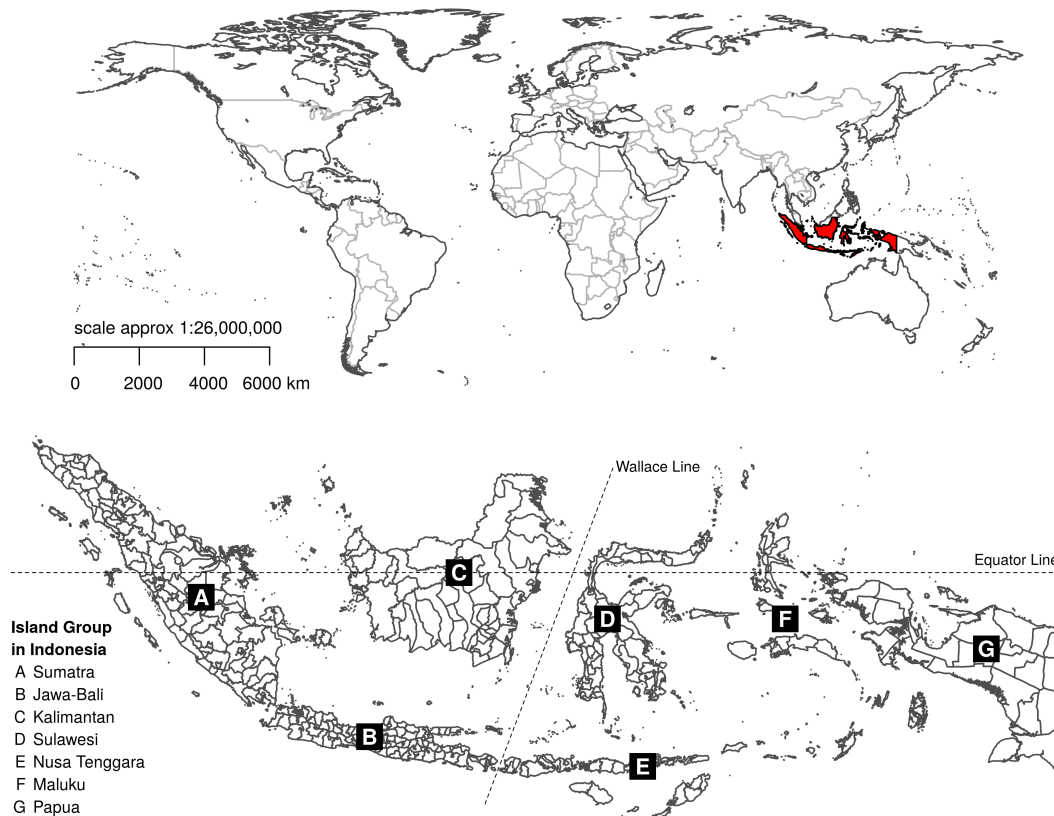
founding and selection biases. The scarcity of experimental and longitudinal data in these settings further makes unbiased estimates of socio-economic gradient in health difficult to obtain; inconsistent findings are not only rampant but also hardly reconcilable as socio-economic status is often loosely defined, poorly measured, and incomparably operationalised in existing studies (Cooper et al., 2012; Howe et al., 2008, 2012; Worrall et al., 2005). On top of that, even though high-quality population data like the Demographic and Health Survey (DHS) have recently become available, their utilisation remains sub-optimal because geographical information and other valuable material contained in special survey modules that have been collected at enormous cost are often ignored in standard, off-the-shelf statistical analyses (Kandala and Ghilagaber, 2014).

It is out of this context that the work of this PhD has arisen. Analysing large-scale, nationally representative survey data from Indonesia (Figure 1.2) using advanced analytic techniques, this PhD aims to bring recent methodological advances in statistics and econometrics to health inequality research in developing countries. More specifically, it seeks to demonstrate how the utilisation of realistically complex statistical models could help stakeholders maximise the depth of policy-relevant insights that

can be derived from existing cross-sectional survey data. In so doing, it is hoped that policy-makers and society as a whole will get more 'value for money' from these data that were collected at such great expense.

To further delineate the scope of investigation, this PhD deals particularly with three routine tasks that are indispensable for the formulation of sound evidence-based policies in developing countries, namely (1) the unbiased estimation of socio-economic gradient in health, (2) the joint estimation of social and spatial distributions of health, and (3) the measurement of multidimensional well-being. As will be demonstrated in later chapters, a judicious application of advanced analytic techniques to these tasks could significantly enhance policy-makers' ability to justify intervention measures, prioritise limited resources, and monitor the impact of public policies on the overall well-being of the population. To this end, five empirical essays in the domains of physical health, mental health and self-rated health will be presented. The detailed outline of these essays is laid out in Section 1.3; but before that, the next section will give three reasons as to why Indonesian data are used in the first place.

1.2   The case for Indonesia

Among 115 developing countries that exist in the world today, Indonesia is unique for the purpose of health inequality research in several significant ways and therefore is used as a case in point in this doctoral study.

Firstly and most importantly, Indonesia possesses high-quality population data that permit the study of socio-economic inequalities in health using a broadly-agreed, well-defined measure of socio-economic status. In contrast to other data sources in the developing world (Howe et al., 2008), the majority of household surveys from Indonesia are equipped with a detailed household consumption expenditure module, which is considered to be the gold standard for measuring material living standards in low- and middle-income countries (Deaton and Zaidi, 2002; Howe et al., 2012). Furthermore, unlike other data sources, Indonesian samples are large in size (in the range of 1 million individuals for repeated cross-section surveys) and are rarely limited to certain age- or gender-groups; at the same time, they are representative of the general population living at the lowest autonomous administrative units of a highly decentralised political regime. These distinctive features of Indonesian data not only make it possible for investigators to get accurate estimates at a small-area level but also allow for precise estimates to be obtained from variants of data-hungry,

Figure 1.3: Causes of death, poverty and inequality indicators, and the state of health research in Indonesia (sources: World Bank, 2014b; PubMed, 2016)
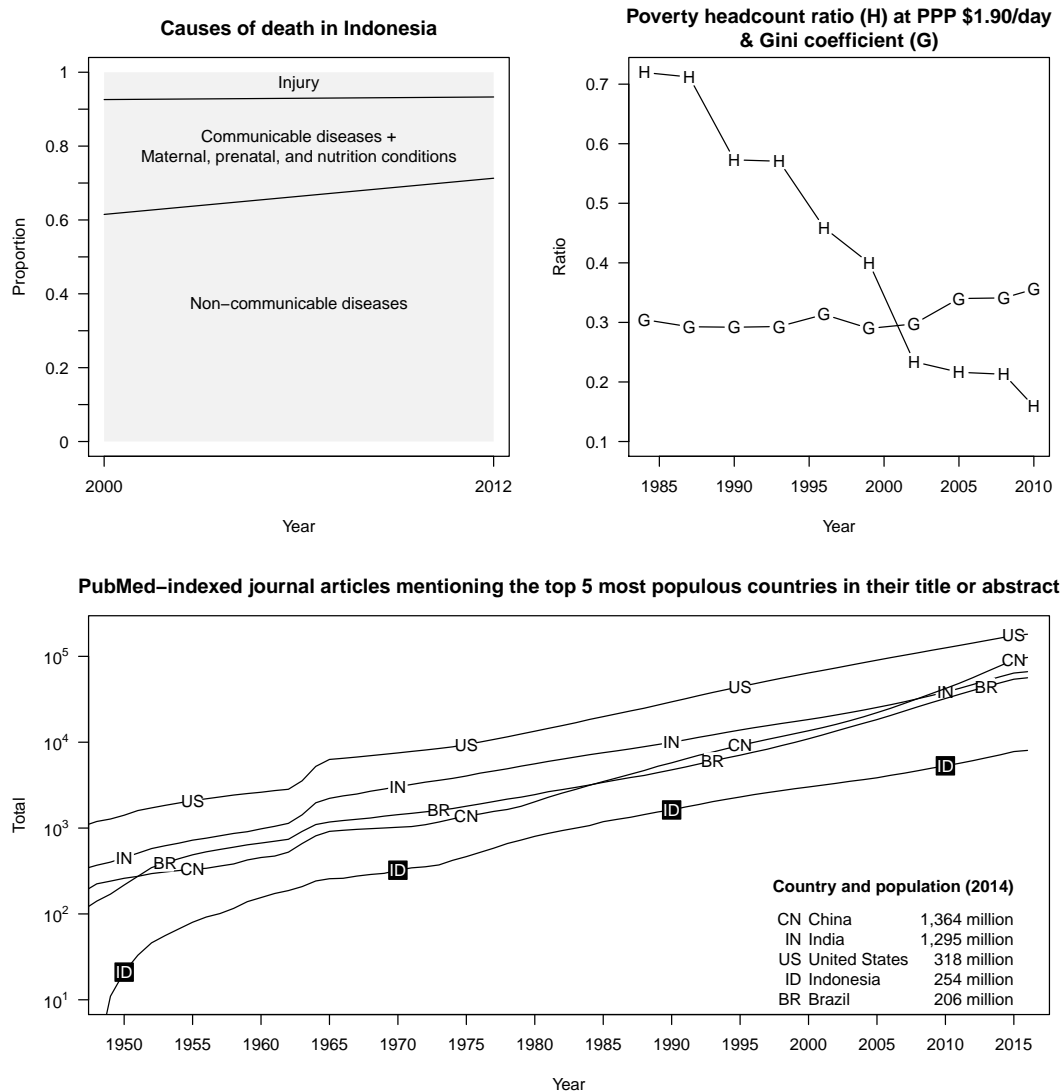
inefficient-but-consistent statistical procedures. These data, therefore, easily lend themselves to the kind of advanced analyses to be demonstrated in this thesis.

Secondly and substantively, Indonesia provides one of the most fertile grounds for the study of health inequalities in resource-poor settings. The 17,000-island equatorial archipelago is among the few places in the world where the double burden of infectious and non-communicable diseases (Boutayeb, 2006; Gillespie and Haddad, 2003; the top-left panel of Figure 1.3) and, until very recently, a pay-for-service healthcare system (Lancet, 2014), meet not only with acute poverty but also with stark income and spatial inequalities (World Bank, 2014a; the top-right panel of Figure 1.3). This situation, although certainly unpleasant for those who must live with its effects, is ripe for academic endeavours because it offers investigators the opportunity to study the

socio-economic patterning of multiple health outcomes within a single developing society.

Finally and equally importantly, despite being the fourth most populous country in the world with its population of 260 million, Indonesia's presence in the global health conversation is still very weak (Horton, 2016; the bottom panel of Figure 1.3). This, in combination with the speed and scale of the economic and epidemiologic transitions Indonesian society is currently undergoing, suggests that 'Indonesia has much to tell (and teach) the world about its [health and medical] experiences' (Horton, 2016: 830). This thesis's deliberate use of Indonesian data, therefore, will certainly enrich the literature on health inequalities in resource-poor settings.

1.3   Structure of the thesis

The novelty of this thesis lies in its emphasis on the application of advanced techniques in the analysis of high-quality population data from Indonesia. In order to showcase the value added by applying recent methodological advances to health inequalities research in developing countries, this thesis is presented as a volume of five self-contained essays that can be read independently of one another. The first two essays (Chapters 2 and 3) study the effect of bias in the empirical estimation of socio-economic gradient in health using observational cross-sectional data. The following two essays (Chapters 4 and 5) demonstrate how to effectively incorporate geographical information into a joint estimation of social and spatial distributions of health so that stakeholders can use the resulting model to inform policy targeting. Finally, the last essay (Chapter 6) suggests a way to monitor the impact of policy interventions on the overall level of well-being using a novel multidimensional poverty measurement method that is sensitive to both the incidence and intensity of multiple deprivations in income, health and education domains.

The outline of the remaining chapters is as follows. Assuming the presence of endogeneity, Chapter 2, DOES POVERTY REDUCE MENTAL HEALTH? AN INSTRUMENTAL VARIABLE ANALYSIS, seeks to estimate the causal effect of poverty on mental health by exploiting a natural experiment induced by weather variability across 440 districts in the Indonesian archipelago. Linear and non-linear instrumental variable as well as control function estimators are used. In addition, sensitivity analyses with respect to distributional assumptions, sample stratification and model specifications are presented. This chapter contributes to the advancement of the mental health literature by

shifting the focus of the social causation hypothesis research from the study of association using small community or facility samples to the study of causal effect using large observational data. The chapter has been published in Hanandita, W., Tampubolon, G., 2014. Does poverty reduce mental health? An instrumental variable analysis. *Social Science & Medicine* 113, 59–67, doi: `10.1016/j.socscimed.2014.05.005`.

Chapter 3, Does Reporting Behaviour Bias the Measurement of Social Inequalities in Self-rated Health in Indonesia? An Anchoring Vignette Analysis, studies the extent to which differential health reporting behaviour biases the estimates of health inequalities by demographic and socio-economic status among older Indonesians. Interpersonal heterogeneity in reporting style is identified by asking respondents to rate a number of vignettes that describe varying levels of health status in six health domains using the same ordinal response scale that is also applied to the self-report health questionnaires. Estimates obtained from a compound hierarchical ordered probit model, which are adjusted to response-scale heterogeneity, are compared to those obtained from an ordinary ordered probit model. Insights gained from this study could have significant implications on future research measuring health inequalities in low- and middle income countries using self-rated health questionnaires. The chapter has been published in Hanandita, W., Tampubolon, G., 2016. Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. *Quality of Life Research* 25 (5), 1137–1149, doi: `10.1007/s11136-015-1152-y`.

Chapter 4, The Double Burden of Malnutrition in Indonesia: Social Determinants and Geographical Variations, investigates the coexistence and the determinants of under- and overnutrition problems using a multilevel multinomial logistic regression technique. Unlike many existing studies, this chapter promotes the formal assessment of the coexistence of under- and overweight within a population using a model-based inferential approach rather than relying on conventional prevalence estimates. Robustness of results is established by means of conducting sex- and urban/rural-stratified analyses as well as by fitting a quantile regression model. In addition to its formal determination of the double burden, this chapter also contributes to the literature through its investigation into whether or not the improvement of living standards constitutes a sufficient policy measure for promoting healthy nutritional status in countries experiencing rapid economic and epidemiologic transitions. The chapter has been published in Hanandita, W., Tampubolon, G., 2015. The double burden of malnutrition in Indonesia: Social determinants and geographical variations. *SSM - Population Health* 1, 16–25, doi: `10.1016/j.ssmph.2015.10.002`.

Chapter 5, GEOGRAPHY AND SOCIAL DISTRIBUTION OF MALARIA IN INDONESIAN PAPUA: A CROSS-SECTIONAL STUDY, maps the social and spatial distributions of malaria in 27 districts on the island of Papua. It also tests the association between poverty and malaria after accounting for differences in socio-demographic and geographical risk factors, while at the same time taking into account the complex multilevel spatial structure of the data. A Bayesian hierarchical logistic regression model with spatial random effects is fitted; sensitivity analyses with respect to the specification of hyperpriors are provided. This study is among the few that perform model-based disease mapping techniques to malaria-endemic areas on the Indonesian side of the island of New Guinea. This chapter has been published in Hanandita, W., Tampubolon, G., 2016. Geography and social distribution of malaria in Indonesian Papua: A cross-sectional study. *International Journal of Health Geographics* 15, 13, doi: `10.1186/s12942-016-0043-y`.

Chapter 6, MULTIDIMENSIONAL POVERTY IN INDONESIA: TREND OVER THE LAST DECADE (2003-2013), examines the extent and patterns of multidimensional poverty in Indonesia from 2003 to 2013 through the application of the Alkire-Foster method, a novel poverty measurement technique that is sensitive to both the incidence and intensity of multiple deprivations. An Indonesian version of the Multidimensional Poverty Index is constructed by augmenting the existing consumption poverty measure with information on health and education. A characterisation of the temporal trend, an evaluation of regional disparity, and an investigation into the extent of inequality among the poor are presented. This chapter provides one of the most comprehensive multidimensional poverty evaluations in Indonesia; it has been published in Hanandita, W., Tampubolon, G., 2016. Multidimensional poverty in Indonesia: Trend over the last decade (2003–2013). *Social Indicators Research* 128 (2), 559–587, doi: `10.1007/s11205-015-1044-0`.

Finally, Chapter 7 summarises the findings and concludes the thesis. Supplementary data and programming scripts used in proceeding chapters are provided in the Appendices.

<div align="center">❈ ❈ ❈</div>

Chapter 2

Does poverty reduce mental health? An instrumental variable analysis

Abstract: That poverty and mental health are negatively associated in developing countries is well known among epidemiologists. Whether the relationship is causal or associational, however, remains an open question. This paper aims to estimate the causal effect of poverty on mental health by exploiting a natural experiment induced by weather variability across 440 districts in Indonesia (N = 577,548). Precipitation anomaly in two climatological seasons is used as an instrument for poverty status, which is measured using per capita household consumption expenditure. Results of an instrumental variable estimation suggest that poverty causes poor mental health: halving one's consumption expenditure raises the probability of suffering mental illness by 0.06 point; in terms of elasticity, a 1% decrease in consumption brings about 0.62% more symptoms of common mental disorders. This poverty effect is approximately five times stronger than that obtained prior to instrumenting and is robust to alternative distributional assumptions, model specifications, sample stratification and estimation techniques. An individual's mental health is also negatively correlated with district income inequality, suggesting that income distribution may have a significant influence upon mental health over and above the effect of poverty. The findings imply that mental health can be improved not only by influencing individuals' health knowledge and behaviour but also by implementing a more equitable economic policy.

Keywords: poverty, mental health, Indonesia, weather, precipitation anomaly, instrumental variable, gmm, control function

2.1 Introduction

The negative association between poverty and mental health in developing countries has been increasingly documented. Research from various parts of the world generally shows that low levels of income, education, and assets as well as low social class are correlated with a higher probability of having common mental disorders (Lund et al., 2010). However, empirical evidence regarding the causal effect of the association remains scarce. Few studies have investigated the strength or the direction of causality between poverty and mental health in developing countries, although such study clearly benefits the formulation of public policy aimed at improving the health of the population. In encouraging study of this topic in the United States, Stowasser et al. (2011: 2) note that '...if causal links between wealth and health were confirmed, society would likely benefit from more universal access to health care and redistributive economic policy. Yet, if such causal links were rebutted, resources would be better spent on influencing health knowledge, preferences, and ultimately the behavior of individuals.' Considering both the growing burden of disease attributed to mental illness (IHME, 2013) and tightly constrained health budgets (Patel, 2007), it is important to understand whether poverty reduces mental health in developing countries.

The fact that poverty is negatively associated with mental health in low- and middle-income countries is hardly surprising, but to reach a convincing estimate of its causal effect is certainly not an easy task. Two-way or simultaneous causation may come into play (Smith, 1999), inflating the estimated effect and making it impossible for researchers looking at observational data to separate the effect of wealth on mental health (social causation hypothesis) from that of the reverse (social selection hypothesis). Secondly, the observed wealth-health relationship may be confounded by unobserved common causes that accidentally induce a spurious correlation. Genetic frailty, early childhood environment, family background and preference or taste for lifestyle may impact both an individual's ability to work (and hence accumulate wealth) and his or her susceptibility to mental illness (Stowasser et al., 2011). The study on the mental health effect of poverty may also suffer from what is generally known as the attenuation bias. More often than not, wealth is measured with error, as a noisy, low signal-to-noise ratio variable which could trivially result in a downward-biased parameter estimate (Cameron and Trivedi, 2005). Because these endogeneity problems might be working at the same time, it is difficult to predict the magnitude and direction of the potential bias resulting from their presence *a priori*. In addition, the small number of population data available in developing countries remains a major obstacle for public health research.

The aim of this paper is therefore to address these issues. We apply instrumental variable and control function estimators to a large (N = 987,205), nationally representative dataset from Indonesia, namely the *Riset Kesehatan Dasar* (Riskesdas) 2007. We use seasonal precipitation anomaly, defined as the average deviation of monthly precipitation from its half-century (1951–2000) normals in all 440 *kabupaten* (districts) in Indonesia, as an instrument for poverty status. The identifying assumptions are that precipitation anomaly strongly predicts per capita household expenditure in a largely agricultural economy (relevance condition), is randomly assigned hence unrelated to any potential unobserved confounders (validity condition), and is exerting its influence upon mental health only through its effect on consumption expenditure (exclusion restriction). Conditional on these partially testable assumptions, the instrumental variable approach allows the analyst to isolate the exogenous variation of poverty, thus allowing for the derivation of a consistent estimate of the mental health effect of poverty in the presence of endogeneity. This study is one of the few population-based studies that attempts to look beyond the simple correlation between poverty and mental health in the context of low- and middle-income countries.

2.1.1 Poverty and mental health: association and causality

The two-way causation between poverty and mental health has been recognised for quite some time. The consensus among epidemiologists seems to suggest that the social causation hypothesis (wealth → health) is more plausible for explaining high-prevalence mental disorders such as depression and anxiety disorders, whilst the social selection hypothesis (health → wealth) is probably more relevant for low-prevalence mental disorders like schizophrenia (Goldberg and Morrison, 1963; Muntaner et al., 2004; Saraceno et al., 2005). Despite the intuitive logic behind this consensus (Adler and Ostrove, 1999; Dohrenwend et al., 1992), it is important to note that there have been only sparse empirical attempts to separate the competing causal directions (Muntaner et al., 2004).

Amid the paucity of population data equipped with reliable income and mental health measures, research in low- and middle-income countries have so far been able to investigate only the associational nature between wealth and mental health. Researchers often have to rely on small community or facility samples which are not only prone to the self-selection bias but also limit the application of multivariate statistical tools. The majority of community and facility studies conducted throughout the developing world suggests that poverty is positively associated with mental illness

(Lund et al., 2010). Population-based studies (Dzator, 2013; Hamad et al., 2008; Myer et al., 2008) also support this finding, although they have not yet addressed the endogeneity issues; in the Indonesian setting in particular, Tampubolon and Hanandita (2014) recently confirmed the association using data from the Indonesia Family Life Survey 2007.

In contrast to the associational nature of studies conducted in developing countries, investigations of the causal effect of wealth on mental health began to appear as early as the mid-1990s in developed countries. Acknowledging the dual relationship between health and economic status (Smith, 1999) as well as the potential error in measuring income and the possibility of confounding due to unmeasured variables, Ettner (1996) took a set of variables (work experience, state-wide unemployment rate, parental education, spousal and spouse's parents' education, and spousal work experience) as instruments for individual income in the United States (N = 8,000; see also illustration in Figure 2.1). She applied a two-stage instrumental variable estimator and found that reduced income leads to worse mental health (as indicated by higher Center for Epidemiologic Studies Depression Scale (CES-D) scores). The effect was four times stronger after instrumenting for income, although Meer et al. (2003) and Frijters et al. (2005) later cast doubt on the validity of her instrument set. Using the same identification strategy but with a different instrument set (age, inheritance, time in current job, mother's education, fraction of household income earned by the respondent, hours watching TV and rural-urban residence), Zimmerman and Katon (2005) did not detect any statistically significant effect of financial status (debt-to-asset ratio) on mental health (CES-D score). They admitted that the non-significance might be due to the poorly performing instrument set, although one could argue that the application of an instrumental variable estimator to a small sample (N = 2,000) might well account for the finding.

Three points in the existing literature are worth noting. First, the negative association between poverty and mental health is generally found in both developed and developing countries, but there is a marked difference with respect to the weight of the evidence. Among the studies conducted in low- and middle-income countries, there has been a lack of the investigation into the causal relationship that has been performed in high-income countries. Perhaps the only study that has sought to do so is the one conducted by Chin (2010), which did not find a statistically significant income effect on mental health in Malawi (F = 10.34; N = 2,400). Second, although the small-sample bias as well as the inefficiency properties of instrumental variable estimator have been well studied (Bound et al., 1995; Cameron and Trivedi, 2005),

Figure 2.1: Addressing endogeneity problem with instrumental variable (solid line = causal relationship, dashed line = mutual dependence on irrelevant common causes)
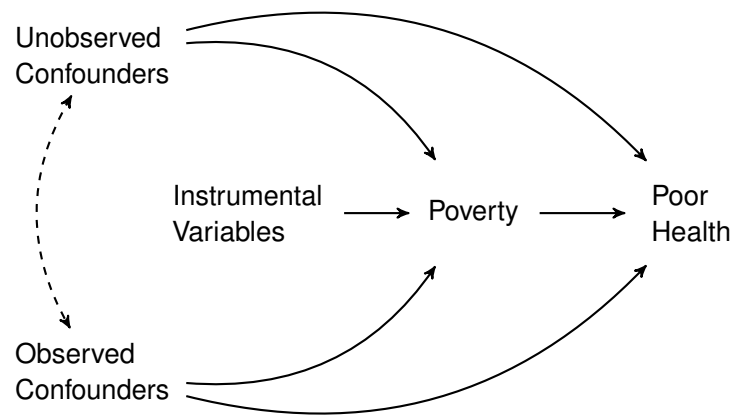
previous applications were mostly limited to small datasets; sometimes, in addition, they were carried out with a rather weak instrument set. Finally, instrumental variable analysis offers a way to address endogeneity problems, but due to the limited availability of data collected in developing countries, it is likely that researchers would not have the privilege of exploring instrument sets like those used in the two US studies reviewed above. This does *not* mean, however, that there is no way for researchers working with data from developing countries to implement the technique.

2.1.2 Weather variability as a source of exogenous variation

One promising instrument for individual income to be used in developing countries is the variability of rainfall over time and across places. It is not difficult to see that, in predominantly agriculture-dependent economies, the amount of precipitation in a given locality should be positively correlated with crop production, hence strongly determining individual income or consumption expenditure. Levine and Yang (2006: 5) showed that 'higher local rainfall leads to higher rice output in Indonesian districts' which 'occurs contemporaneously (in the same calendar year), rather than with a lag', although the effect seems to be statistically significant only 'in districts that are not major cities'. This is also supported by Kishore et al. (2000), who looked at the impact of rainfall anomalies during the 1997–1998 El Niño event in Indonesia. In Africa, analysis conducted using Ugandan data has also resulted in similar findings: higher rainfall is correlated with higher production of coffee, bananas and peas as well as higher GDP (Björkman-Nyqvist, 2013). In fact, Miguel et al. (2004) found that positive rainfall shock is generally associated with positive GDP growth in 41 African

countries. At the individual level, the positive correlation between rainfall shock and individual income has been confirmed in the Philippines (Yang and Choi, 2007), Malawi (Chin, 2010), Tanzania (Savage and Fichera, 2013) and Thailand (Paxson, 1992) as well. Studies consistently show that positive rainfall shock can be generally interpreted as a positive exogenous income shock for individuals living in developing countries.

Working under the assumption that rainfall shock is a random variate uncorrelated with any unobserved common causes and is exerting its influence on the health outcome of interest only through its effect on the instrumented variable, researchers have been able to estimate the causal effect of individual economic status on general health status and subjective well-being in Malawi (Chin, 2010) as well as on body mass index, self-rated health, height-for-age, weight-for-age and vaccination coverage in Tanzania (Savage and Fichera, 2013). In the Philippines, Glewwe and King (2001) used rainfall shock in combination with the price of salt to identify the impact of early childhood nutritional status on cognitive development. These demonstrate the utility of weather variability as a natural experiment.

In the next section we describe the data, measures and statistical methods used to estimate the causal effect of poverty on mental health in Indonesia.

2.2 Methods

2.2.1 Data

The data is drawn from the *Riset Kesehatan Dasar* (Riskesdas) 2007. Managed by the Ministry of Health of the Republic of Indonesia, the Riskesdas study is the largest public health study ever conducted in the country. The 2007 wave includes 987,205 individuals from 258,366 households residing in all 440 districts and is representative of the Indonesian population (Kemenkes, 2008). Its size and coverage clearly distinguish the Riskesdas dataset from the Indonesia Family Life Survey (IFLS) dataset (30,000 individuals living in 260 districts) that was previously analysed by Tampubolon and Hanandita (2014) (see Figure A.2 in Appendix A). Informed consent was obtained prior to interview, and participants' confidentiality was strictly protected. Further details regarding ethical and sampling procedures are available in Kemenkes (2008). For our purposes, individuals younger than 15 years old were excluded from the analysis because of their ineligibility for the mental health questionnaire (Kemenkes, 2008);

also excluded were those who reported a history of schizophrenia. These exclusions yield a complete-case final sample size of 577,548 individuals.

2.2.2 Measure of mental health

Mental health is measured using the 20-item Self-Reporting Questionnaire (SRQ-20), which was specifically developed as an instrument for detecting non-psychotic mental disorders in primary health care settings (Harding et al., 1980). The instrument has favourable psychometric properties and has been validated in many developing countries including Vietnam, Rwanda, Mongolia, China and others (Beusenberg and Orley, 1994; Chen et al., 2009; Ghubash et al., 2001; Giang et al., 2006; Pollock et al., 2006; Scazufca et al., 2009; Scholte et al., 2011; Stratton et al., 2013). In the Riskesdas 2007 study, eligible respondents were asked to report whether or not they experienced the 20 symptoms of non-psychotic mental disturbances (exact wording is provided in Appendix A). Responses were coded using a binary 'yes/no' indicator, and mental health scores were derived by summing the individual items. This yields a mental health score whose theoretical value ranges from 0 (mentally healthy) to 20 (severely depressed). In accordance with a validation study conducted by Ganihartono (1996), a cut-off point of 6 is used. Therefore, individuals are classified as having clinically significant symptoms of common mental disorders (probable caseness) if their mental health scores are equal to or higher than six. Both the raw and dichotomised mental health scores are analysed in the following statistical analysis.

2.2.3 Measure of poverty

Poverty is measured using the log-transformed per capita household consumption expenditure, which reflects 'a household's ability to meet (or exceed) their material needs and to access services' (Howe et al., 2012: 876). This measure is relatively accurate for measuring standards of living in Indonesia because of its ability to capture the monetary welfare of the self-employed or informal workers (Deaton and Zaidi, 2002) who constitute the majority (60–70%) of the Indonesian labour force (Nazara, 2010). Household expenditure is insensitive to intermittent income shock; it is thus capable of delivering a good approximation for permanent income (Deaton and Zaidi, 2002; see also Cutler and Katz, 1992 and Poterba, 1989 on why expenditure is preferred to income). Substantively, as demonstrated by Ecob and Smith (1999), the log transformation allows analysts to capture the log-linear or the proportional

Figure 2.2: Precipitation anomaly in March-April-May (MAM) 2007 season
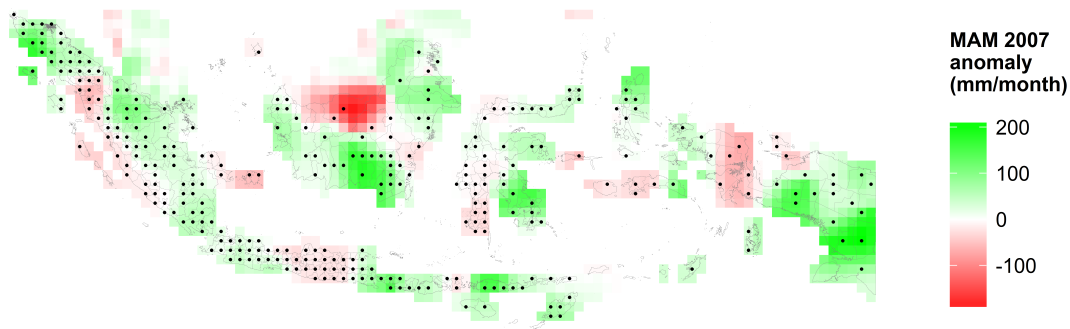
relationship between individual poverty level and (mental) health. The transformation also makes the distribution more symmetric, hence reducing the influence of outliers.

2.2.4 Measure of weather variability

We use precipitation anomaly to instrument for the endogenous poverty variable. Precipitation anomaly data is obtained from the Global Precipitation and Climatology Centre (GPCC) (Schneider et al., 2014), which is operated by the German Meteorological Service (DWD). The specific dataset used in this paper is the GPCC Land-Surface Full Data Reanalysis Version 6 dataset at 0.5° resolution (Meyer-Christoffer et al., 2011; Schneider et al., 2011). The 0.5° latitude by 0.5° longitude spatial grid is approximately equal to an area of 55 × 55 kilometres at the equator, exactly where Indonesia is located. We then matched the centroid of every district in Indonesia to its corresponding grid in order to obtain the measure of precipitation anomaly (millimetre/month) in four climatological seasons (December-January-February (DJF or winter), March-April-May (MAM or spring), June-July-August (JJA or summer) and September-October-November (SON or autumn)) for the year 2007. This is depicted in Figure 2.2.

2.2.5 Control variables

In the models we include standard individual- and household-level socio-demographic controls, measures of health-related behaviours and two district-level contextual variables. Age is treated as a categorical variable with six factors (the reference is 15–24 years old). Gender is a dummy variable representing the female gender. Marital status

is treated as a categorical variable with being married as the reference. Education is also a categorical variable; the reference is less than middle school, which in the Indonesian context is equivalent to not having completed the nine-year compulsory education (*wajib belajar*). Employment status is a dummy variable indicating whether or not an individual is unemployed. Physical activity is treated as a dummy variable denoting those who reported less physical activity (the derivation of this measure is given in Kemenkes, 2008). Frequent smoker (smokes every day), heavy drinker (drinks ≥ 5 days in a week) and having chronic illness (any of the following: cardiovascular disease, diabetes, cancer, stroke or hypertension) are all entered as dummy variables. Household size is a continuous covariate, while household residential location is a dummy variable indicating those who reside in an urban area. The district-level covariates are deprivation index and income inequality (Gini index). The deprivation index measures the lack of basic social facilities in each district (see Appendix A). It was calculated from the *Potensi Desa* (Podes) 2008 dataset, which covers all 75,410 villages across the archipelago. The Gini index, in a 0–1 scale, was derived from the *Survei Sosial Ekonomi Nasional* (Susenas) 2007 dataset using the method described by Milanovic (1997). These two contextual variables are entered as continuous covariates.

2.2.6   Modelling techniques

Mental health is modelled as a function of individual-, household- and district-level determinants. We analyse both the dichotomous (probable caseness) and continuous parameterisations of the SRQ-20 score in order to avoid the loss of information due to misclassification error (Zimmerman and Katon, 2005: 1202). Statistical analysis is carried out in two steps: initially, we treat poverty as a predetermined variable, ignoring its potential endogeneity; in the second step, we use precipitation anomaly to instrument for log per capita household consumption expenditure. Both linear and Poisson regression models are fitted for the continuous outcome, while linear probability and probit models are applied to the dichotomous outcome. The Poisson model offers a convenient way of addressing the skewed and non-negative nature of the SRQ-20 score (Gould, 2011; Nichols, 2010; Santos Silva and Tenreyro, 2006), whereas the linear model offers a number of diagnostic tools that are useful for testing the exogeneity of the suspected endogenous variable as well as for measuring the strength of the instrument set. The Generalized Method of Moments (GMM) estimator is used to estimate the endogenous linear, linear probability and Poisson models. The linear and linear probability models share an identical $E[z_i(y_i - x_i'\beta)] = 0$ moment

condition, while the Poisson model uses the additive error $\mathrm{E}[\boldsymbol{z}_i(y_i - \exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\})] = \mathbf{0}$ moment condition (Windmeijer and Santos Silva, 1997). On the other hand, the endogenous probit model is fitted using the Maximum Likelihood (ML) estimator exploiting the joint normality of the correlated error terms (Cameron and Trivedi, 2010). The precise mathematical expression of these models and a note on the use of linear model are provided in Appendix A.

To obtain a more intuitive interpretation of parameter estimates and to allow for a straightforward comparison with the linear probability model, we report average marginal effect instead of a raw regression coefficient for probit model. In all cases, sampling weight is used in order to obtain nationally representative parameter estimates (Kemenkes, 2008) and standard errors are clustered by district using a generalised Huber/White robust variance estimator (Cameron and Trivedi, 2010; Snijders and Bosker, 2012) to allow for arbitrary heteroscedasticity and autocorrelation within districts. Continuous covariates are centred to their respective grand mean (log per capita household expenditure, Gini index) or to a representative value (household size of 3, deprivation index equals 0) so that the intercept can be given a meaningful interpretation. Finally, we conduct robustness analysis in three ways: (1) we test the stability of the poverty effect by taking out some of the control variables, (2) we refit the models with urban-rural stratification, and finally (3) we re-estimate the models with a different but closely related control function estimator (Imbens and Wooldridge, 2007), as well as with a random effects estimator.

2.3 Results

2.3.1 Descriptive and bivariate analysis

Table 2.1 shows that the distribution of mental health scores is, as expected, extremely right skewed with 11.5% of the study participants categorised as having clinically significant symptoms of common mental disorders. This figure is very close to the official tabulation (11.6%) provided by the Ministry of Health (Kemenkes, 2008). Bivariate analysis confirms the general findings in social epidemiology. The odds of having a clinically significant mental disorder symptomatology are higher among women and among those who are old, divorced, widowed, less educated or unemployed. Individuals who reported engaging in less physical activity, suffering from chronic illness, being a heavy drinker or living in a rural area also tend to have higher

odds. Reduced odds are found among those who are consumption-rich, living in an egalitarian district, and those who have a big family. That frequent smokers seem to have lower odds may somewhat counter-intuitive, but it should be noted that this may be an artefact resulting from confounding. This is formally addressed by multivariate analysis presented later.

Figure 2.3 shows the spatial distribution of common mental disorders, facility deprivation, income inequality and precipitation anomaly across 440 districts in Indonesia. The hotspots appearing in the topmost panel seem to reflect the mental health costs of devastating earthquakes and tsunamis that occurred in Sumatra (Irmansyah et al., 2010) and in the islands of East Nusa Tenggara. The hotspot in central Sulawesi, however, might reflect the aftermath of the Poso conflict (Tol et al., 2010). The second panel shows the concentration of basic social facilities in Java and Bali islands, vividly portraying the consequences of the long-standing Java-centric development agenda. The third panel presents district-level income inequality with the Asmat district in Papua, the West Jakarta district in Java, and the Luwu Timur district in Sulawesi being the three most unequal districts. Finally, the last panel of Figure 2.3 depicts the precipitation anomaly during the June-July-August (JJA) 2007 climatological season. Java, Maluku and some parts of southern Sumatra and central Papua were drier than the normal years, whereas Kalimantan, Sulawesi, Halmahera and the rest of Sumatra were generally wetter than normal. The highest precipitation anomaly for the season was recorded in West Papua.

2.3.2 Multivariate analysis without instrumenting for poverty variable

The results of multivariate analysis, assuming exogenous poverty, are presented in Table 2.2 under the headings 'Linear', 'Poisson', 'LPM' and 'Probit'. Log per capita household expenditure is found to be statistically significant and negatively related with symptoms of mental illness in all four models. Parameterising the SRQ-20 score as a continuous variable, the linear model estimates that, *ceteris paribus*, a doubled per capita household expenditure is associated with a $0.21 \times \ln(2) = 0.16$ point reduction in the SRQ-20 score (better mental health), though it must be noted that this modelling technique does not take the skewness and the non-negativity of the SRQ-20 into account. For that reason, we fit a Poisson model, which suggests that a 1% increase in consumption expenditure is associated with a 0.11% reduction in symptoms of mental illness. This estimate means that the change in an individual's

Table 2.1: Sample characteristics and bivariate relationships

Variable	Summary statistics	Odds ratio
Mental health:		
SRQ-20 score	2.22 ± 3.29	n.a
Probable caseness (SRQ-20 ≥ 6)	11.5%	n.a
Age:		
15–24	23.0%	1.00
25–34	23.0%	1.02 ± 0.02
35–44	21.2%	1.14 ± 0.02‡
45–54	15.8%	1.41 ± 0.03‡
55–64	9.0%	1.95 ± 0.06‡
65+	8.0%	3.70 ± 0.15‡
Gender:		
Male	48.1%	1.00
Female	51.9%	1.67 ± 0.02‡
Marital status:		
Married	68.6%	1.00
Never Married	23.1%	0.74 ± 0.02‡
Divorced	1.8%	1.66 ± 0.05‡
Widowed	6.5%	2.63 ± 0.06‡
Education:		
Less than middle school	53.1%	1.00
Middle school	20.3%	0.58 ± 0.01‡
High school	21.2%	0.49 ± 0.02‡
College	5.4%	0.42 ± 0.02‡
Employment status:		
In employment or schooling	88.9%	1.00
Unemployed	11.1%	2.04 ± 0.05‡
Physical activity:		
Adequate	70.0%	1.00
Less	30.0%	1.24 ± 0.03‡
Smoking behaviour:		
Occasional or non-smoker	72.4%	1.00
Frequent smoker	27.6%	0.82 ± 0.01‡
Drinking behaviour:		
Light or non-drinker	99.5%	1.00
Heavy drinker	0.5%	1.21 ± 0.10†
Chronic illness:		
No	90.8%	1.00
Yes	9.2%	2.99 ± 0.06‡
Household size	4.59 ± 1.90	0.94 ± 0.01‡
Residential location:		
Rural	62.5%	1.00
Urban	37.5%	0.88 ± 0.04†
Per capita household exp., log	12.50 ± 0.52	0.78 ± 0.03‡
District deprivation index	-0.03 ± 1.03	0.97 ± 0.04
District inequality (Gini index)	0.25 ± 0.04	8.15 ± 0.75‡
Rainfall anomaly:		
March April May (MAM) 2007	24.82 ± 48.67	n.a
June July August (JJA) 2007	19.77 ± 61.45	n.a

Note: Sampling weight is not applied. * p < 0.10, † p < 0.05, ‡ p < 0.01.

Figure 2.3: Spatial distribution of common mental disorders, facility deprivation, income inequality and precipitation anomaly across 440 districts in Indonesia
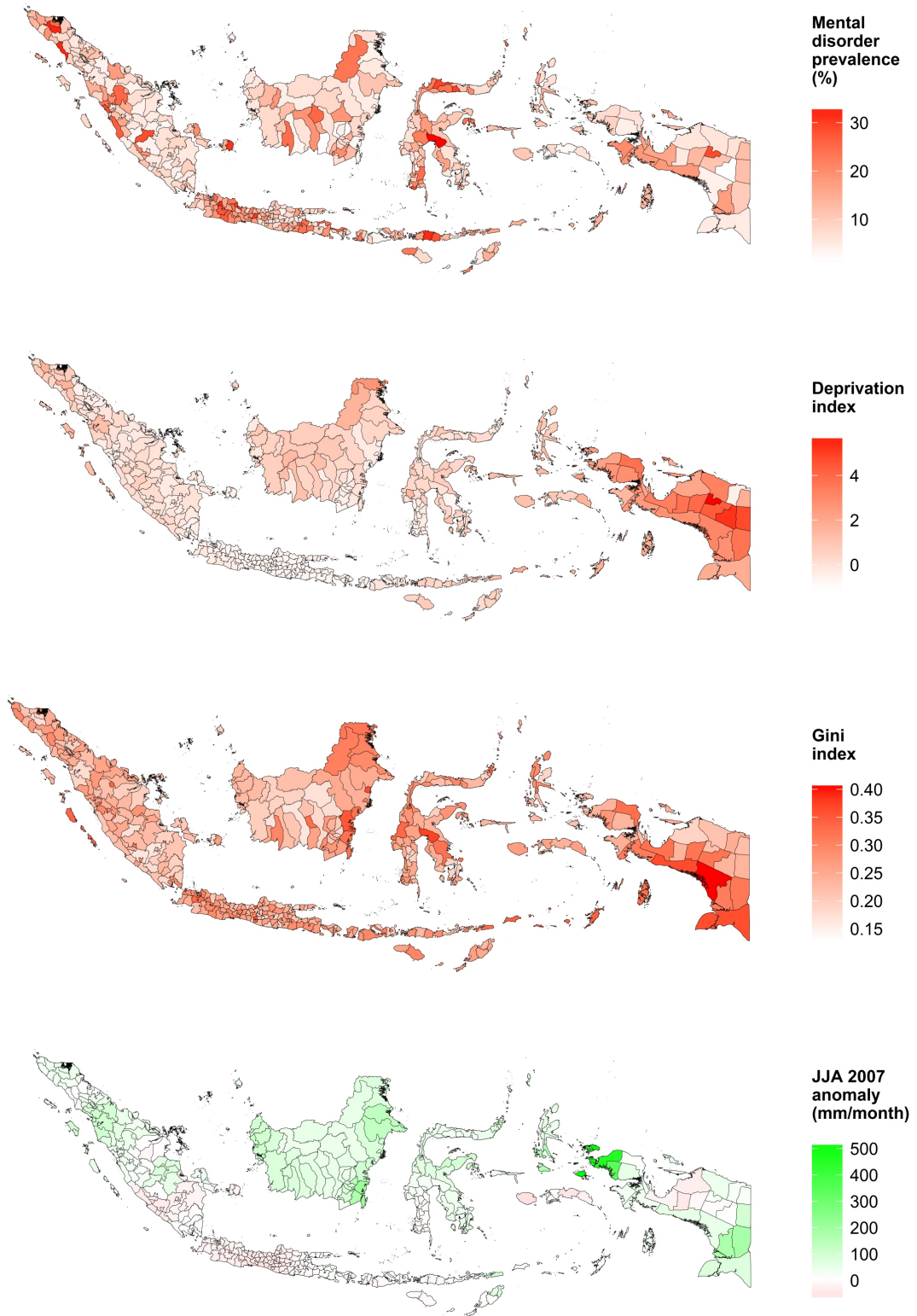
mental health status is relatively inelastic to the change in his or her consumption. The negative association remains consistent even when mental health is treated as a dichotomous variable. Both linear probability and probit models suggest that, for a typical Indonesian, a doubled consumption expenditure is associated with an approximately $0.02 \times \ln(2) = 0.01$ lower probability of having clinically significant symptoms of common mental disorders. These results demonstrate that the negative relationship between poverty and mental health is robust to distributional assumption and to alternative parameterisation.

Although frequent smoking was found to be associated with better mental health in the bivariate analysis presented previously, this is no longer the case in the multivariate analysis. After adjusting for potential confounders, frequent smokers are now estimated to have a 0.02 higher probability of having clinically significant symptoms of common mental disorders compared to those who smoke occasionally and those who do not smoke at all. Of the two contextual variables, only income inequality is statistically significant. A 0.1 point increase in the district-level Gini index (rising inequality) is associated with a 0.03 higher probability of having mental illness. This correlation provides weak support for the hypothesis that income distribution exerts a significant effect on the mental health of the population over and above the effect of individual income (Wilkinson and Pickett, 2010). The estimates for other covariates generally remain similar to those reached through the simple bivariate analysis.

2.3.3 Multivariate analysis after instrumenting for poverty variable

The results of multivariate analysis, assuming endogenous poverty, are presented in Table 2.2 under the headings 'Linear-IV', 'Poisson-IV', 'LPM-IV' and 'Probit-IV'. Apart from the assumption, these models are identical to those discussed above. We specify precipitation anomaly in two climatological seasons (MAM and JJA 2007) as instruments for per capita household consumption expenditure. Both climatological seasons coincide with the onset of the dry season in Indonesia; the JJA season also covers the start of the Riskesdas fieldwork (August 2007). The statistically significant result of test of the endogeneity of the instrumented regressor allows us to reject the null hypothesis that the poverty variable can actually be treated as exogenous. In investigating the strength of the instruments, we reject the null hypothesis that the instrument set is weak: the Kleibergen-Paap rank Wald F-statistic (F = 21.04) is well above the 10% critical value of the Cragg-Donald statistic (F = 19.80), although we

Table 2.2: Estimation results, Riskesdas 2007 national sample aged 15 or older (N = 577,548)

Predictors	Mental Health Score				Probable Caseness			
	Linear	Linear-IV	Poisson	Poisson-IV	LPM	LPM-IV	Probit	Probit-IV
Log(PCE)	-0.24 ± 0.05‡	-1.31 ± 0.55†	-0.11 ± 0.03‡	-0.62 ± 0.26†	-0.02 ± 0.00‡	-0.09 ± 0.04†	-0.02 ± 0.00‡	-0.09 ± 0.05*
Age 25–34	0.02 ± 0.02	0.03 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	-0.00 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.00‡	-0.00 ± 0.00†
Age 35–44	0.09 ± 0.03‡	0.14 ± 0.04‡	0.05 ± 0.01†	0.07 ± 0.02‡	-0.00 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	0.00 ± 0.00
Age 45–54	0.21 ± 0.04‡	0.34 ± 0.07‡	0.10 ± 0.02‡	0.16 ± 0.04‡	0.00 ± 0.00	0.01 ± 0.01†	0.00 ± 0.00	0.01 ± 0.01*
Age 55–64	0.41 ± 0.05‡	0.51 ± 0.07‡	0.18 ± 0.02‡	0.23 ± 0.03‡	0.02 ± 0.00‡	0.03 ± 0.01‡	0.02 ± 0.00‡	0.03 ± 0.01‡
Age 65+	0.81 ± 0.06‡	0.83 ± 0.07‡	0.30 ± 0.03‡	0.31 ± 0.03‡	0.06 ± 0.01‡	0.06 ± 0.01‡	0.04 ± 0.01‡	0.04 ± 0.01‡
Female	0.62 ± 0.03‡	0.63 ± 0.03‡	0.30 ± 0.01‡	0.30 ± 0.01‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.05 ± 0.00‡
Never married	-0.09 ± 0.03†	-0.01 ± 0.05	-0.06 ± 0.01‡	-0.03 ± 0.02	-0.00 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	0.00 ± 0.00
Divorced	0.35 ± 0.04‡	0.33 ± 0.05‡	0.13 ± 0.02‡	0.11 ± 0.02‡	0.03 ± 0.00‡	0.03 ± 0.00‡	0.03 ± 0.00‡	0.02 ± 0.00‡
Widowed	0.26 ± 0.03‡	0.25 ± 0.04‡	0.05 ± 0.01‡	0.05 ± 0.01‡	0.03 ± 0.00‡	0.03 ± 0.00‡	0.01 ± 0.00‡	0.01 ± 0.00‡
Middle school	-0.30 ± 0.03‡	-0.13 ± 0.09	-0.14 ± 0.01‡	-0.06 ± 0.04	-0.02 ± 0.00‡	-0.01 ± 0.01	-0.02 ± 0.00‡	-0.01 ± 0.01
High school	-0.44 ± 0.03‡	-0.09 ± 0.17	-0.22 ± 0.02‡	-0.05 ± 0.08	-0.03 ± 0.00‡	-0.01 ± 0.01	-0.04 ± 0.00‡	-0.01 ± 0.02
College	-0.55 ± 0.04‡	0.03 ± 0.29	-0.29 ± 0.02‡	-0.01 ± 0.14	-0.04 ± 0.00‡	-0.00 ± 0.02	-0.05 ± 0.00‡	-0.01 ± 0.03
Unemployed	0.25 ± 0.03‡	0.18 ± 0.04‡	0.11 ± 0.01‡	0.08 ± 0.02‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Less physical activity	0.15 ± 0.04‡	0.23 ± 0.05‡	0.07 ± 0.02‡	0.10 ± 0.02‡	0.01 ± 0.00‡	0.02 ± 0.00‡	0.01 ± 0.00‡	0.02 ± 0.00‡
Frequent smoker	0.25 ± 0.02‡	0.24 ± 0.02‡	0.13 ± 0.01‡	0.12 ± 0.01‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Heavy drinker	0.54 ± 0.11‡	0.52 ± 0.11‡	0.25 ± 0.04‡	0.24 ± 0.05‡	0.04 ± 0.01‡	0.04 ± 0.01‡	0.04 ± 0.01‡	0.04 ± 0.01‡
Chronic illness	1.26 ± 0.04‡	1.33 ± 0.05‡	0.44 ± 0.01‡	0.48 ± 0.02‡	0.11 ± 0.00‡	0.11 ± 0.00‡	0.08 ± 0.00‡	0.08 ± 0.01‡
Household size	-0.04 ± 0.01‡	-0.12 ± 0.04†	-0.02 ± 0.00‡	-0.06 ± 0.02†	-0.00 ± 0.00‡	-0.01 ± 0.00†	-0.00 ± 0.00‡	-0.01 ± 0.00†
District deprivation	-0.00 ± 0.05	-0.03 ± 0.06	-0.00 ± 0.03	-0.01 ± 0.03	-0.00 ± 0.00	-0.01 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.00
District inequality	3.59 ± 1.07‡	4.18 ± 1.17‡	1.71 ± 0.50‡	1.83 ± 0.54‡	0.28 ± 0.09†	0.32 ± 0.10‡	0.28 ± 0.09‡	0.32 ± 0.10‡
Urban	-0.07 ± 0.06	0.24 ± 0.19	-0.04 ± 0.03	0.11 ± 0.09	-0.00 ± 0.00	0.02 ± 0.01	-0.00 ± 0.00	0.02 ± 0.02
Intercept	1.68 ± 0.06‡	1.50 ± 0.11†	0.52 ± 0.03‡	0.41 ± 0.07‡	0.07 ± 0.00‡	0.06 ± 0.01‡	-0.00 ± 0.00	0.02 ± 0.02
Estimator	OLS	GMM	GMM	GMM	OLS	GMM	ML	ML
Instruments' validity		0.97		0.83		0.73		
Log(PCE)'s exogeneity		3.33*				3.33*		
Instruments' strength		21.04				21.04		2.16

Note: * p < 0.10, † p < 0.05, ‡ p < 0.01.

note that this critical value is appropriate only for the i.i.d. normal sample (Baum et al., 2007). That precipitation anomaly strongly predicts consumption expenditure over and above the effect of other exogenous covariates should not be particularly surprising. This can be explained by the fact that, in 2007, nearly half (41%) of the members of the Indonesian labour force were employed in the agriculture sector while at the same time only 16% of the agricultural land was covered by irrigation infrastructure (World Bank, 2014b). A test of overidentifying restrictions also returns favourable results. The non-significant Hansen's J-statistic seems to suggest that both instruments give the same information.

Having assessed the quality of the instruments, we now interpret the results. With continuous parameterisation, the linear model estimates that a doubled per capita household expenditure reduces the SRQ-20 score by approximately 1 point (better mental health), while the Poisson model estimates that a 1% increased consumption leads to a 0.62% decrease in symptoms of mental illness. In concordance, dichotomous parameterisation suggests that raising one's consumption twofold brings about a 0.06 point lower probability of having clinically significant symptoms of common mental disorders. These estimated effects are approximately five times stronger than those obtained prior to instrumenting for per capita household expenditure, hinting that perhaps the bias due to measurement error rather than simultaneity or reverse causality was more dominant. Notice that now, after instrumenting for poverty status, individuals' mental health status becomes moderately elastic to the change in consumption expenditure. A similar pattern was found earlier by Ettner (1996) who analysed data from the US. Overall, except for education variables, whose estimated effects have become statistically indistinguishable from zero, all other covariates remain in the same direction as they were prior to instrumenting for the poverty variable.

Table 2.3 displays the sensitivity of the estimated poverty effect to the set of control variables entered into the model. It appears that the effect is robust to the choice of model specification. In the appendix accompanying this article, we further refit the model with (1) urban-rural stratification, (2) control function estimator and (3) random effects estimator. The finding remains: consumption-poor individuals have a higher probability of suffering from mental illness.

Table 2.3: Estimates of Log(PCE) in different specifications

Specifications	Mental Health Score		Probable Caseness	
	Linear-IV	Poisson-IV	LPM-IV	Probit-IV
Full model	-1.31 ± 0.55†	-0.62 ± 0.26†	-0.09 ± 0.04†	-0.09 ± 0.05*
Without unemployed	-1.32 ± 0.55†	-0.63 ± 0.26†	-0.09 ± 0.04†	-0.09 ± 0.05*
Without less physical activity, frequent smoker, heavy drinker and chronic illness	-1.12 ± 0.57†	-0.51 ± 0.26†	-0.08 ± 0.04*	-0.08 ± 0.05
Without deprivation and inequality	-1.45 ± 0.62†	-0.70 ± 0.30†	-0.12 ± 0.05†	-0.11 ± 0.06*
Without unemployed, less physical activity, frequent smoker, heavy drinker and chronic illness	-1.12 ± 0.57*	-0.50 ± 0.26*	-0.08 ± 0.04*	-0.08 ± 0.05
Without unemployed, less physical activity, frequent smoker, heavy drinker, chronic illness, deprivation and inequality	-1.17 ± 0.64*	-0.53 ± 0.29*	-0.10 ± 0.05*	-0.09 ± 0.06

Note: * $p < 0.10$, † $p < 0.05$, ‡ $p < 0.01$.

2.4 Discussion and conclusion

Despite the claim that poverty causes mental illness in low- and middle-income countries, empirical evidence remains scarce. Little has been done to address the question of whether the observed wealth-health relationship is causal or just associational. The present study attempts to fill this gap by exploiting seasonal precipitation anomaly as a form of natural experiment that randomly determines poverty status in Indonesia. Results suggest that poverty causes poor mental health. Holding all other covariates constant, halving one's consumption expenditure raises the probability of having mental illness by 0.06 point, or, in terms of elasticity, a 1% decrease in consumption brings about 0.62% more symptoms of common mental disorders. This study finds that the effect of poverty on mental health is approximately five times stronger than is conventionally estimated, which may be indicative of the fact that measurement error rather than reverse causality was the main source of bias (Ettner, 1996). The effect is robust to varying distributional assumptions, model specifications, estimation techniques and sample stratification. This supports the general finding in social epidemiology (Lund et al., 2010).

The present study also investigates the association between district-level income inequality and mental health. It is consistently estimated that income inequality correlates negatively with mental health over and above the effect of poverty. Individuals

living in unequal districts are found to have a higher probability of suffering from mental illness than those who live in more egalitarian districts. This is consistent with the recent finding of Filho et al. (2013), who conducted a multilevel study in the Brazilian context. This also weakly supports the broader idea of the income inequality hypothesis put forward by Wilkinson and Pickett (2010). Additionally, the present study found that women, older people and those who are divorced or widowed tend to have a higher probability of suffering common mental disorders. This is, again, consistent with the existing literature on mental health in developing countries. Finally, negative health behaviours such as less physical activity, frequent smoking and heavy drinking are all related to lower levels of mental health.

This study has a number of limitations. The first pertains to the core assumption of instrumental variable estimation. For this method to work properly, one must maintain three strong assumptions, namely the relevance condition, the validity condition and the exclusion restriction. Not all of these are testable. While it has been shown through the weak identification test that seasonal precipitation anomaly strongly predicts per capita household expenditure (hence satisfying the relevance condition), there is no empirical test capable of examining the exclusion restriction (Freedman, 2009, 2010; Hernán and Robins, 2006). This must be established *a priori*. The quality of an instrumental variable estimation is only as good as its story; here it rests ultimately on the untestable assumption that precipitation anomaly is indeed a random variate perfectly uncorrelated with any determinants of mental health, and that it does not affect an individual's mental health except through its influence upon consumption expenditure. While this assumption is plausible in the context of agriculture-dependent societies, Sarsons (2015) recently cautions that individual income in well-irrigated regions is less sensitive to rainfall fluctuations. The second limitation relates to the possible interpretation of the causal parameter recovered by instrumental variable estimation, namely as a local average treatment effect (LATE) (Angrist and Pischke, 2008). Under the LATE interpretation, the causal parameter obtained in this study is simply the average effect of poverty on mental health for individuals whose income fluctuates in accordance with the randomisation provided by the natural experiment (the average treatment effect of the compliers). Of course, generalising this causal effect to the entire population of Indonesia requires additional layers of assumption, but given that a large proportion of the Indonesian workforce is employed in the largely rain-dependent agriculture sector, we believe that even the LATE parameter is worthy of consideration. This study is also limited by the cross-sectional nature of the data. Future studies may take advantage of a longitudinal design so that temporal order can be incorporated into the model.

Despite its limitations, the present study contributes to the literature on mental health in developing countries in several ways. First, this study is among the few studies that attempt to address the endogeneity problem in the estimation of the mental health effect of poverty. Second, using a large and representative data from Indonesia, this study demonstrates that the adverse effect of poverty on mental health is not merely attributed to the self-selection bias that threatens small-sample community or facility studies. Third, considering both the use of a standard mental health and poverty measure and the fact that Indonesia is the most populous developing country after China and India, this study provides a finding that is suitable for cross-national comparison. Finally, the present study shows that poverty remains an important determinant of mental health irrespective of whether it is treated as an exogenous or as an endogenous variable. Indonesian policy-makers now have reason to believe that poverty alleviation efforts can have considerable impact on the mental health of the population. Mental health can be improved not only by influencing individuals' health knowledge and behaviour but also by implementing a more equitable economic policy. Policy-makers may also want to consider a greater public investment in the long-neglected mental health service sector, which would certainly benefit the nation as a whole given that the burden of mental illness is borne not only by the patients but also by their family members. Additionally, research has shown that mental illness is costly for a nation's economy (Lund et al., 2013). Furthermore, according to the referral scheme of Indonesia's recently launched version of the universal healthcare system (Lancet, 2014), every prospective patient is required to report to the nearest primary care centre prior to visiting a hospital; mental healthcare service, then, must be surely made available at the lowest level of the referral hierarchy. Unless such a provision is available, the mental health of Indonesians will continue to be overlooked.

❀ ❀ ❀

Chapter 3

Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis

Abstract: Studies on self-rated health outcomes are fraught with problems when individuals' reporting behaviour is systematically biased by demographic, socio-economic or cultural factors. Analysing the data drawn from the Indonesia Family Life Survey (IFLS) 2007, this paper aims to investigate the extent of differential health reporting behaviour by demographic and socio-economic status among Indonesians aged 40 and older (N = 3,735). Interpersonal heterogeneity in reporting style is identified by asking respondents to rate a number of vignettes that describe varying levels of health status in targeted health domains (mobility, pain, cognition, sleep, depression and breathing) using the same ordinal response scale that is applied to the self-report health question. A compound hierarchical ordered probit model is fitted to obtain health differences by demographic and socio-economic status. The obtained regression coefficients are then compared to the standard ordered probit model. We find that Indonesians with more education tend to rate a given health status in each domain more negatively than their less-educated counterparts. Allowing for such differential reporting behaviour results in relatively stronger positive education effects. There is a need to correct for differential reporting behaviour using vignettes when analysing self-rated health measures in older adults in Indonesia. Unless such an adjustment is made, the salutary effect of education will be underestimated.

Keywords: self-rated health, socio-economic status, reporting heterogeneity, anchoring vignette, Indonesia

3.1 Introduction

Both resource constraints and the multidimensionality of health concepts being studied often necessitate the collection of self-rated health (SRH) data. SRH measures, which ask individuals to report their health status either in general or on a specific health domain using an ordinal response scale, require no specialist intervention during data collection, are relatively cheap and quick to obtain and are feasible to implement in large-scale surveys. In addition to the belief that SRH can capture aspects of health that cannot be tapped by objective measure (Wallace and Herzog, 1995), research has shown that SRH is highly correlated with assessments provided by health professionals (Ferraro, 1980) and that is also a strong predictor of mortality (Idler and Benyamini, 1997) as well as healthcare utilisation (van Doorslaer et al., 2004).

Notwithstanding these benefits, the use of SRH in the study of socio-economic inequalities in health becomes fraught with serious problems when individuals have different expectations, knowledge or standards of what constitutes good health. For example, when experiencing an identically severe health problem, poor individuals may paradoxically report better health than their richer counterparts (Figure 3.1) simply because the poor have a much higher tolerance to health problems than the rich (Sen, 2002). This is known in the literature as 'reporting heterogeneity' (Shmueli, 2002), 'differential item functioning' (King et al., 2004), 'response category cut-point shift' (Murray et al., 2002), 'scale of reference bias' (Groot, 2000), or simply 'differential reporting' (Lindeboom and van Doorslaer, 2004).

To address this problem, the anchoring vignette method has been proposed (Tandon et al., 2002; King et al., 2004; King and Wand, 2007; Wand, 2013). By means of this method, researchers can identify the individual-specific reporting style by asking respondents to rate a number of vignettes (hypothetical scenarios) that describe varying levels of health status in a health domain using the same ordinal response scale that is applied to the self-report health. Then, if one is willing to assume that, apart from random error, each vignette is perceived in the same way by all respondents (vignette equivalence assumption) and that they apply exactly the same standard to judge both their own health status and those of the vignettes (response consistency assumption), one can fit a compound hierarchical ordered probit (CHOPIT) model (King et al., 2004) to identify health inequalities that are free from bias due to heterogeneous reporting style.

Using anchoring vignette, it has been shown that among older individuals in eight

Figure 3.1: Illustration of response-scale heterogeneity

In general, how is your health?
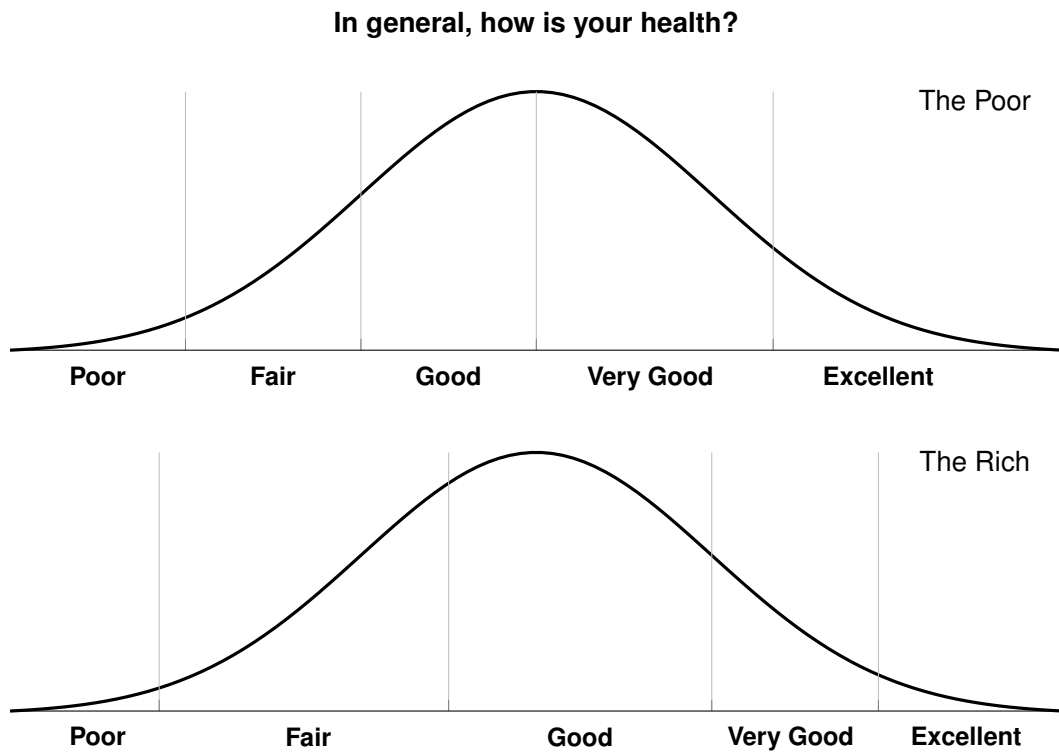
European countries, there is strong evidence for the existence of differential health reporting by education level. Bago d'Uva et al. (2008a) found that highly educated older Europeans tend to have higher expectation of health than their less educated peers and suggested that accounting for differences in the reporting of health is important because 'measured health inequalities by education are often underestimated, and even go undetected, if no account is taken of these reporting differences.' (Bago d'Uva et al., 2008a: 1375). However, when the authors analysed data from three most populous developing countries (China, India and Indonesia) they found that in Indonesia and India 'there are either no differences in reporting by education or the better educated are more likely to report very good health' (Bago d'Uva et al., 2008b: 362). This finding defies conventional expectation; the authors then speculated that perhaps the Chinese sample, which has a higher level of education than the Indonesian and Indian, were more able to comprehend the vignette exercise.

Motivated by these mixed findings, this paper aims to investigate whether there is evidence for differential reporting behaviour by demographic and socio-economic status (SES) among Indonesians. We analyse data from the fourth wave of the Indonesia Family Life Survey (IFLS 2007), which is among the very few population studies conducted in developing countries that employed a vignette rating module. The present study departs from the existing application of anchoring vignette method

in Indonesia (Bago d'Uva et al., 2008b) in its use of a newer dataset and of fewer and simpler vignettes, as well as in its analysis of a more homogeneous age group.

3.2 Methods

3.2.1 Data

The data is drawn from the IFLS 2007, which is a multi-purpose household longitudinal study that collects information from more than 30,000 individuals from 12,000 households living in 260 districts in Indonesia and is representative of about 83% of the entire population (RAND, 2007). The IFLS 2007 is the only IFLS wave that has vignette module. Because the module was administered to only a fraction of study participants, the sample of this study is, by design, limited to 3,735 adults aged 40 and older. These individuals were asked to report their self-assessment of health, but only one-third of them (1,245 individuals) were subjected to the vignette rating questionnaire. Further details regarding sampling and ethical procedure are available in the IFLS's documentation (RAND, 2007).

3.2.2 Measures

Survey respondents were asked to evaluate their own health in six health domains (mobility, pain, cognition, sleep, depression and breathing) using the question 'Overall in the last 30 days, how much of a problem did you have with ...?'. Responses were recorded using a five-category ordinal scale: (1) none, (2) mild, (3) moderate, (4) severe, and (5) extreme. In addition to this self-assessment, randomly selected respondents were also asked to evaluate the health status of hypothetical persons described in the vignettes. For each domain, three vignettes of varying severity were presented; respondents were then asked to think about these persons' experiences as if they were their own and to rate the health status of the persons portrayed in the hypothetical scenarios in the same way they had rated their own health earlier. Vignettes were presented in the order of mild–moderate–severe health problem and responses were recorded using the same response scale applied to the SRH. For ease of understanding, we reverse-coded the response scale so that a score of 5 represents very good health and a score of 1 represents very poor health. The exact wording of these questionnaires and vignettes is provided in Appendix B.

The SES variables are education (entered as a dummy variable representing those who completed the 9-year compulsory education) and the logarithm of per capita household asset value. We opted to use these SES indicators rather than the usual indicators of income or expenditure because many respondents were already at the retirement age (56 or older). In this case, education is particularly relevant because it is probably the best measure of SES for older adults (Grundy and Holt, 2001). In later life, education serves as a good proxy for permanent income and is less endogenous than income as it is usually fixed early in life (Grundy and Holt, 2001). Per capita household asset value was measured from the total value of land, property, vehicles, poultry, livestock, fish ponds, hard stem plants, household appliances, household furniture and utensils, savings, deposit, stocks, receivables and jewellery owned by the household members. Like education, assets are also considered as less endogenous than income due to their accumulative nature (Mu, 2014).

We also include respondents' age groups (40–49, 50–59, 60–69, 70+), gender, marital status (married and unmarried), family size (dummy variable for those living with more than four household members), and urban or rural residential location.

3.2.3 Modelling techniques

For each health domain, we first fit an ordered probit (OPROBIT) model (Greene and Hensher, 2010) to estimate the effect of demographic and SES variables on health. Then, we refit the same specification with a compound hierarchical ordered probit (CHOPIT) model (King et al., 2004) that generalises the OPROBIT by allowing cut-points or thresholds to be different across individuals (Figure 3.2).

The CHOPIT model is comprised of two components: the self-assessment and the vignette rating component. In the self-assessment equation, we write the unobserved perceived level of health as:

$$y_i^* \quad \sim \quad N(\mu_i, 1) \tag{3.1}$$

$$\mu_i \quad = \quad X_i \beta \tag{3.2}$$

with subscript i denotes individuals responding to SRH questionnaire. Individuals' actual health level μ_i varies as a linear function of observed covariates X_i with parameter vector β. Respondents then turn their perceived level of health y_i^* into reported

ordinal category y_i via the following observation mechanism:

$$y_i = k \quad \text{if} \quad \tau_i^{k-1} < y_i^* < \tau_i^k, \quad k = 1, \ldots, K \tag{3.3}$$

where

$$-\infty = \tau_i^0 < \tau_i^1 < \tau_i^2 < \ldots < \tau_i^K = \infty \tag{3.4}$$

To allow for individual-specific response category cut-point shift, thresholds τ_i are modelled as a linear function of observed covariates X_i with parameter vector γ and are identified in the model using information obtained from the vignette rating exercise.

$$\tau_i^1 = X_i \gamma^1 \tag{3.5}$$

$$\tau_i^k = \tau_i^{k-1} + X_i \gamma^k, \quad \text{for} \quad k = 2, \ldots, K \tag{3.6}$$

In the vignette rating equation, we write the perceived level of health of the person described in vignette j evaluated by survey respondent i as:

$$z_{ij}^* \sim N(\theta_j, \sigma_j^2) \tag{3.7}$$

The actual health level of the person described in the vignette (θ_j) is assumed to be identical for every respondent, hence formalising the 'vignette equivalence' assumption. As in the self-assessment part of the model, respondents then turn the perceived level of health z_{ij}^* into the same K ordinal category via similar mechanism:

$$z_{ij} = k \quad \text{if} \quad \tau_{ij}^{k-1} < z_{ij}^* < \tau_{ij}^k, \quad k = 1, \ldots, K \tag{3.8}$$

Thresholds in the vignette rating equation are determined by the same γ parameter as in the self-assessment part, but note that the sample used in each model component need not be identical (two different samples drawn from the same large population may be used). The appearance of the same γ parameter vector in both self-assessment and vignette rating components thus formalises the 'response consistency' assumption. Figure 3.2 shows a graphical illustration comparing CHOPIT with the ordinary OPROBIT model. We shall note that (1) X, y and z are observed; (2) y^*, μ, z^*, θ and τ are unobserved; and that (3) β and γ are vectors of parameters to be estimated from the data.

For identification and model comparability purposes, the standard ordered probit normalisation restriction (intercept is fixed at zero; variance is set to one) (Wand et al., 2011) is imposed upon both OPROBIT and CHOPIT models. Then, formal

Figure 3.2: Illustration of OPROBIT and CHOPIT models
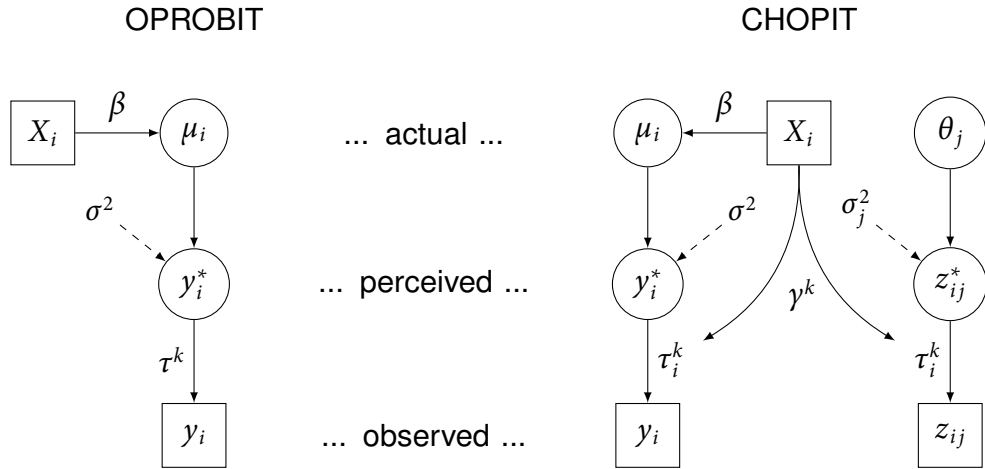
tests of reporting homogeneity (H_0 : all $\gamma = 0$) and parallel cut-point shift (H_0 : $\gamma^1 = \gamma^2 = \ldots = \gamma^{K-1}$) (Jones et al., 2013) are performed after acquiring the estimate of the CHOPIT model, accompanied by graphical illustrations when necessary. To facilitate interpretation, we also compute the partial effect of relevant variables on the probability of reporting very good health ($\Pr\left[y_i = K | X, \gamma, \beta\right] = 1 - \Phi\left[\tau_i^{K-1} - X_i\beta\right]$) (Jones et al., 2013).

Only complete observations are used in the modelling exercise, yielding a sample size of 3,069 individuals in the SRH equations (82% of the original sample) and 939–1,130 individuals in the vignette rating equations (75–90% of the original sample).

3.3   Results

We begin with a description of the sample. The mean age is 53.95 (SD = 10.81, median = 52, IQR = 16); half of the sample (52.8%) are female and 20% are unmarried. The majority of the sample (77.4%) live with at least five household members; about half (49.18%) live in urban area and only one-third (37.92%) completed the 9-year compulsory education. Per capita household asset value is log-normally distributed with a mean equal to USD 1,660 (SD = 3,800, median = 721, IQR = 1,368). The well-behaved histograms in Figure 3.3 show that respondents seem to understand the vignette rating exercise very well: the ratings of moderate health problems are symmetrically distributed, while those of mild and severe health problems are left- and right-skewed, respectively. Overall, there is no marked difference between the characteristics of the SRH sample and those of the vignette sample.

Figure 3.3: Distribution of vignette ratings (1 = extreme, 2 = severe, 3 = moderate, 4 = mild, 5 = none)
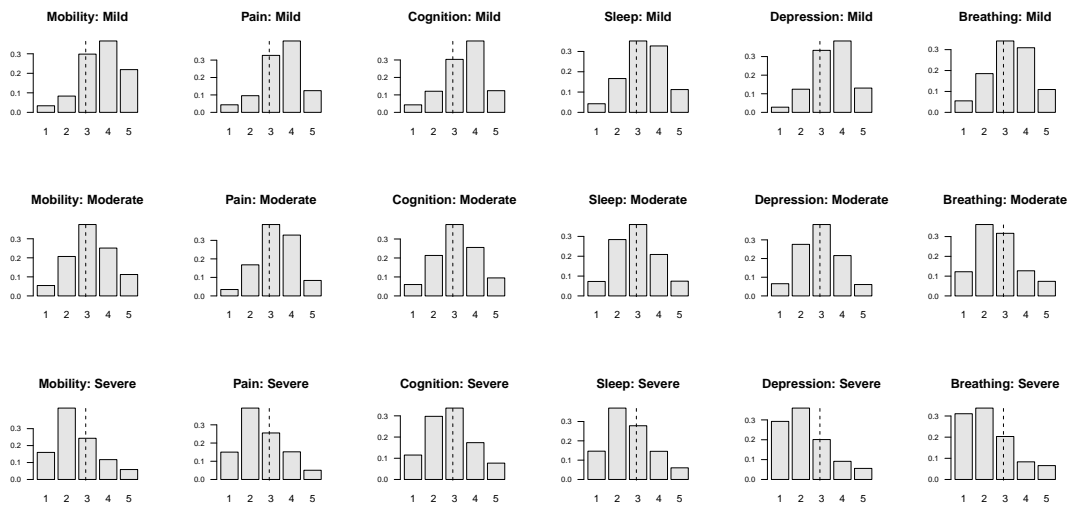
Table 3.1: Test of reporting homogeneity and parallel cut-point shift

Test	Mobility	Pain	Cognition	Sleep	Depression	Breathing
Reporting homogeneity	50.70*	93.86‡	82.28‡	99.03‡	105.46‡	98.81‡
Parallel cut-point shift	32.16	66.99‡	33.40	53.06‡	67.98‡	46.12†

Note: Reported are χ^2 statistic with 36 degrees of freedom (reporting homogeneity) and 27 degrees of freedom (parallel cut-point shift); * $p < 0.10$, † $p < 0.05$, ‡ $p < 0.01$.

The regression coefficients obtained from the OPROBIT model are represented by hollow circles plotted in the left panel of both Figures 3.4 and 3.5. Assuming that respondents apply identical thresholds, the results suggest a general trend that: (1) health deteriorates with age in a possibly non-linear fashion (except in the depression domain), (2) women report worse health than men (except in the breathing domain), and (3) the better educated are healthier than those with minimal education attainment (except in the depression domain). Being unmarried is associated with lower health status in the sleep and depression domains, but there is no evidence for such association in other domains. The models show that there seems to be no statistically discernible effect of family size and urban-rural residential location on health in all six domains. Wealth, however, seems to have a positive impact on health in the mobility, cognition, sleep and depression domains if only to a very small degree. This can be understood as monetary welfare is no longer a good indicator of SES in later life.

What happen when we relax the reporting homogeneity assumption by fitting a CHOPIT model? Regression coefficients predicting the latent health index in each domain (β) are shown using solid circles in the left panels of Figures 3.4 and 3.5, while

Figure 3.4: Estimation results for mobility, pain and cognition domains (main coefficients (β) in left panel, threshold coefficients (γ) in right panel, intercepts in threshold equation not shown)
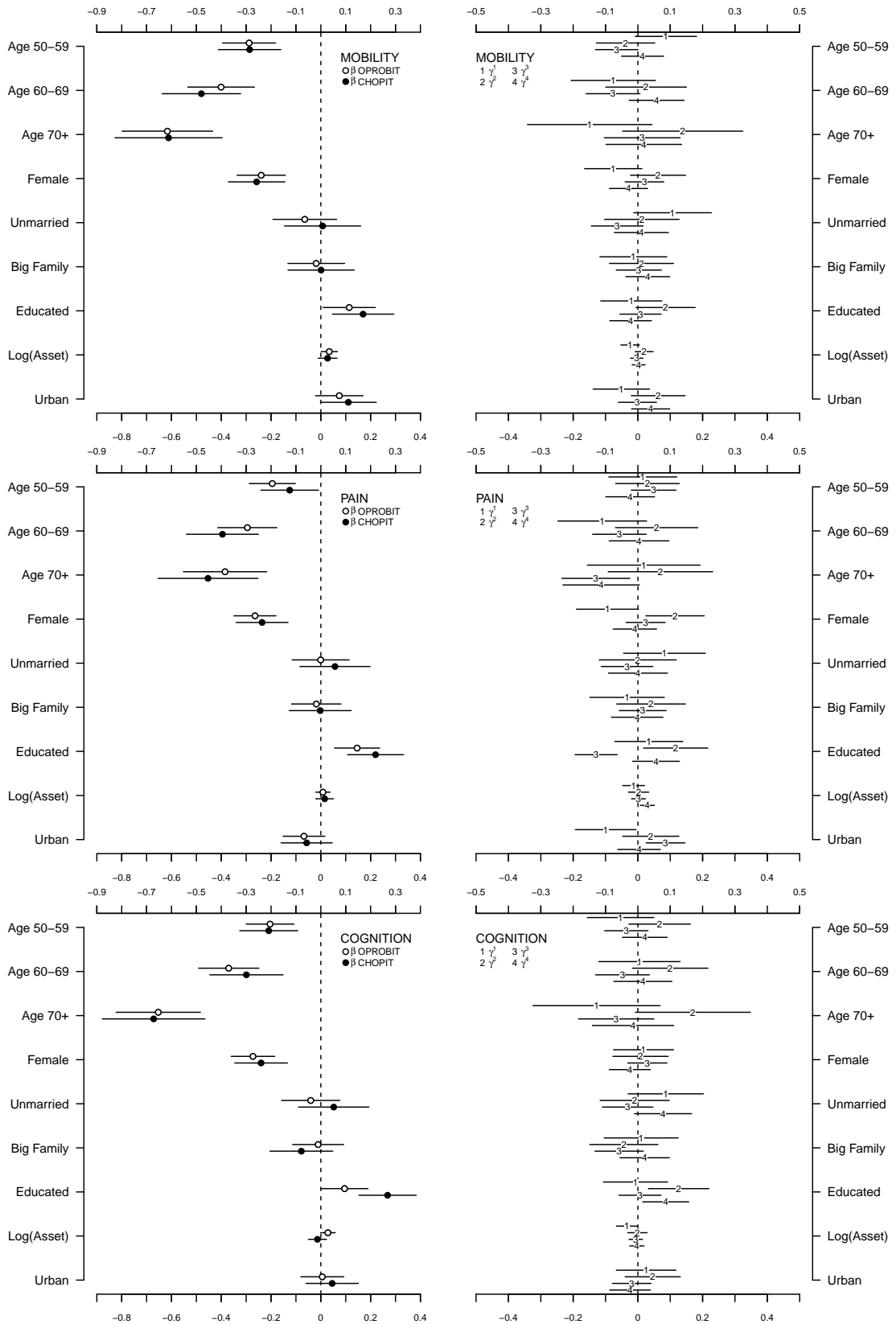
Figure 3.5: Estimation results for sleep, depression and breathing domains (main coefficients (β) in left panel, threshold coefficients (γ) in right panel, intercepts in threshold equation not shown)
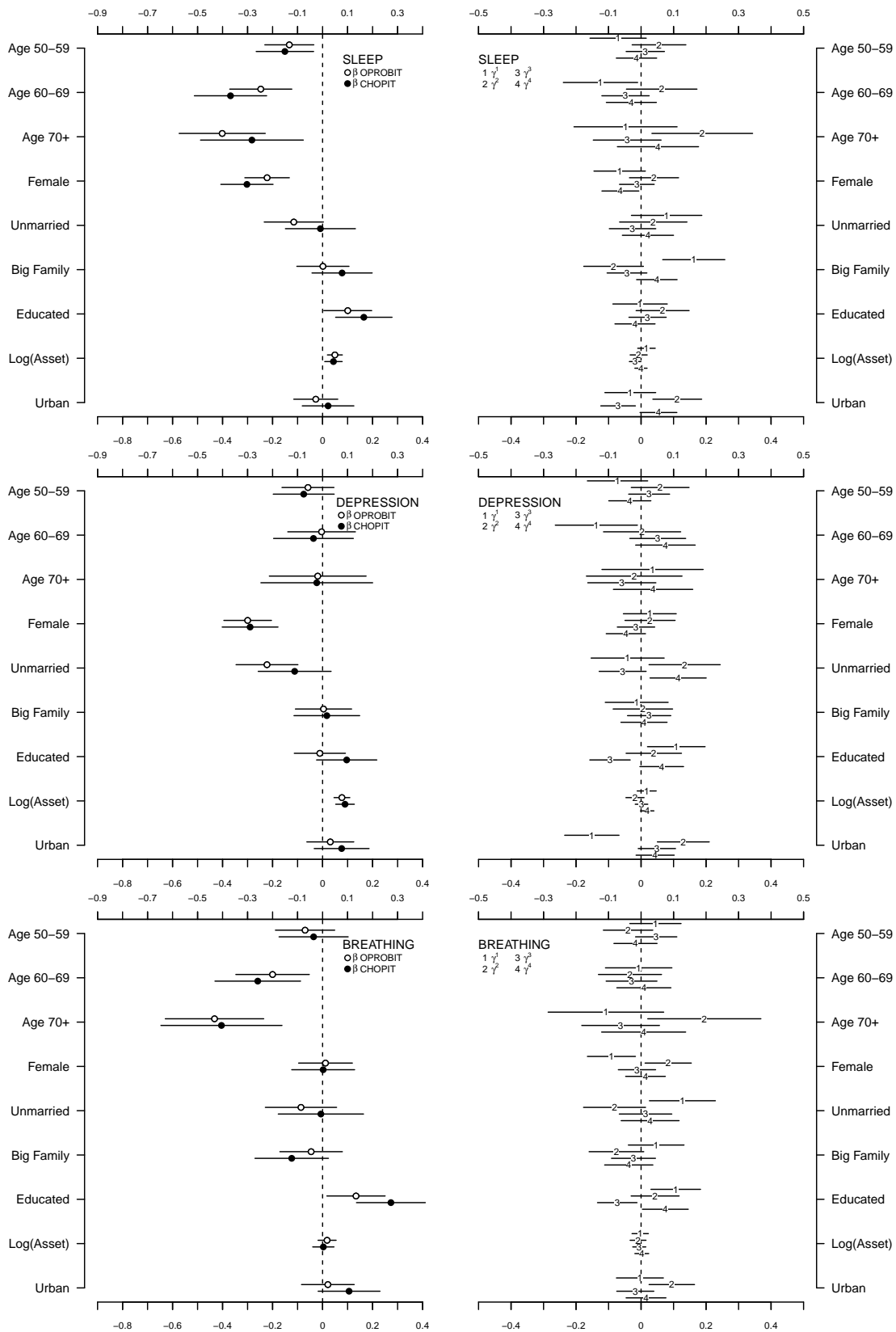
Figure 3.6: Estimated location of vignette rating (θ_j)
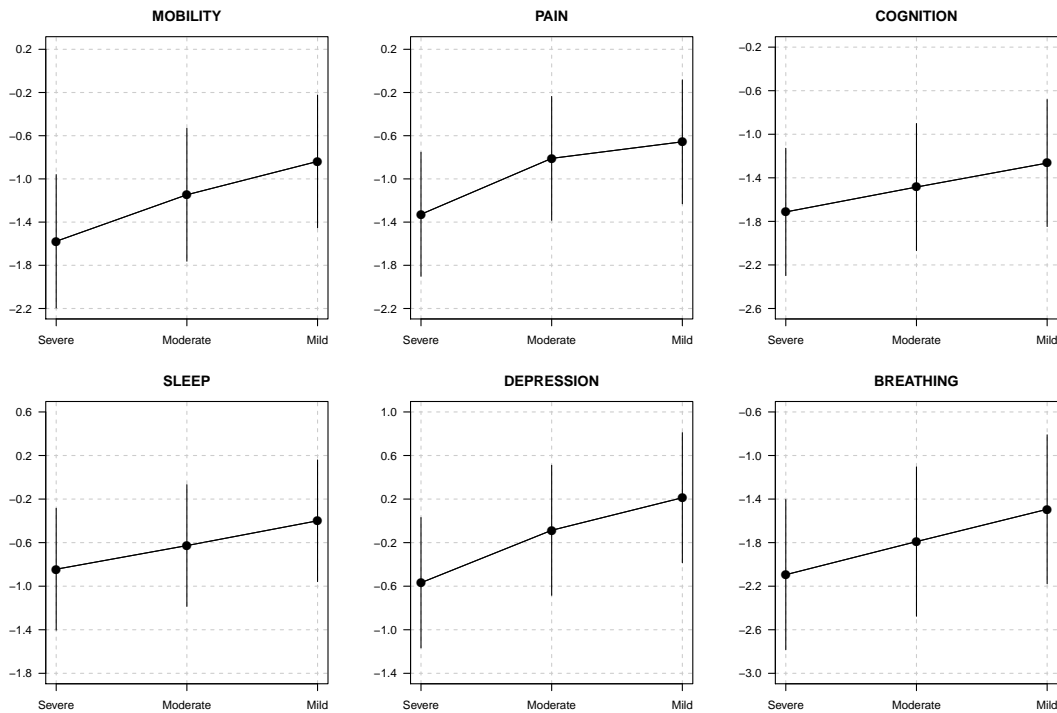
those predicting the individual-specific thresholds (γ) are shown using numbers in the right panels of the figures. An omnibus test of reporting homogeneity in each domain (Table 3.1) rejects the joint null hypothesis that all coefficients in the threshold equation are equal to zero at conventional significance levels, indicating that respondents do not necessarily apply identical cut-points when transferring their latent health indices onto the ordinal categories. In other words, there seems to be disagreement as to what constitutes good health among the respondents; some may have higher or lower standards than others. The statistically significant results of a global test of parallel cut-point shift in each domain (except in mobility and cognition; see Table 3.1) further indicate that respondents' reporting behaviour depends on the covariates in a complex way. The relationship between the thresholds and the covariates is not necessarily characterised by a simple linear function. Respondents, however, seem to agree on the levels of health described in the vignettes. As shown in Figure 3.6, the estimated vignette locations in the latent health space are in concordance with the intended ordering. This confirms the earlier exploratory analysis presented in Figure 3.3.

Allowing for interpersonal differences in reporting style does alter the point estimate of each β coefficient (Figures 3.4 and 3.5), but with the exception of that of education, the correction is practically negligible. In fact, when we test for reporting homogeneity by each covariate, only education variable is consistently statistically significant in all

six health domains (Table 3.3). After adjusting for reporting heterogeneity, the 95% confidence intervals of age, gender, family size, wealth and urban/rural residential location still overlap largely with those of the OPROBIT model, and their interpretation remains. For marital status, the adjustment brings significant change in the sleep and depression domains where the health-protective effect of being married diminishes after correcting for the lower expectation of health among married individuals.

Table 3.3: Test of reporting homogeneity by each covariate

Variable	Mobility	Pain	Cognition	Sleep	Depression	Breathing
Age 50–59						
Age 60–69	◎	○		◎	◎	
Age 70+		◎		◎		
Female		○		◎		
Unmarried	○		○	◎	△	○
Big Family				△		
Educated	○	△	△	○	△	△
Log(Asset)			△			
Urban		○		△	△	△

Note: ○ p < 0.10, ◎ p < 0.05, △ p < 0.01.

Figure 3.7: Effect of education on vignette ratings' cut-points
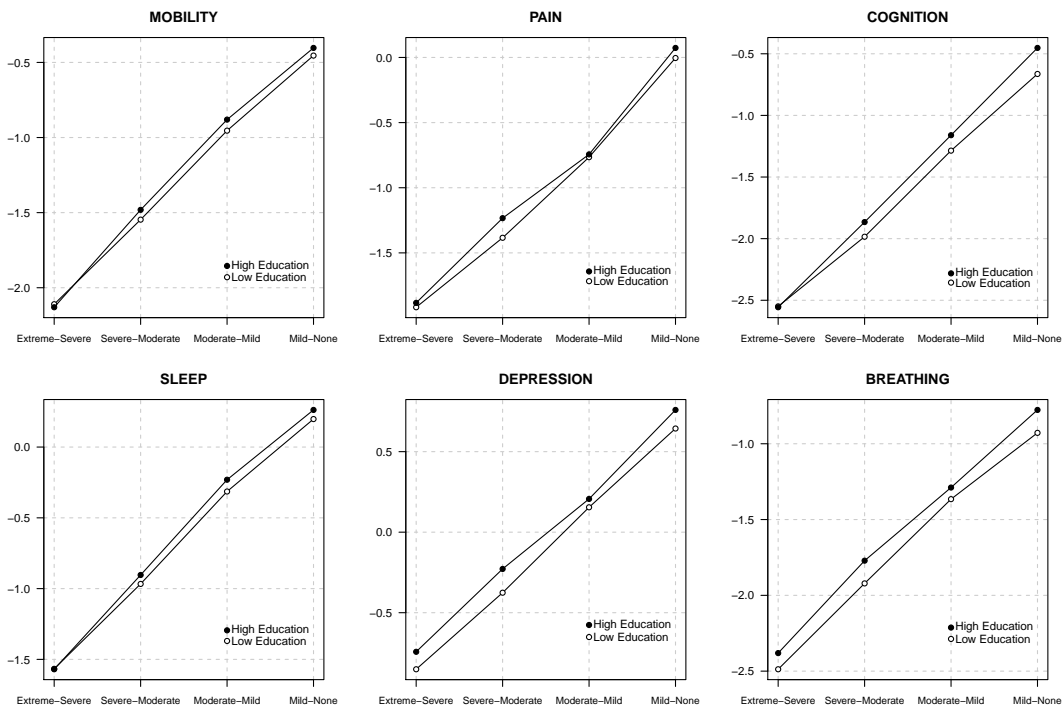
A significant correction is observed with regards to education. The positive education effect in some threshold equations across health domains (shown in the right panels of Figures 3.4 and 3.5) suggests that Indonesians with high levels of educational

Table 3.4: Test of parallel cut-point shift by each covariate

Variable	Mobility	Pain	Cognition	Sleep	Depression	Breathing
Age 50–59						
Age 60–69						
Age 70+		◎		○		
Female		○		○		
Unmarried					△	
Big Family				○		○
Educated		△	△		◎	◎
Log(Asset)		○				
Urban		◎		△	△	○

Note: ○ p < 0.10, ◎ p < 0.05, △ p < 0.01.

attainment tend to rate a given health status more negatively than their less-educated counterparts. This is consistent with the educated being better informed; they have higher health standards. Thus, adjusting for this difference magnifies the positive effect of education on health status in all domains (Table 3.2). Most notably, the adjustment raises the estimated difference in the probability of reporting very good health between the well- and less educated Indonesians in cognition and breathing domains by two- to threefold. The CHOPIT coefficients also tend to be more precisely estimated. Figure 3.7 shows how education level alters respondents' thresholds, which are used to transfer the latent health index onto the ordinal categories. The plots suggest that reporting behaviour depends on education in a rather complex way, reiterating the results of the test of parallel cut-point shift (Tables 3.1 and 3.4). Finally, following the method of Voňková and Hullegie (2011), we test whether or not the adjustment to reporting heterogeneity is sensitive to the choice of vignettes used in the model by refitting the CHOPIT model with a single vignette at a time, predicting the latent health index and then calculating the Pearson correlation coefficient between pairs of predicted values in each domain. As shown in Figure 3.8, the strong correlations suggest that the adjustment is insensitive to the choice of hypothetical scenarios.

3.4   Discussion and conclusion

Applying anchoring vignette methodology to a sample of older Indonesians, this study investigates the extent of differential reporting behaviour by demographic and socio-economic status in six health domains. We find that allowing for interpersonal heterogeneity in response style consistently magnifies the positive effect of education on health in all domains. One plausible interpretation of this finding is that educated Indonesians, who are likely to be well-informed and aware of their well-being, have

higher standards or expectations with regards to health than their less-educated counterparts. This indicates that health disparity by education might actually be wider than it is usually reported. Unless an adjustment is made for this systematic differential, the salutary effect of education will be underestimated. This finding is in line with an earlier observation in Europe (Bago d'Uva et al., 2008a), but it contradicts a previous study showing the overestimation of education effect among the general population in Indonesia (Bago d'Uva et al., 2008b). Such a divergence might result from our (1) use of fewer and simpler vignettes, (2) analysis of a more homogeneous age group, and/or (3) use of a newer dataset. We also find significant modification in the effect of marital status in the sleep and depression domains. The detrimental effect in these domains of being unmarried diminishes after correcting for the higher expectations of health prevalent among unmarried individuals. Otherwise, we find little difference when calibrating the effects of other demographic variables. Overall, these findings suggest that policy-maker cannot only rely on people's perception of health when attempting to measure the reality. Studies on self-reported health outcomes particularly in developing countries should consider administering vignettes and using them to arrive at unbiased report on health inequality.

Table 3.2: Partial effects of education on the probability of reporting very good health

Domain	OPROBIT	CHOPIT
Mobility	0.03 ± 0.01†	0.04 ± 0.02‡
Pain	0.06 ± 0.02‡	0.08 ± 0.02‡
Cognition	0.03 ± 0.02†	0.09 ± 0.02‡
Sleep	0.04 ± 0.02†	0.06 ± 0.02‡
Depression	-0.00 ± 0.02	0.03 ± 0.02*
Breathing	0.03 ± 0.01†	0.06 ± 0.01‡

Note: * $p < 0.10$, † $p < 0.05$, ‡ $p < 0.01$.

The generalisability of this study is limited by the restricted age group being analysed as well as by the small sample size. Future studies may collect more extensive vignette data so that statistical inferences can be extended to general population and so that stratified analysis by age, gender or urban/rural residential location can be performed. We also note that the validity of the anchoring vignette method hinges critically on the maintenance of both vignette equivalence and response consistency assumptions. A number of studies have investigated the plausibility of these assumptions; some have found positive supports (King et al., 2004; Rice et al., 2011; van Soest et al., 2011), while others report possible violations (Bago d'Uva et al., 2011b; Bolt et al., 2014; Datta Gupta et al., 2010; Hirve et al., 2013). In this study, there is always the possibility that these assumptions are violated. Vignette equivalence, for example, might not hold if some individuals perceive one of the vignettes more as being in a serious health condition because he or she has experienced or taken care of a family member who went through similar illness. Also, unmeasured respondents' past experience of

Figure 3.8: Correlations among pairs of predicted health index in each domain
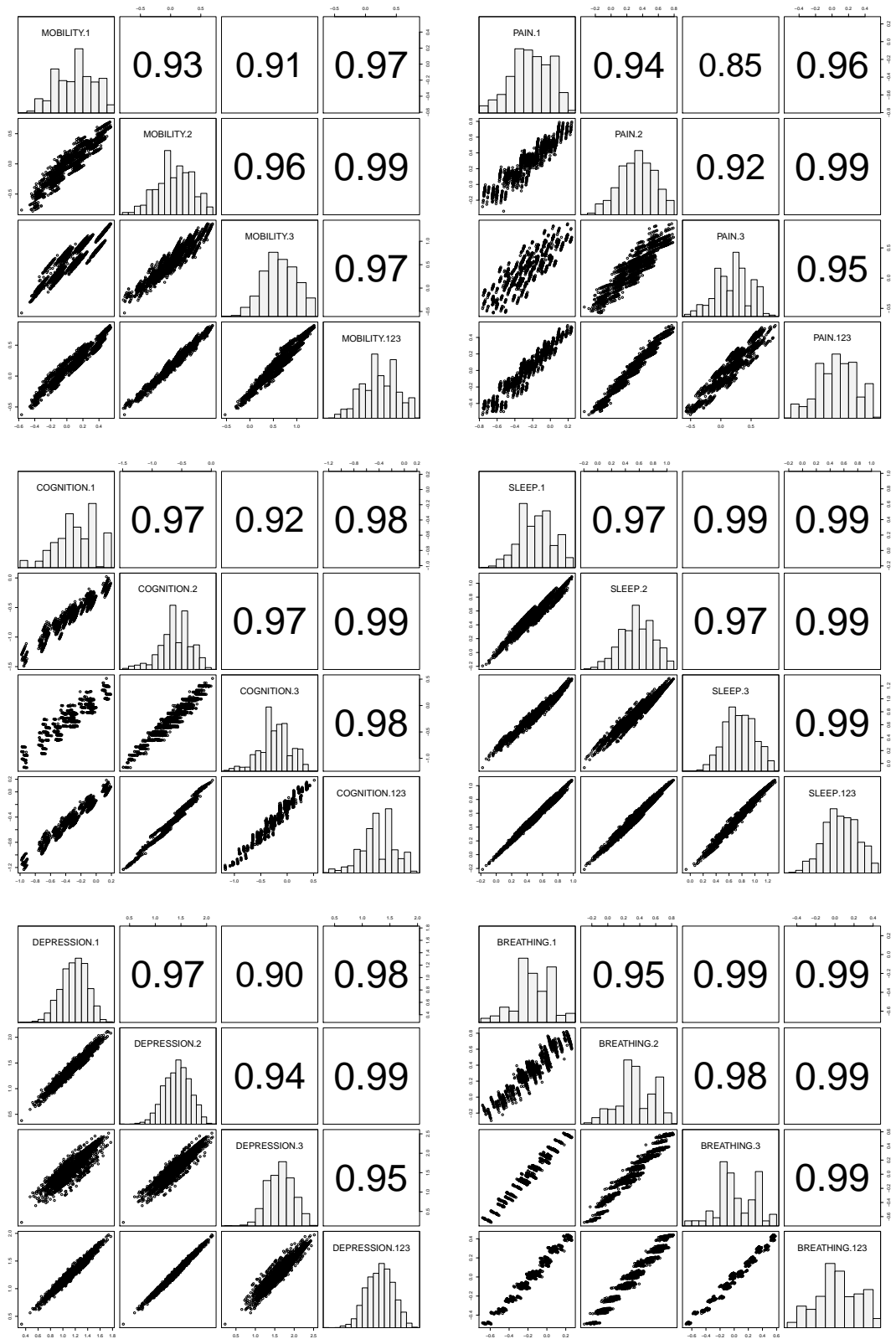
adverse events, surgery or major illness, could have an effect on their perception of the vignettes as well as on their response to SRH questionnaire. While we have not provided a direct test for these assumptions, we are at least reassured that our analysis is insensitive to the choice of vignettes used in the model. Furthermore, by asking survey respondents to rate the vignettes as if they assess their own health condition, the IFLS study has at least tried to reinforce the response consistency assumption during data collection stage.

Anchoring vignette is a promising method that offers a direct way of handling interpersonal incomparability in self-report measure. Although methodologists have extended the original anchoring vignette method (King et al., 2004) to accommodate more complex situations (Bago d'Uva et al., 2011a; Bolt et al., 2014; Kapteyn et al., 2007; Paccagnella, 2011; Peracchi and Rosetti, 2012; van Soest et al., 2011; Wand, 2013), adequate attention should also be given to the fundamental matters of question wording (Au and Lorgelly, 2014; Hirve et al., 2013) and ordering (Buckley, 2008; Hopkins and King, 2010). We believe that, given its cost-effectiveness and feasibility in large-scale surveys, SRH and anchoring vignette have the potential to play a greater role in public health research in now-decentralised Indonesia, where more than 500 local administrations must struggle with a scarcity of competent health workers (Rokx et al., 2010; Thabrany, 2006) as well as with the high cost of collecting objective health measures.

❊ ❊ ❊

Chapter 4

The double burden of malnutrition in Indonesia: Social determinants and geographical variations

Abstract: The presence of simultaneous under- and overnutrition has been widely documented in low- and middle-income countries, but global nutritional research has seen only a few large-scale population studies from Indonesia. We investigate the social determinants as well as the geographical variations of under- and overnutrition in Indonesia using the largest public health study ever conducted in the country, the National Basic Health Research 2007 (N = 645,032). Multilevel multinomial logistic regression and quantile regression models are fitted to estimate the association between nutritional status and a number of socio-economic indicators at both the individual and district levels. We find that: (1) education and income reduce the odds of being underweight by 10–30% but at the same time increase those of overweight by 10–40%; (2) independent from the compositional effect of poverty, income inequality is detrimental to population health: a 0.1 increase in the Gini coefficient is associated with an 8–12% increase in the odds of an individual's being both under- and overweight; and (3) the effects that these determinants have upon nutritional status are not necessarily homogeneous along the continuum of body mass index. Equally important, our analysis reveals that there is substantial spatial clustering of areas with elevated risk of under- or overnutrition across the 17,000-island archipelago. As of 2007, undernutrition in Indonesia remains a 'disease of poverty', while overnutrition is one of affluence. The income inequality accompanying Indonesia's economic growth may aggravate the dual burden of under- and overnutrition. A more equitable economic policy and a policy that improves living standards may be effective for addressing the double burden.

Keywords: double burden of malnutrition, Indonesia, social determinants, multilevel model, quantile regression

4.1 Introduction

The simultaneous presence of under- and overnutrition within populations of developing countries undergoing rapid economic transition has been widely documented (Gillespie and Haddad, 2003; Jehn and Brewis, 2009). The changes in dietary intake patterns and leisure-time activities associated with industrialisation and urbanisation are known to have contributed to an increased prevalence of obesity in numerous countries (Popkin, 1998, 1999); at the same time, the problem of undernutrition remains undefeated. This dual burden, which may also exist within a single household (Doak et al., 2005; Lee et al., 2012), is costly for the health as well as the economy of a nation. Undernutrition impairs cognition (Sandjaja et al., 2013) and physical development (Mani, 2012), reduces economic productivity (Victora et al., 2008), raises the mortality rate, and even induces an intergenerational cycle of malnutrition (Barker, 1997); on the other extreme of the nutritional spectrum, overnutrition is known to increase the risk of non-communicable diseases, inflate health care costs (Cawley and Meyerhoefer, 2012; Withrow and Alter, 2011), and reduce overall quality of life.

The body of nutritional epidemiology and development economics research suggests that, over and above the biological aspects of age and sex, socio-economic status, along with a number of ecological factors such as urban environment, area-level economic development and income inequality, seems to consistently determine the social distribution of malnutrition (Doak et al., 2005; Ha et al., 2011; Lee et al., 2012; Rahmanian et al., 2014; Roemling and Qaim, 2013; Shafique et al., 2007; Subramanian et al., 2007; Vaezghasemi et al., 2014). Notwithstanding the increasing number of studies in this stream of research, the literature, however, does not yet include sufficient evidence from Indonesia, which is the most populous developing country after China and India. To date, empirical evidence tends to come from South Asia, Africa and Latin America (see for example Corsi et al., 2011 or Jehn and Brewis, 2009). Little is known about the double burden of malnutrition in Indonesia, despite the fact that it is in a state of rapid economic and epidemiologic transition where industrialisation, urbanisation and political decentralisation are met with rising income inequality, widening regional disparities and a diminishing rate of poverty reduction (World Bank, 2014a). All existing studies focusing on Indonesia (Doak et al., 2005; Oddo et al., 2012; Roemling and Qaim, 2013; Vaezghasemi et al., 2014; Winkvist et al., 2000) have thus far (1) dealt specifically with the coexistence of under- and overnutrition within the same households (double burden households), (2) concentrated only on particular population subgroups (women) or small geographical areas (relatively affluent western Indonesia), or (3) failed to account for the influence of macro-level

contextual factors. A large-scale population study covering the entire 17,000-island archipelago is, to our knowledge, non-existent as 'there is little awareness of the double burden of malnutrition issues, be it in the government, the public or professional circles' (Shrimpton and Rokx, 2013: 6; see also WHO, 2010).

Exploiting the fact that a large, nationally representative sample has recently become available, this paper aims to investigate the social determinants as well as the geographical variations of under- and overnutrition among adults aged 15 years and older living in 440 districts in Indonesia. In particular, we are interested in understanding (1) the pattern of association between an individual's socio-economic position and his or her nutritional status; (2) the influence of contextual factors at the district level on one's probability of being under- or overweight; and (3) the geographical distribution of the risk of malnutrition within the archipelago after accounting for the effects of observable socio-demographic determinants. Because understanding *who* gets the diseases and *where* the diseases strike is imperative for tackling the double burden (UNSCN, 2006: 7), insights gained from this analysis are of high relevance for the formulation of evidence- or need-based intervention measures—especially for policy targeting in Indonesia as well as in other parts of the developing world.

4.2   Methods

4.2.1   Data

The data are drawn from the Riset Kesehatan Dasar (National Basic Health Research; henceforth 'Riskesdas') 2007. Managed by the Ministry of Health of the Republic of Indonesia, Riskesdas is the largest public health research initiative ever carried out in the country. The repeated cross-sectional study includes 987,205 individuals from 258,366 households residing in all 440 districts and is thus representative of the Indonesian population (Kemenkes, 2008). Its size and geographical coverage clearly distinguish Riskesdas from the Indonesia Family Life Survey (IFLS) dataset (30,000 individuals living in 260 districts) that was analysed in some earlier studies (Doak et al., 2005; Roemling and Qaim, 2013). Hence, in addition to the benefit of additional statistical power, Riskesdas also offers the opportunity for researchers to extend their inferences to the deprived and usually neglected islands of the archipelago (Sulawesi, Maluku, Halmahera, Nusa Tenggara and Papua). Informed consent was obtained prior to interview and participants' confidentiality was strictly protected. Further

details regarding ethical and sampling procedures are available through Kemenkes (2008).

Included in the sample of this study are adults aged 15 and older. After excluding pregnant women and individuals of extreme height (less than 100 cm or more than 200 cm) or weight (less than 25 kg or more than 200 kg), the final sample size was 645,032 individuals. This corresponds to approximately 97% of all adults who participated in the Riskesdas 2007 study.

4.2.2 Measures

The dependent variable is adult nutritional status as indicated by body mass index (BMI). BMI is calculated by dividing an individual's weight (in kilograms) by his or her squared height (in metres); following the standard adopted by the government of Indonesia (Kemenkes, 2008), the individual is then classified as 'underweight' (BMI < 18.5), 'normal' ($18.5 \leq$ BMI < 25), 'overweight' ($25 \leq$ BMI < 27), or 'obese' (BMI $\geq$ 27). However, for the sake of computational feasibility as well as ease of understanding, we collapse the last two categories (see also Gurrici et al., 1998 and WHO Expert Consultation, 2004 for discussions regarding BMI cut-off points for obesity in the Indonesian context). Both the categorical representation of nutritional status and the continuous measure of BMI are used in the following statistical analysis.

The individual-level socio-economic explanatory variables of interest are education (indicator variables for primary education or less, secondary school, high school and college or more), employment status (dummy indicators for those who are not employed or in school) and per capita household expenditure (PCE) serving as a proxy for individual income. In Indonesia, as in many parts of developing world, the individual income measure is usually not available (reliable) due to the high prevalence of both self- and seasonal employment (60–70% in Indonesia; Nazara, 2010). The literature (Deaton and Zaidi, 2002; Howe et al., 2012) suggests that PCE is capable of delivering a good approximation for permanent income due to its insensitivity to intermittent income shock that is inherent in informal economy. Both the logarithmic and the quintile representations of PCE are used in the analysis.

At the district level, we include continuous measures of income inequality, level of economic development (median PCE in million Indonesian rupiah) and index of deprivation. Income inequality is measured using the Gini index on a scale of 0–1 and was derived from the PCE measure available in the Survei Sosial Ekonomi

Table 4.1: Exploratory factor analysis of district deprivation index

Proportion of village without ...	Factor loading	Summary statistics	
Communication facilities	0.86	Explained variance	88%
Electricity	0.81	Cronbach's α	0.82
Street lighting	0.76	Eigenvalue	3.58
Healthcare facilities	0.75	KMO	0.80
TV signal coverage	0.73	N	454
Education facilities	0.65		
Entertainment facilities	0.30		

Nasional (National Socio-economic Survey) 2007 dataset using the method described by Milanovic (1997). Subsequently, to aid with interpretation, this Gini index is multiplied by a factor of 10 before being used in any statistical modelling exercises. The deprivation index was calculated from the Potensi Desa (Village Census) 2008 dataset, covering all 75,410 villages across the archipelago. Factor loadings, proportion of shared variance as well as other statistics obtained during the derivation of the index are shown in Table 4.1. It is noteworthy, at this point, that the inclusion of measures of area-level economic development and facility deprivation alongside the income inequality variable allows researchers to separate the contextual effect of income inequality from the compositional effect of poverty (Subramanian and Kawachi, 2004).

In the statistical models described next, we also control for survey respondent age group (15–24, 25–34, 35–44, 45–54, 55–64, or 65+), sex (dummy variable for female survey respondents), marital status (married, never married, divorced or widowed), self-report physical activity (indicator variable for those reporting inadequate physical activity according to the criteria set by Kemenkes, 2008), urban/rural residential setting (dummy variable for urban residency), and number of household members. Continuous covariates are either centred to their respective grand means (log per capita household expenditure, Gini index) or to a representative value (household size of 3, deprivation index equals 0) so that the intercept can be meaningfully interpreted. Accordingly, for categorical variables the references are: married male aged 15–24 with primary school or less education, currently employed or in school, living in rural area with income at the poorest quintile and engaging in adequate physical activity.

4.2.3 Modelling techniques

In order to predict the nutritional status of individual i residing in district j with three possible nominal outcomes s = {underweight, normal, overweight} and unknown

intra-cluster correlation induced by hierarchical dependence (Figure 4.1), we specify the following generalised linear mixed model (GLMM) with logit link-function (Goldstein, 2011; Rabe-Hesketh and Skrondal, 2012):

$$\log\left[\frac{\Pr(y_{ij} = s)}{\Pr(y_{ij} = \text{normal})}\right] = X_{ij}\beta^{(s)} + u_j^{(s)}, \quad s = \text{underweight, overweight} \quad (4.1)$$

$$\begin{pmatrix} u_j^{(s)} \\ u_j^{(s+1)} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u^{(s)}}^2 & \rho\sigma_{u^{(s)}}\sigma_{u^{(s+1)}} \\ \rho\sigma_{u^{(s)}}\sigma_{u^{(s+1)}} & \sigma_{u^{(s+1)}}^2 \end{pmatrix}\right] \quad (4.2)$$

In this specification, X is the matrix of explanatory variables at both individual (x_{ij}) and district (x_j) level that also includes a constant term and cross-level interaction terms ($x_{ij} \times x_j$). The unknown parameter vector $\beta^{(s)}$ captures the average effect of each explanatory variable on the probability of an adult being underweight or overweight relative to having a normal BMI. To facilitate interpretation, $\beta^{(s)}$ is reported as a relative risk (odds) ratio ($RRR = \exp\{\beta^{(s)}\}$).

Figure 4.1: Illustration of hierarchically correlated data
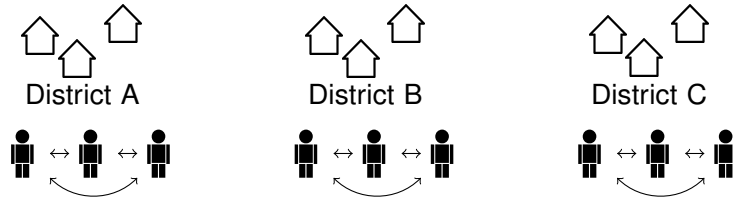
The $u_j^{(s)}$ is the contrast- and district-specific random effect that is assumed to be uncorrelated with X and is normally distributed with zero mean and variance to be estimated from the data. A parameter capturing the correlation (ρ) between random effects $u_j^{(s)}$ and $u_j^{(s+1)}$ is also obtainable from the model and is particularly useful for measuring the strength as well as the direction in which the risks of under- and overnutrition covary within a single district. Such an interpretation has been used in some earlier studies in India (Subramanian and Smith, 2006; Subramanian et al., 2007); in fact, Corsi et al. (2011) have recently called for a wider use of this parameter to arrive at a formal way of assessing the existence of the double burden of malnutrition within a given geographical area. Furthermore, the fact that the estimated random effect $u_j^{(s)}$ is independent from the influence of observed socio-demographic characteristics is also helpful for the purpose of risk mapping or ranking (see Ackerson et al., 2008 for such an application to Indian data). It is important to note, however, that the standard multinomial logit model maintains the assumption of the independence of irrelevant alternatives (IIA), meaning that 'adding or deleting alternatives does not affect the odds among the remaining alternatives' (Long and Freese, 2006: 243). This should

not be a particularly serious problem for the present study because the outcomes can plausibly be assumed to be distinct from one another (McFadden, 1973) and, more formally, because here we relax the restrictive IIA properties via the introduction of correlated random effects $u_j^{(s)}$ and $u_j^{(s+1)}$ into the model (Grilli and Rampichini, 2006; Hensher et al., 2005).

As an alternative to the multinomial outcome modelling exercise, which may suffer from a loss of information due to the arbitrariness of cut-off points, we also specify a quantile regression model (Koenker, 2005; Koenker and Hallock, 2001) that uses the continuous representation of BMI as the outcome variable. The model is given as follows:

$$Q_q(y_i) = X_i\beta^{(q)} + e_i^{(q)}, \quad q = 0.05, 0.10, \ldots, 0.95. \tag{4.3}$$

In this specification, $Q_q(y_i)$ denotes the q-th conditional quantile of BMI, X is the matrix of predictors with a constant term included, $\beta^{(q)}$ is the vector of parameters capturing the effect of each explanatory variable on the q-th conditional quantile while holding all other covariates constant, and $e_i^{(q)}$ is the asymmetrically weighted absolute residual. Unlike in the linear model, neither specific distributional assumption nor homoscedasticity is assumed for the error term, making this non-parametric modelling technique relatively robust to the influence of outliers.

The fact that one can obtain $\beta^{(q)}$ estimates for a range of conditional quantiles and allow each predictor to have an impact on both the location and scale parameters of the model is useful for the purpose of understanding the heterogeneity in the relationship between BMI and its determinants. This possibility of obtaining a more complete picture of change in the conditional distribution of BMI is undoubtedly of particular interest from a public health perspective where monitoring both the upper and lower extremes of BMI is critical. It should be noted, however, that, unlike in mean regression, the conditional quantile is not generally equal to its unconditional one (Firpo et al., 2009; Jolliffe, 2011). For purposes of computational feasibility with our large dataset, we address the clustering of individuals within districts by means of specifying a cluster-robust variance-covariance estimator (Machado et al., 2014; Santos Silva and Parente, 2013) instead of fitting a multilevel quantile regression model (Geraci, 2014; Geraci and Bottai, 2014).

4.3 Results

4.3.1 Descriptive and bivariate analysis

Table 4.2 presents descriptive statistics and measures of bivariate association between nutritional status and its predictors. BMI is approximately normally distributed (mean = 22.05 $\mathrm{kg/m^2}$, median = 21.52 $\mathrm{kg/m^2}$), albeit with some positive excess of kurtosis. The estimated national prevalence of underweight is 14.4% while that of overweight is 17.9%; despite our additional data cleaning procedure (Section 4.2.1), these figures remain very close to the official tabulation released by the Ministry of Health (14.8% and 19.1%, respectively; Kemenkes, 2008). These clearly show that, in 2007, one in three Indonesian adults was potentially suffering from nutritional problems and that the double burden of malnutrition in the country consisted relatively equally of both extremes of nutritional status.

In the sample, sex is distributed equally; and the majority of survey respondents (92%) are of working age (15–64 years-old). About two-thirds of them are married; half have not completed the nine-year compulsory education; and most (70%) report adequate physical activity. Two-thirds of adults participating in the study live in a rural area; the average number of household members across residential settings is 4.6 persons per household; and the unemployment rate is at about 11%. Median monthly individual income is 258,421 Indonesian rupiah (USD 26), while the mean of the corresponding figure at the district level is IDR 265,638 (USD 27). Income inequality ranges from 0.13 (most egalitarian) to 0.40 (least egalitarian) with the mean equal to 0.25.

Bivariate association is presented in the last two columns of Table 4.2. As can be expected from a dataset that has large statistical power, nearly all parameters are precisely estimated. The odds of being both under- and overweight generally increase with being older (notably at age 65 and older), female, having inadequate physical activity, and living in a less egalitarian neighbourhood. Marriage, education, employment and income clearly protect Indonesians from being underweight, but they also increase the probability of being overweight. Larger household size is negatively associated with overnutrition, but there is no statistically discernible effect on undernutrition. Consistent with the pattern observed across the world, urban environments in Indonesia also seem to be obesogenic. A rather unexpected result, however, comes from the deprivation index. A priori, we would expect the coefficient for deprivation to have a positive sign in the underweight equation, yet at this early stage of analysis, our bivariate exploration suggests that the more deprived a region is, the smaller the

Table 4.2: Sample description and bivariate analysis (N = 645,032)

Variable	Descriptive statistic	Unadjusted odds ratio Underweight	Overweight
Nutritional status:			
Body mass index	22.05 ± 3.81		
Normal	67.7%		
Underweight	14.4%		
Overweight	17.9%		
Age group:			
Age 15–24	22.9%	1.00	1.00
Age 25–34	22.7%	0.40 ± 0.01	2.72 ± 0.04
Age 35–44	21.3%	0.33 ± 0.01	4.26 ± 0.07
Age 45–54	16.0%	0.45 ± 0.01	4.39 ± 0.09
Age 55–64	9.1%	0.80 ± 0.02	3.44 ± 0.08
Age 65+	8.0%	1.55 ± 0.03	2.21 ± 0.06
Sex:			
Male	48.8%	1.00	1.00
Female	51.2%	1.15 ± 0.01	1.89 ± 0.03
Marital status:			
Married	68.3%	1.00	1.00
Never married	23.4%	2.07 ± 0.03	0.29 ± 0.01
Divorced	1.8%	1.51 ± 0.04	0.84 ± 0.02
Widowed	6.5%	2.63 ± 0.04	0.91 ± 0.02
Education:			
Primary school or less	53.3%	1.00	1.00
Middle school	20.3%	0.92 ± 0.01	0.92 ± 0.01
High school	21.1%	0.68 ± 0.01	1.22 ± 0.02
College	5.3%	0.50 ± 0.01	1.78 ± 0.05
Employment status:			
In employment or schooling	88.9%	1.00	1.00
Unemployed	11.1%	2.07 ± 0.03	0.65 ± 0.01
Physical activity:			
Adequate physical activity	70.1%	1.00	1.00
Less physical activity	29.9%	1.47 ± 0.02	1.18 ± 0.02
Residential setting:			
Rural	62.6%	1.00	1.00
Urban	37.4%	0.95 ± 0.02	1.78 ± 0.04
Household size and income:			
Household size	4.59 ± 1.90	1.00 ± 0.00*	0.97 ± 0.00
Log(PCE)	12.50 ± 0.51	0.74 ± 0.01	1.81 ± 0.03
District characteristics:			
Median PCE (million Rupiah)	0.27 ± 0.08	0.32 ± 0.06	11.45 ± 2.01
Deprivation (standardised)	-0.03 ± 1.03	0.91 ± 0.02	0.81 ± 0.03
Inequality	0.25 ± 0.04	1.02 ± 0.03*	1.33 ± 0.05

Note: * $p > 0.10$; standard errors are adjusted for the clustering of individuals within 440 districts.

The prevalence of obesity as defined by BMI ≥ 30 kg/m² is 3.44%.

odds of the residents being both over- *and* underweight. Whether this is simply an artefact of confounding is to be tested in the multivariate analysis presented next.

4.3.2 Multilevel multinomial logistic regression analysis

Having identified potential risk factors for under- and overweight through a simple bivariate procedure that does not take confounding into account, we now fit a series of multilevel multinomial logistic regression models to estimate the independent effect of each predictor on nutritional status (Table 4.3). The analysis is conducted in a stepwise manner: first, we fit an age-sex adjusted model (Null Model) before introducing the complete set of explanatory variables in the second model (Full Model 1); we further characterise the relationship between individual income and nutritional status by replacing the logarithmic parametrisation with indicators of income quintile (Full Model 2); finally, we consider the possibility of effect modification by interacting the female indicator with individual income and income inequality (Interaction Model). Goodness of fit is assessed by means of monitoring the Akaike/Bayesian information criterion statistic (AIC/BIC) such that models with smaller AIC/BIC are preferred over those with larger statistic.

The age-sex adjusted model (Null Model) shows that, compared to their male counterparts, Indonesian women are more vulnerable to both under- and overnutrition. Undernutrition seems to be more prevalent in early adulthood (15–24 years old) and later life (65 years old and older) than in middle age. In contrast, the risk of overnutrition seems to increase with age, peak at 45–54 years old, and then gradually decrease throughout the life course although the odds of being overweight are still about two times greater among the elderly than the youngest adults (15–24 years old). The random part of the model tells us that there seems to be a small negative correlation ($\rho = -0.19$) between the district-specific effects determining the probability of being under- or overweight. This means that places with high risk of undernutrition tend to be the ones with low risk of overnutrition; in other words, the double burden of malnutrition does *not* generally exist within the same districts in Indonesia. These age, sex and geographical patterns persist even when additional variables are introduced into subsequent models.

In fully specified models (Full Model 1, Full Model 2, Interaction Model), it is estimated that being underweight is negatively associated with being married, having a high education level, being employed, having a large household size, and having

Table 4.3: Adjusted odds ratio obtained from multilevel multinomial logistic models

Predictors	Null Model		Full Model 1		Full Model 2		Interaction Model	
	Underweight	Overweight	Underweight	Overweight	Underweight	Overweight	Underweight	Overweight
Intercept	0.30 ± 0.01	0.06 ± 0.00	0.19 ± 0.00	0.06 ± 0.00	0.22 ± 0.01	0.05 ± 0.00	0.19 ± 0.00	0.06 ± 0.00
Individual characteristics:								
Age 25–34	0.39 ± 0.00	2.77 ± 0.04	0.57 ± 0.01	1.99 ± 0.03	0.57 ± 0.01	1.99 ± 0.03	0.57 ± 0.01	1.99 ± 0.03
Age 35–44	0.32 ± 0.00	4.46 ± 0.06	0.51 ± 0.01	3.02 ± 0.05	0.51 ± 0.01	3.02 ± 0.05	0.51 ± 0.01	3.02 ± 0.05
Age 45–54	0.43 ± 0.01	4.64 ± 0.06	0.66 ± 0.01	3.15 ± 0.05	0.66 ± 0.01	3.17 ± 0.05	0.66 ± 0.01	3.15 ± 0.05
Age 55–64	0.77 ± 0.01	3.62 ± 0.06	1.09 ± 0.02	2.55 ± 0.05	1.09 ± 0.02	2.56 ± 0.05	1.09 ± 0.02	2.54 ± 0.05
Age 65+	1.48 ± 0.02	2.24 ± 0.04	1.78 ± 0.03	1.71 ± 0.04	1.78 ± 0.03	1.72 ± 0.04	1.78 ± 0.03	1.72 ± 0.04
Female	1.12 ± 0.01	1.95 ± 0.01	1.11 ± 0.01	2.00 ± 0.02	1.11 ± 0.01	2.00 ± 0.02	1.11 ± 0.01	2.10 ± 0.02
Never married			1.79 ± 0.02	0.49 ± 0.01	1.79 ± 0.02	0.49 ± 0.01	1.79 ± 0.02	0.49 ± 0.01
Divorced			1.27 ± 0.04	0.73 ± 0.02	1.27 ± 0.04	0.73 ± 0.02	1.27 ± 0.04	0.72 ± 0.02
Widowed			1.24 ± 0.02	0.84 ± 0.01	1.25 ± 0.02	0.84 ± 0.01	1.24 ± 0.02	0.84 ± 0.01
Middle school			0.91 ± 0.01	1.12 ± 0.01	0.91 ± 0.01	1.13 ± 0.01	0.91 ± 0.01	1.13 ± 0.01
High school			0.79 ± 0.01	1.16 ± 0.01	0.78 ± 0.01	1.18 ± 0.01	0.79 ± 0.01	1.16 ± 0.01
College			0.72 ± 0.02	1.23 ± 0.02	0.71 ± 0.02	1.27 ± 0.02	0.72 ± 0.02	1.21 ± 0.02
Unemployed			1.10 ± 0.01	0.99 ± 0.02*	1.10 ± 0.01	0.99 ± 0.02*	1.09 ± 0.01	0.99 ± 0.02*
Less physical activity			1.20 ± 0.01	1.10 ± 0.01	1.19 ± 0.01	1.11 ± 0.01	1.19 ± 0.01	1.10 ± 0.01
Household size			0.98 ± 0.00	1.03 ± 0.00	0.98 ± 0.00	1.03 ± 0.00	0.98 ± 0.00	1.03 ± 0.00
Urban			1.01 ± 0.01*	1.35 ± 0.01	1.00 ± 0.01*	1.36 ± 0.01	1.01 ± 0.01*	1.35 ± 0.01
Log(PCE)			0.75 ± 0.01	1.61 ± 0.02			0.78 ± 0.01	1.98 ± 0.03
2nd PCE quintile					0.92 ± 0.01	1.18 ± 0.01		
3rd PCE quintile					0.89 ± 0.01	1.34 ± 0.02		
4th PCE quintile					0.80 ± 0.01	1.51 ± 0.02		
5th PCE quintile					0.72 ± 0.01	1.81 ± 0.02		
District characteristics:								
Median PCE			0.87 ± 0.18*	1.57 ± 0.33	0.34 ± 0.07	7.34 ± 1.53	0.87 ± 0.18*	1.59 ± 0.33
Deprivation			0.91 ± 0.02	0.97 ± 0.02*	0.92 ± 0.02	0.97 ± 0.02*	0.91 ± 0.02	0.97 ± 0.02*
Inequality			1.08 ± 0.04	1.09 ± 0.04	1.08 ± 0.04	1.12 ± 0.04	1.11 ± 0.04	1.09 ± 0.04
Interaction terms:								
Female × Log(PCE)							0.93 ± 0.02	0.71 ± 0.01
Female × Inequality							0.94 ± 0.02*	0.99 ± 0.02*
Between-district variance	0.12	0.20	0.10	0.11	0.10	0.12	0.10	0.11
Correlation between RE	-0.19		-0.19		-0.18		-0.19	
N	645,027		578,512		578,512		578,512	
AIC	1,018,045		891,342		891,623		890,755	
BIC	1,018,238		891,849		892,198		891,307	

Note: * p > 0.10.

81

Figure 4.2: Spatial distribution of malnutrition across 440 districts in Indonesia
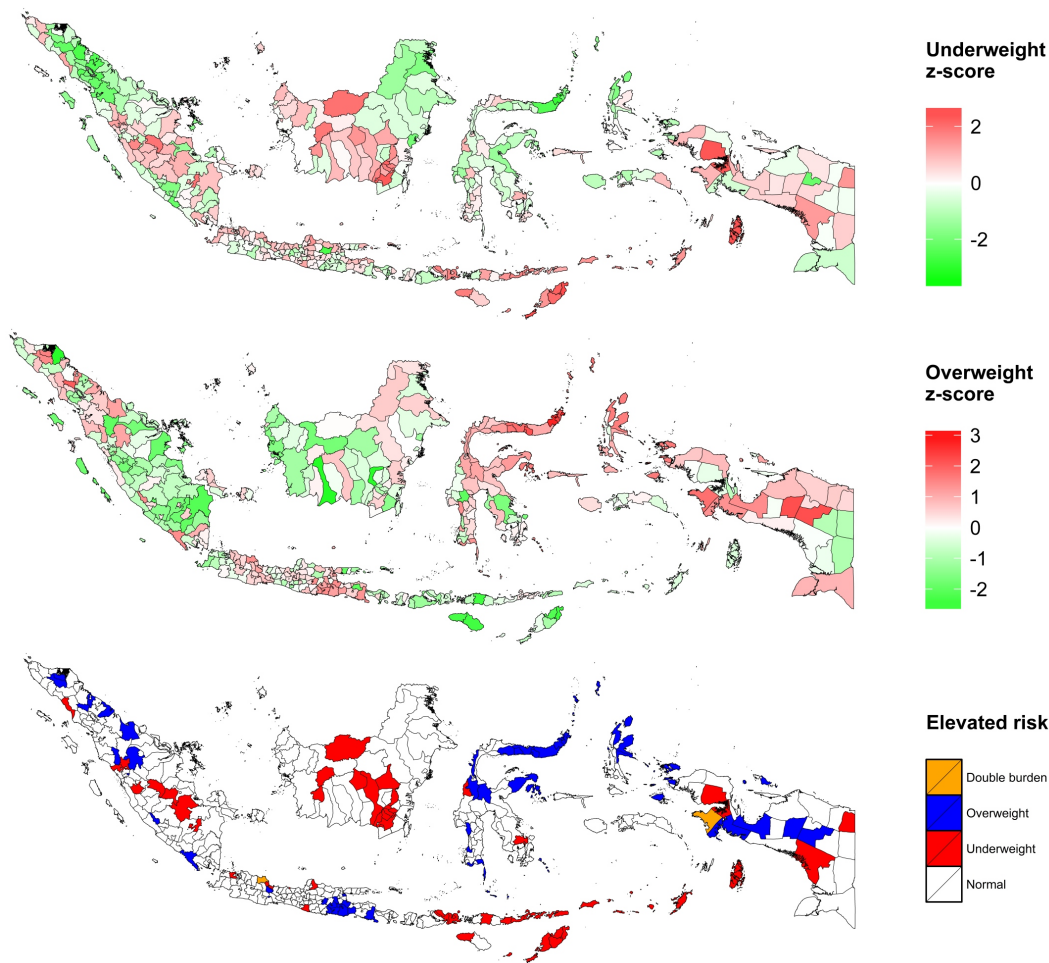
high income; yet these factors are also generally associated with greater odds of being overweight. The monotonicity of income effect is clearly demonstrated in Full Model 2, although a curvilinear parametrisation as introduced in Full Model 1 appears to be more parsimonious. This implies that, as of 2007, undernutrition in Indonesia remains a 'disease of poverty', while overnutrition is one of affluence. Having enough physical activity and living in an egalitarian area seems to protect Indonesians from both extremes of malnutrition, but area-level economic development only appears to aggravate the overnutrition problem and does not seem to aid in alleviating undernutrition even after controlling for facility deprivation and urban/rural residential location.

Finally, the Interaction Model tests whether women's nutritional vulnerability is modified by income level or income inequality. We found some evidence indicating that this is indeed the case. The model shows that as individual income increases, the nutritional gap between men and women narrows in both the underweight and overweight

equations. The gap also diminishes as the level of income inequality increases in the underweight equation, but a similar effect is imprecisely estimated in the overweight equation. In essence, this tells us that the effect of income is more pronounced among women than men and that adults of both sexes are equally deprived when they live in less egalitarian environments. In all models, urban areas are consistently obesogenic while, rather paradoxically, facility deprivation remains negatively associated with undernutrition. Ultimately, in order to ascertain whether these relationships are robust across disaggregations by sex and urban/rural location, we perform stratified analyses. Table 4.4 shows that these findings are indeed consistent.

Having investigated the determinants of nutritional status, we now attempt to understand the geographical distribution of the risk of malnutrition within the Indonesian archipelago by means of extracting the standardised random effects (empirical Bayes modes instead of means, for computational feasibility) for each contrast (Ackerson et al., 2008) in the best fitting model (Interaction Model) and plotting them in the top and middle panels of Figure 4.2. In this mapping exercise we manually impute the estimated random effects for two districts (Puncak Jaya and Pegunungan Bintang) with the value of their nearest neighbours (Jayawijaya and Yahukimo) because of a lack of individual income data in these districts. It is then evident from the maps that the risks of under- and overnutrition are indeed spatially segregated across the islands in Indonesia. Clusters of areas with high undernutrition vulnerability are observable in South Sumatra, Central and South Kalimantan (Borneo), Java (north coast), and Nusa Tenggara (Lesser Sunda) islands; areas particularly vulnerable to overnutrition appear in North Sumatra, West and East Java, North and Central Sulawesi (Celebes), Halmahera, and Papua. Further, in the bottom panel we identify areas with elevated risk of dual malnutrition (Z-score > 1). Only two out of 440 districts are categorised as double burden districts (Indramayu in West Java and Fak-Fak in West Papua); the number of districts classified as underweight and overweight is 54 and 66, respectively. Finally, Table 4.5 presents the top 10 most nutritionally vulnerable districts. It is apparent at this point that if evidence- or need-based interventions are to be prescribed, then the islands of Nusa Tenggara (containing four of the 10 districts most vulnerable to undernutrition) and Sulawesi (containing eight of the 10 districts most vulnerable to overnutrition) must be the primary targets.

Table 4.4: Adjusted odds ratio obtained from stratified models

Predictors	Underweight				Overweight			
	Male	Female	Urban	Rural	Male	Female	Urban	Rural
Intercept	0.23 ± 0.01	0.24 ± 0.01	0.27 ± 0.01	0.19 ± 0.01	0.04 ± 0.00	0.13 ± 0.00	0.08 ± 0.00	0.05 ± 0.00
Individual characteristics:								
Age 25–34	0.53 ± 0.01	0.61 ± 0.01	0.56 ± 0.01	0.57 ± 0.01	1.87 ± 0.05	1.92 ± 0.04	2.02 ± 0.08	1.99 ± 0.04
Age 35–44	0.48 ± 0.01	0.54 ± 0.01	0.42 ± 0.01	0.56 ± 0.01	2.78 ± 0.08	2.84 ± 0.05	3.27 ± 0.14	2.85 ± 0.06
Age 45–54	0.58 ± 0.02	0.75 ± 0.02	0.45 ± 0.02	0.78 ± 0.02	3.00 ± 0.09	2.90 ± 0.06	3.73 ± 0.17	2.78 ± 0.06
Age 55–64	0.94 ± 0.03	1.22 ± 0.03	0.70 ± 0.03	1.30 ± 0.03	2.59 ± 0.08	2.23 ± 0.05	3.12 ± 0.15	2.18 ± 0.06
Age 65+	1.65 ± 0.05	1.88 ± 0.05	1.27 ± 0.05	2.04 ± 0.05	1.78 ± 0.07	1.48 ± 0.04	2.07 ± 0.11	1.49 ± 0.05
Female			1.02 ± 0.02*	1.15 ± 0.01			1.77 ± 0.02	2.23 ± 0.02
Never married	1.62 ± 0.03	1.96 ± 0.03	1.69 ± 0.04	1.78 ± 0.03	0.60 ± 0.01	0.39 ± 0.01	0.50 ± 0.01	0.51 ± 0.01
Divorced	1.30 ± 0.07	1.28 ± 0.04	1.25 ± 0.07	1.29 ± 0.05	0.76 ± 0.05	0.70 ± 0.02	0.74 ± 0.03	0.71 ± 0.03
Widowed	1.34 ± 0.05	1.26 ± 0.03	1.18 ± 0.04	1.28 ± 0.03	0.84 ± 0.04	0.81 ± 0.02	0.86 ± 0.02	0.80 ± 0.02
Middle school	0.91 ± 0.01	0.90 ± 0.01	0.88 ± 0.02	0.93 ± 0.01	1.26 ± 0.02	1.06 ± 0.01	1.08 ± 0.02	1.13 ± 0.02
High school	0.70 ± 0.01	0.88 ± 0.02	0.74 ± 0.01	0.79 ± 0.01	1.57 ± 0.02	0.97 ± 0.01	1.09 ± 0.02	1.27 ± 0.02
College	0.56 ± 0.02	0.86 ± 0.03	0.66 ± 0.02	0.76 ± 0.03	1.89 ± 0.04	0.90 ± 0.02	1.12 ± 0.02	1.53 ± 0.04
Unemployed	1.15 ± 0.02	1.05 ± 0.02	1.14 ± 0.02	1.10 ± 0.02	1.00 ± 0.03*	1.01 ± 0.03*	0.95 ± 0.02	0.98 ± 0.02*
Less physical activity	1.31 ± 0.02	1.10 ± 0.01	1.16 ± 0.01	1.21 ± 0.01	1.22 ± 0.02	1.02 ± 0.01	1.08 ± 0.01	1.11 ± 0.01
Household size	0.98 ± 0.00	0.97 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	1.03 ± 0.00	1.02 ± 0.00	1.02 ± 0.00	1.04 ± 0.00
Urban	1.06 ± 0.02	0.95 ± 0.01			1.36 ± 0.02	1.35 ± 0.02		
2nd PCE quintile	0.93 ± 0.02	0.92 ± 0.01	0.92 ± 0.01	0.92 ± 0.01	1.20 ± 0.02	1.16 ± 0.02	1.17 ± 0.03	1.19 ± 0.02
3rd PCE quintile	0.89 ± 0.02	0.88 ± 0.01	0.88 ± 0.02	0.88 ± 0.01	1.39 ± 0.03	1.30 ± 0.02	1.32 ± 0.03	1.36 ± 0.02
4th PCE quintile	0.83 ± 0.02	0.78 ± 0.01	0.80 ± 0.02	0.80 ± 0.01	1.58 ± 0.03	1.44 ± 0.02	1.43 ± 0.03	1.56 ± 0.03
5th PCE quintile	0.76 ± 0.02	0.69 ± 0.01	0.71 ± 0.02	0.73 ± 0.01	2.00 ± 0.04	1.64 ± 0.03	1.65 ± 0.04	1.93 ± 0.03
District characteristics:								
Median PCE	0.38 ± 0.09	0.29 ± 0.06	0.24 ± 0.06	0.33 ± 0.09	8.09 ± 1.81	6.41 ± 1.43	4.48 ± 1.06	12.31 ± 3.44
Deprivation	0.88 ± 0.02	0.96 ± 0.02	0.86 ± 0.03	0.93 ± 0.02	1.03 ± 0.02*	0.94 ± 0.02	1.03 ± 0.03*	0.95 ± 0.02
Inequality	1.07 ± 0.04*	1.08 ± 0.04	1.11 ± 0.05	1.05 ± 0.05*	1.13 ± 0.05	1.11 ± 0.05	1.13 ± 0.05	1.13 ± 0.05
Between-district variance	0.11	0.11	0.10	0.12	0.12	0.13	0.11	0.14
Correlation between RE	−0.32	−0.18	−0.05	−0.24	−0.32	−0.18	−0.05	−0.24
N	283,218	304,643	213,942	364,570	283,218	304,643	213,942	364,570

Note: * $p > 0.10$.

Table 4.5: Top 10 most nutritionally vulnerable districts

	Undernutrition			Overnutrition	
Rank	District	Island	Rank	District	Island
1	Belu	Nusa Tenggara	1	Kota Tomohon	Sulawesi
2	Rote Ndao	Nusa Tenggara	2	Kota Bitung	Sulawesi
3	Kepulauan Aru	Papua	3	Minahasa Selatan	Sulawesi
4	Teluk Bintuni	Papua	4	Minahasa	Sulawesi
5	Banjar	Kalimantan	5	Jayawijaya	Papua
6	Timor Tengah Utara	Nusa Tenggara	6	Bone Bolango	Sulawesi
7	Hulu Sungai Utara	Kalimantan	7	Kota Manado	Sulawesi
8	Timor Tengah Selatan	Nusa Tenggara	8	Minahasa Utara	Sulawesi
9	Kapuas Hulu	Kalimantan	9	Karo	Sumatra
10	Tebo	Sumatra	10	Kota Gorontalo	Sulawesi

4.3.3 Quantile regression analysis

The previous modelling exercises have implicitly assumed that the relationship between nutritional status and its predictors is homogeneous along the continuum of BMI. In this section, we relax this assumption by allowing each predictor to have an impact on both the location and the scale of conditional BMI distribution. The result of fitting a quantile regression model with the Full Model 1 specification is presented in Figure 4.3. In the figure, the X-axis represents the conditional quantile of BMI, while the Y-axis indicates the estimated regression coefficient; a bold black line shows the independent effect of each explanatory variable on the respective conditional quantile with its associated 95% point-wise confidence interval shown in grey shade; the three solid black circles represent the conditionally underweight (the 0.1th quantile), normal (the 0.5th quantile) and overweight (the 0.9th quantile). The goal of this modelling exercise is to find out for whom the effect of each covariate is particularly relevant. A flat line means that the effect is equal for all individuals, irrespective of their nutritional status. A monotonically increasing or decreasing line indicates that the effect becomes gradually more pronounced in one extreme of nutritional status. A U-shaped line suggests that the effect is different between individuals with BMIs in the normal range and those at both extremes of the nutritional spectrum. Finally, any line crossing the zero Y-axis shows that there is a divergence in the direction (a positive-to-negative reversal, or vice versa) of an effect.

As shown in Figure 4.3, being married, having a high education level, being employed and having one additional household member are associated with a constant positive increase of BMI. In contrast, the effects of income, age and urban environment on BMI are monotonically positive with magnitudes that become increasingly stronger

Figure 4.3: Quantile regression estimates (BMI quantiles in X-axis; β in Y-axis)
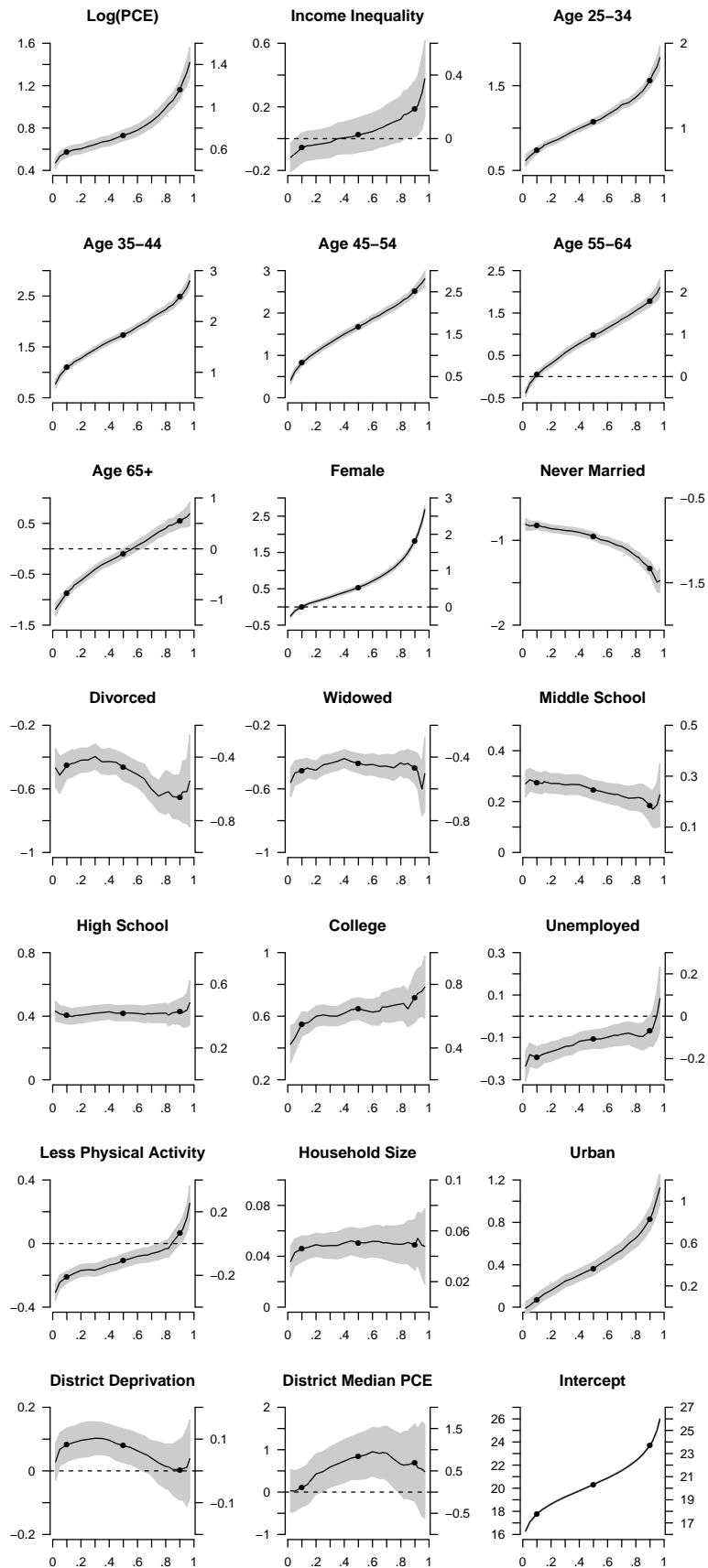
as one moves from the underweight to the overweight sub-population. An exception, though, is the oldest age group (65 years old and older). Among the underweight, later life is associated with a lower BMI, while among the overweight, it is associated with a higher BMI; this is, however, of little consequence for normal individuals. A roughly similar pattern is observable for sex, physical activity and income inequality. This means that being female, having inadequate physical activity, being in the oldest age group and living in a less egalitarian area are especially detrimental for the under- and overweight sub-populations. U-shaped relationships are observable for the effects of deprivation and area-level economic development. This suggests that a positive change in these variables is associated with a higher BMI; it is, however, only statistically significant among individuals with BMIs in the normal range. The relatively straight intercept estimates show that BMI is approximately normally distributed, which indeed confirms the result of our earlier descriptive analysis (Section 4.3.1). Overall, regardless of the differences in modelling assumptions, the picture obtained from the quantile regression model largely mirrors that of the multilevel multinomial logistic models.

4.4   Discussion and conclusion

Analysing a nationally representative dataset, this paper investigates the social determinants as well as the geographical variations of the double burden of malnutrition in 440 districts in Indonesia. The main objectives of this research are to study (1) how individuals' socio-economic positions relate to nutritional status, (2) how contextual factors at the district level influence individuals' nutritional status, and (3) how the risks of under- and overnutrition are distributed around the Indonesian archipelago after adjusting for the effects of observable socio-demographic determinants.

We found that, in 2007, the prevalence of under- and overweight was 14.4% and 17.9%, respectively. These figures indicate that one in three Indonesian adults faces a potential nutritional problem and that the double burden of malnutrition is shared roughly equally by both under- and overnutrition problems. We found that education, employment, and income protect Indonesians from undernutrition but that they also increase the probability of being overweight. Individual income as measured using per capita household expenditure seems to exhibit a monotonically decreasing and increasing effect on the likelihood of being under- and overweight, respectively. This suggests that undernutrition in Indonesia remains a disease of the poor while overnutrition is one of the affluent, a finding consistent with the general trend observed in

other low and lower-middle income countries but not among upper-middle and high income countries (Jolliffe, 2011; Popkin, 2001; Subramanian et al., 2009).

The risk of under- and overnutrition seems to be spatially clustered within the islands of Indonesia. Clusters of districts with high undernutrition vulnerability are located in South Sumatra, Central and South Kalimantan, Java (north coast), and Nusa Tenggara islands; susceptibility to overnutrition is observed particularly in North Sumatra, West and East Java, North and Central Sulawesi, Halmahera, and Papua. We found little evidence to suggest that the double burden of malnutrition exists within the same districts in Indonesia. Areas with high risk of undernutrition tend to be the ones with low risk of overnutrition; in fact, only Indramayu district in West Java and Fak-Fak district in West Papua are identified as double burden districts. To some extent, this is perhaps a relief from the point of view of policy-makers, for whom the burden of under- and overnutrition coexisting within the same districts might have presented a somewhat difficult situation. As previous research has already pointed out, though, despite appearing to be a transitory phenomenon, the double burden of malnutrition does indeed appear in a significant portion of individual Indonesian households (Doak et al., 2005; Oddo et al., 2012; Roemling and Qaim, 2013; Vaezghasemi et al., 2014).

While finding little evidence for the presence of double burden districts, we have identified the existence of 'doubly vulnerable' population sub-groups. Our analysis shows that the elderly, women, individuals engaging in insufficient physical activity, and individuals living in highly unequal districts are vulnerable to both under- and overnutrition problems. We suspect that, for the elderly, this is due to the changes in metabolic function and lifestyle as well as the psychological challenges associated with ageing (Hickson, 2006). For women, the double vulnerability seems to be consistent with explanations provided by the biological, social and cultural aspects of malnutrition (Brown and Konner, 1987; Delisle, 2008). On the one hand, some suggest that women's propensity to obesity is driven by the difficulty of maintaining a healthy weight after the high nutritional requirements of childbearing (pregnancy and lactation) subside; in some parts of the developing world, the tendency to obesity is further shaped by the ideal body image maintained by society (fatness as a symbol of maternity, nurturance and affluence). On the other hand, researchers also document that women in some poor societies are often subjected to gender discrimination in intra-household food allocation, hence posing a greater risk of undernutrition (Frongillo and Bégin, 1993; Molini and Nubé, 2007; Thomas, 1990).

Regarding the adverse effect of income inequality on nutritional status, Subramanian

et al. (2007) suggest in their study of Indian society that income inequality can be a marker of both resource maldistribution and inefficient public policy. It is likely that unequal areas are the places where the privileged over-consume while the underprivileged face food insecurity. Equally likely is that, due to the low social cohesion as well as other negative externalities associated with a highly skewed income distribution, public policy in a less egalitarian society is prone to manipulation by vested interests, resulting in poor provision of the amenities that are vital for combating malnutrition.

In our research, we also found a paradoxical protective effect of facility deprivation on undernutrition. We initially suspected this to be an artefact of confounding, but it remains unresolved even after fitting multivariate models. While puzzling, this is not an isolated observation (Thomas and Strauss, 1992; Wolff and Maliki, 2008). Perhaps this is attributable to the endogenous, non-random spatial distribution of government programs as a result of the historical priority on placing health facilities and interventions in less healthy areas (Pitt et al., 1995). Unfortunately, this puzzle cannot simply be addressed using the cross-sectional data we have at hand; it may therefore be pursued further in future research.

Other limitations of this study must now be acknowledged. The cross-sectional data that we have do not permit us to incorporate the temporal dimension into our analysis. As a consequence, this study only provides a snapshot capturing the determinants and geographical variations of the double burden of malnutrition in Indonesia in the year 2007. It is known that the burden of obesity gradually shifts to the poor as a nation progresses economically (Brown and Konner, 1987; Popkin, 1998). Whether such a shift has begun to occur in Indonesia is indeed an interesting subject to study, but carrying out the relevant research obviously necessitates the availability of newer data. Another limitation is that the statistical models fitted in this study did not explicitly account for spatial-contextual autocorrelation which may, to some extent, affect the precision as well as the smoothness of the estimated risks. The importance of undertaking such an endeavour cannot be underestimated, but it clearly deserves its own avenue in the vast literature of spatial epidemiology.

Despite these limitations, this study does, however, contribute to the literature in several ways. This study is among the few to consider the double burden of malnutrition in Indonesia from the perspective of the general population. As noted earlier, all existing studies have focused rather specifically on Indonesian women (Winkvist et al., 2000) or households (Doak et al., 2005; Oddo et al., 2012; Roemling and Qaim, 2013; Vaezghasemi et al., 2014). This study also adds to the literature by showing that the influence of contextual macro-economic conditions (income inequality and level

of economic development) is not negligible with regard to the nutritional well-being of individuals (Block et al., 2004). In addition, this study provides the literature with a principled characterisation of the spatial distribution of nutritional vulnerability within the 17,000-island Indonesian archipelago which, we believe, is indispensable for the purpose of policy targeting. Of course, in the absence of good data, this study would not have been able to offer the present analysis.

If any policy implications for dealing with the double burden of malnutrition are to be suggested from the findings of this study, then they should include the following points. Raising the overall level of the socio-economic status of the population through education, employment, and income-enhancing opportunities can help to improve purchasing power, which, in turn, enables individuals to afford enough food to fulfil their needs. That alone, however, is not sufficient; we have already seen that the risk of overnutrition also increases with every improvement in socio-economic conditions. Therefore, there is a need for a wider public educational campaign that promotes behavioural changes especially in, but not limited to, the spheres of physical activity, dietary pattern and gender equality (Roemling and Qaim, 2012). Furthermore, the need for better nutritional education (Webb and Block, 2004) in academic curricula cannot be overstated as it has become apparent that, at least in our models, more schooling is not always correlated with better nutritional status. Better nutritional education, of course, will not only facilitate behaviour change but also help to shape a healthier body image in society. Simultaneously, as it has been projected that by the year 2030 more people in developing countries will live in cities than in rural areas (Cohen, 2006), the obesogenic urban environment must also be addressed. A recent assessment of Indonesia's built environment indicates an environment 'that is fairly unfriendly to pedestrian physical activity with limited access to healthy foods' (Shrimpton and Rokx, 2013: 3). This hints that improvement in nutritional health can also be achieved through the provision of a healthier urban planning initiative.

Furthermore, as much as nutritional well-being is determined by genetic predisposition and individual behaviour, it is also a matter of social justice. While the effects of inequality may appear relatively minor, they affect millions of Indonesians. An economic policy that promotes equity and quality of development as opposed to one that emphasises growth *per se* is much desired. This entails the aim not only to narrow the gap between the haves and the have-nots within a region, but also to distribute the fruit of development fairly between regions. As shown in the nutritional vulnerability map (Figure 4.2), it is no coincidence that places with high risk of undernutrition tend to be the ones that are difficult to access and that have an inefficient distribution

system and low market penetration. Indonesians living in these remote areas, no matter how much spending power they have, still find it difficult to achieve diversified, nutritionally balanced diets relative to those living in other parts of the archipelago. Perhaps it is not too late to remind ourselves that an efficiently functioning market and distribution system constitutes a necessary condition for a nation's nutritional well-being. Lastly, it is also worth noting that, if any interventions are to be initiated, then islands in east Indonesia should now clearly be the top priority for policy-makers.

❀ ❀ ❀

Chapter 5

Geography and social distribution of malaria in Indonesian Papua:
A cross-sectional study

Abstract: Despite being one of the world's most affected regions, only little is known about the social and spatial distributions of malaria in Indonesian Papua. Existing studies tend to be descriptive in nature; their inferences are prone to confounding and selection biases. At the same time, there remains limited malaria-cartographic activity in the region. Analysing a subset (N = 22,643) of the National Basic Health Research 2007 dataset (N = 987,205), this paper aims to quantify the district-specific risk of malaria in Papua and to understand how socio-demographic/economic factors measured at individual and district levels are associated with individual's probability of contracting the disease. We adopt a Bayesian hierarchical logistic regression model that accommodates not only the nesting of individuals within the island's 27 administrative units but also the spatial autocorrelation among these locations. Both individual and contextual characteristics are included as predictors in the model; a normal conditional autoregressive prior and an exchangeable one are assigned to the random effects. Robustness is then assessed through sensitivity analyses using alternative hyperpriors. We find that rural Papuans as well as those who live in poor, densely forested, lowland districts are at a higher risk of infection than their counterparts. We also find age and gender differentials in malaria prevalence, if only to a small degree. Nine districts are estimated to have higher-than-expected malaria risks; the extent of spatial variation on the island remains notable even after accounting for socio-demographic/economic risk factors. Although we show that malaria is geography-dependent in Indonesian Papua, it is also a disease of poverty. This means that malaria eradication requires not only biological (proximal) interventions but also social (distal) ones.

Keywords: malaria, map, Papua, Indonesia, Bayesian, spatial, multilevel

5.1 Introduction

Malaria, a mosquito-borne infectious disease that inflicts devastating health (Guyatt and Snow, 2001; Holding and Snow, 2001) and economic (Gallup and Sachs, 2001; Sachs and Malaney, 2002; Teklehaimanot and Meija, 2008) costs on society, remains a major problem in Indonesian Papua (Webster, 2001). This region of mixed-parasite endemicity is located in the easternmost part of the Indonesian archipelago (Figure 5.2) and is classified by the World Health Organization (WHO) as hyper-endemic area with annual parasite incidence (API) greater than 10% (nationwide API <1%; WHO, 2012b) and parasite prevalence (PP) as high as 50–75% (nationwide PP <1%; WHO, 2014). Malaria accounts for a considerable proportion (15–34%) of total hospital workload in the region (Karyana et al., 2008); mortality due to severe anaemia (Douglas et al., 2014) as well as multi-drug resistance with high rate of therapeutic failure (65–95%) have been documented (Sumawinata et al., 2003; Tjitra et al., 2008). In 2007, the Ministry of Health of the Republic of Indonesia (Kemenkes, 2008) estimated that the infectious disease was prevalent among one-fifth (22.25%) of the Papuan population—a figure that is seven times higher than the national average (Figure 5.1). Perhaps nothing can highlight the seriousness of this situation better than the fact that while malaria prevalence for the whole Indonesian archipelago decreased from 2.9% in 2007 (Kemenkes, 2008) to 1.9% in 2013 (Kemenkes, 2013), the figure for Papua actually increased to 24% over the same period.

Defeating malaria is certainly a high priority for Indonesian policy-makers; they have not only set the year 2030 as the deadline for malaria elimination in the country (Kemenkes, 2009) but have also entrusted local Papuan administrators with the responsibility for preventing and combating endemic diseases through the enactment of the 2001 Papua Special Autonomy Law No. 21 (LNRI No. 135, 2001). Notwithstanding these political commitments, challenges to disease control in Papua remain. Principal among them is that the spatial distribution of malaria, which is vital for guiding efficient and equitable allocation of the limited resources available for intervention, is still understudied. To date, the only risk map available for the region is the one produced by the Malaria Atlas Project (MAP), which, while informative, was unfortunately based on community blood surveys carried out in non-randomly selected locations (Elyazar et al., 2011a, 2012). Moreover, because the risk estimate in the existing malaria maps is presented as a continuous surface obtained from geostatistical models that are blind to political boundaries, there is no straightforward way to obtain a single summary (Wall, 2004) for each local administrative unit in Papua. Policy-makers in now-decentralised Indonesia (Hill, 2014) are therefore deprived of an intuitive

Figure 5.1: Malaria prevalence in 33 Indonesian provinces in 2007 (%), sorted by island group's longitude (left to right = west to east, low to high prevalence; source: Kemenkes, 2008)
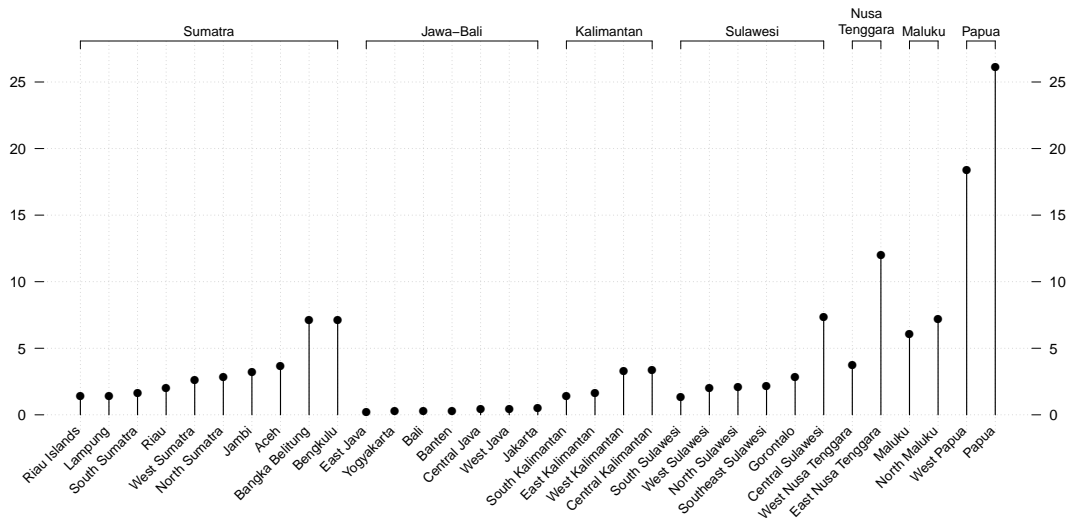
tool for prioritising development projects or other forms of intervention that are funded by transfers from central to local governments (the *Kabupaten/Kota* or the district/municipality).

This scarcity of malaria-cartographic activity is further complicated by the fact that, unlike in Africa, the social and environmental determinants of malaria in Papua have not yet been thoroughly examined. Existing knowledge—that the risk of contracting the disease seems to be higher among non-native Papuans (Barcus et al., 2007; Tjitra et al., 2008), children and young adults (Douglas et al., 2014), as well as rural (Barcus et al., 2007) and lowland dwellers (Douglas et al., 2014)—was in fact elicited from simple descriptive or bivariate analyses performed on small community or facility samples that are prone to both confounding and selection biases. So, although Papua is reputed to be one of the most malaria-ridden regions in the world (CDC, 2010), to date, only little is known about the social and spatial aspects of the disease. Without precise knowledge of where in Papua malaria strikes and which population subgroup it hits the hardest, it is likely to be difficult for Indonesian policy-makers to meet the 2030 elimination target on time.

Analysing large population data (N = 22,643) from the National Basic Health Research 2007 (*Riset Kesehatan Dasar*; Kemenkes, 2008), this study aims to address these gaps. Through the application of a Bayesian hierarchical modelling technique that accounts for both the nesting of individuals within districts (vertical dependence) and the spatial autocorrelation among these areas (horizontal dependence; see Figure 5.3),

Figure 5.2: Setting of the study

District in Papua
 1 Fak-Fak
 2 Kaimana
 3 Manokwari
 4 Raja Ampat
 5 Sorong Selatan
 6 Sorong
 7 Teluk Bintuni
 8 Teluk Wondama
 9 Asmat
10 Biak Numfor
11 Boven Digoel
12 Jayapura
13 Jayawijaya
14 Keerom   21 Puncak Jaya
15 Mappi   22 Sarmi
16 Merauke   23 Supiori
17 Mimika   24 Tolikara
18 Nabire   25 Waropen
19 Paniai   26 Yahukimo
20 Pegunungan Bintang 27 Yapen Waropen

**Island Group
in Indonesia**
 A Sumatra
 B Jawa-Bali
 C Kalimantan
 D Sulawesi
 E Nusa Tenggara
 F Maluku
 G Papua

Figure 5.3: Illustration of hierarchical and spatial dependence
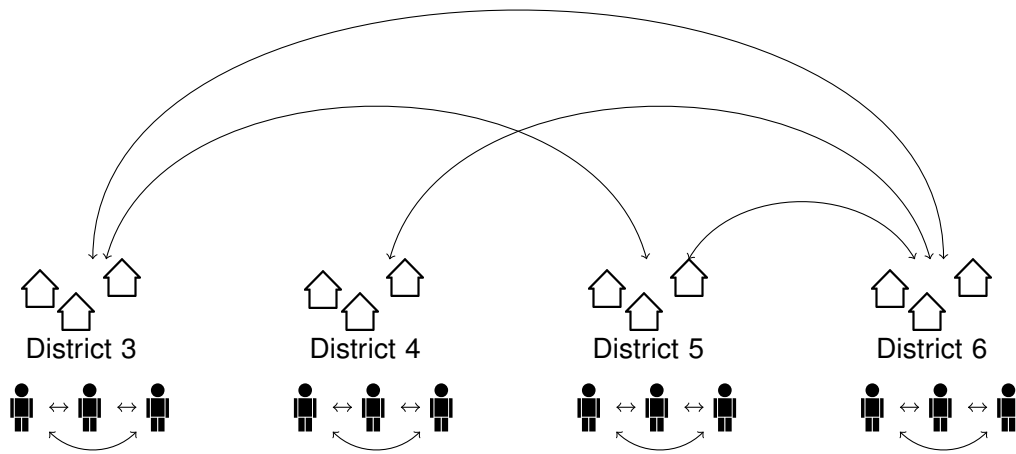
this paper seeks (1) to quantify the district-specific risk of malaria in Papua and (2) to understand how socio-demographic/economic factors measured at individual and district levels are associated with an individual's probability of contracting the disease. The novelty of this paper is threefold. First, in using randomly sampled population data from Indonesia's largest public health study, this paper avoids the problem of confounding and selection biases that beset earlier studies mentioned above. Second, through its spatial analysis of irregular lattice data, this study is able to deliver a single risk summary for each district and municipality in Papua, which is the lowest autonomous administrative unit in the Indonesian political system. Finally, the present study is also distinguished from others in its multilevel analysis of individual and contextual determinants of malaria, avoiding ecological fallacy (Robinson, 1950; Freedman, 2001; Greenland, 2001; Snijders and Bosker, 2012).

The remainder of this paper is structured as follows. The next section describes the study site, data, measures and modelling techniques. Section 5.3 presents the results. Section 5.4 concludes.

5.2 Methods

5.2.1 Study site

This study was carried out in the western half, or the Indonesian side, of the New Guinea island, commonly referred to as the *Papua* or *Irian Jaya* region among Indonesians (Figure 5.2). Lying between latitudes 0–9° South and longitudes 124–141° East,

Figure 5.4: An Indonesian version of the Alkire-Foster multidimensional poverty index (Alkire and Foster, 2011a) for the year 2013 (source: Hanandita and Tampubolon, 2016b)
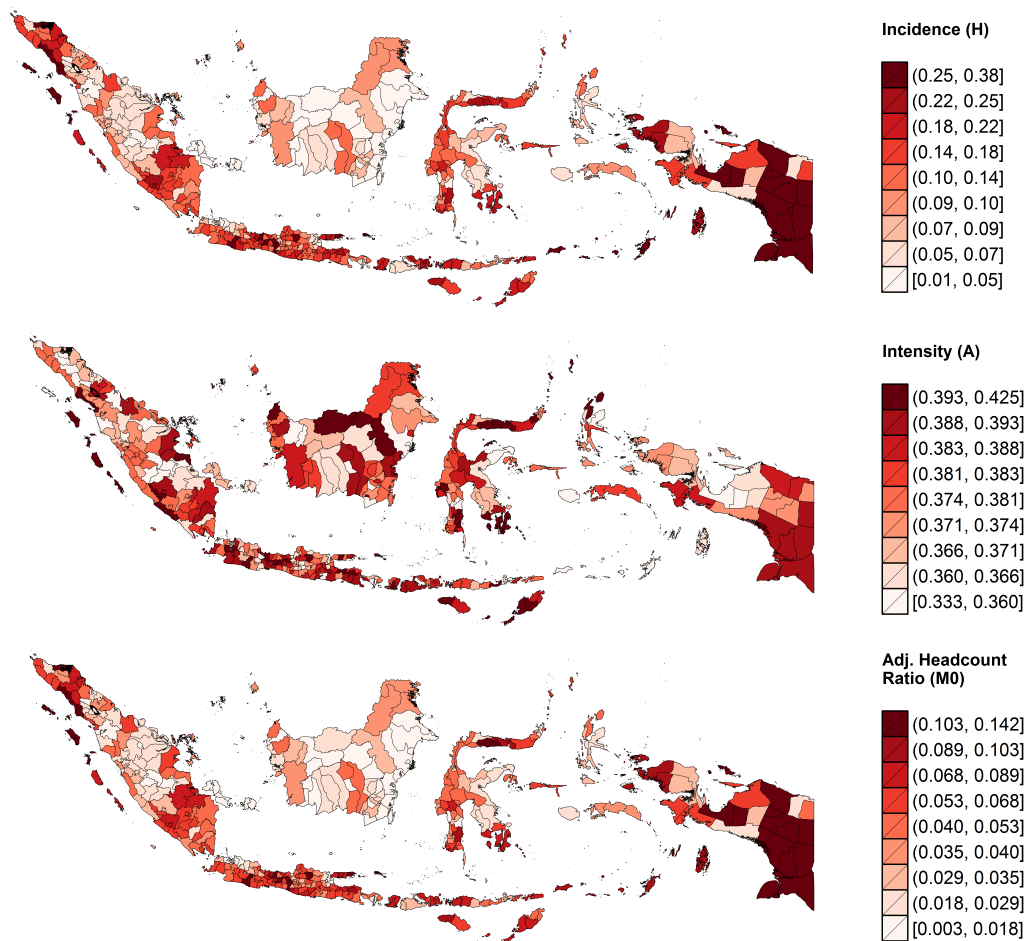
the climate of Papua is entirely tropical, with a dry season typically occurring from April–October and a wet season from October–April. Most of Papua's land area is covered by forests. Apart from a mountain range stretching more than 1,500 kilometres from the west to central east of the island, the topography of Papua is shaped by the extensive presence of swamps, wetlands, mangroves, savannah grasslands, lakes and rivers. Rain persists throughout the year (150–270 days of rain per year), yielding 2,000–3,000 mm of annual rainfall (BPS Papua, 2015; BPS Papua Barat, 2015). The average humidity is 80–90% while the average temperature is about 26° Celsius, with an average maximum of 30° and an average minimum of 22° (BPS Papua, 2015; BPS Papua Barat, 2015).

According to the latest census conducted in 2010 (BPS, 2010), the population of Papua is 3.6 million (2% of Indonesia's population) living in an area of 420,540 km^2 (22% of the country's land area), with a population density of just 9 persons per

square kilometre (the lowest in Indonesia). As many as 70–75% of Papuans live in rural areas (BPS, 2010). Despite hosting one of the planet's largest gold mining operations (the Grassberg mine in Mimika district), Papuan society is plagued by poverty and under-development. As shown in Figure 5.4, Hanandita and Tampubolon (2016b) estimate that approximately a quarter of Papuan adults aged 18 and older were multidimensionally poor in 2013; collectively, they were subjected to about 10% of the total deprivation (in terms of income, illness episodes, morbidity, schooling and literacy) potentially experienced by all adult Indonesian that year. The combination of geographic features, climate conditions and extreme poverty provides a suitable environment for malaria transmission, both biologically and socially (Lowe et al., 2014; Manh et al., 2011; Sachs and Malaney, 2002).

5.2.2 Data

We analyse data drawn from the National Basic Health Research (*Riset Kesehatan Dasar*, Riskesdas) 2007. Involving 987,205 individuals from 258,366 households in 440 districts, Riskesdas is the largest public health study ever conducted by the Ministry of Health of the Republic of Indonesia (Kemenkes, 2008). For our analysis, we selected individuals of all ages living in Papua, yielding a total sample size of 22,643 individuals.

Information on each respondent's malaria status, age, sex, use of insecticide-treated net (ITN), and urban/rural residential location is available from the Riskesdas 2007 dataset. However, because the household consumption expenditure module was not administered to survey respondents living in a number of Papuan districts, we are unable to include a measure of individual income. Instead, we obtain a measure of wealth in the form of each district's median per capita household consumption expenditure (Deaton and Zaidi, 2002; Howe et al., 2012), computed from the National Socio-economic Survey (*Survei Sosial Ekonomi Nasional*, Susenas) 2008 dataset. We also obtain additional information on districts' median household elevation (as a proxy for temperature and precipitation; Weiss et al., 2015) and the proportion of districts' populations living in or near forest (as a proxy for forest density). This contextual information is derived from the Village Census (*Potensi Desa*, Podes) 2008 dataset that covers all 75,410 villages across the Indonesian archipelago.

Spatial polygons and the associated political boundary data are obtained from the freely-accessible GADM database of global administrative area (`www.gadm.org`). Originally, there were 29 districts and municipalities in Papua in 2007, but due to the

lack of spatial polygons for *Kota Sorong* and *Kota Jayapura* municipalities, we have no choice but to regroup study participants living in these locations with those living in *Kabupaten Sorong* and *Kabupaten Jayapura* districts, respectively. This will not come as a surprise to researchers analysing data from Indonesia. Parmanto et al. (2008) write at some length about both the poor quality of the country's spatial data and the government's slow process of updating administrative boundaries.

Although malaria status is fully observed, 871 of the 22,643 individuals selected as our study sample (3.8%) have missing values in other socio-demographic variables (to be described next) and are thus excluded from the subsequent multivariate modelling exercise. This data-cleaning procedure produces a final complete-case sample size of 21,772 individuals, corresponding to 96.2% of the original Papuan sample of the Riskesdas 2007 study. It is further assumed that missingness is non-informative since we found no significant difference between the included and the excluded individuals in terms of both malaria prevalence and other risk factors considered in this study. Informed consent was obtained prior to data collection; study participants' confidentiality was strictly protected by means of anonymisation (Kemenkes, 2008).

5.2.3   Measures and a priori expectations

The outcome variable, namely the individual's malaria status, is coded as a binary variable whose value equals one (malaria-positive) if, within the past month, the study participant had been diagnosed with laboratory-confirmed malaria, suffered from high fever accompanied by chills, sweating, or headache, or took anti-malarial drugs (Kemenkes, 2008). Age is treated as a 7-category ordinal variable indicating the respondent's age group (0–4, 5–14, 15–24, 25–34, 35–44, 45–54, and 55+). Sex, ITN use, and urban/rural residential location are each entered as a dummy variable representing female individuals, respondents who slept under an ITN the night prior to data collection, and those living in rural areas, respectively.

The three contextual variables are operationalised in the following way. Because no district has a median elevation between 200 and 1,200 metres, median household elevation is treated as a dummy variable indicating whether the majority of the district's population lives in lowland (≤ 200 metre above sea level) or highland ($\geq 1,200$ metre). The proportion of a district's population living in or near forest is multiplied by a factor of 10 and used as a continuous variable. For ease of interpretation as well as for capturing a possible non-linear relationship, district median income

is split into quintiles before being entered into the statistical model described next as a set of four dummy variables indicating the relative wealth of each district in Papua. With this set up, we then set the reference individuals (the intercept) in the model to represent urban, ITN non-user, male infants living in the poorest, least densely-forested, highland district.

A priori, we expect that the chance of contracting malaria will be relatively high among individuals living in rural areas and in poor, densely forested, lowland districts of Papua. This is because the extant literature has already hinted that:

- there is an inverse relationship between temperature (hence altitude and latitude) and the length of the *plasmodium* growth-cycle (Alegana et al., 2014; CDC, 2015; Manh et al., 2011; Sachs and Malaney, 2002);

- the micro-climate of forests enhances *anophelines* breeding sites and prolongs their survival as adults (Ernst et al., 2009; Stresman, 2010);

- the pollution and high population density of urban areas entail poor mosquito habitats and low biting frequency (Mmbando et al., 2011; Lowe et al., 2014); and that

- poverty creates conditions (poor housing, lack of health knowledge, negative health behaviours) that favour the spread of infectious diseases and restrict access to prevention and treatment (Haque et al., 2011; Ingstad et al., 2012; WHO, 2012a).

We also expect that the probability of being malaria-positive will be high among those who do not sleep under ITN due to the lack of a physical barrier separating them from the mosquitoes (Opeskin, 2009). Mendis et al. (2001) suggest that the 'male rather than female' as well as the 'working age rather than infant or elderly' infection patterns that are commonly found in South East Asian countries are unlikely to hold in high endemicity areas such as Papua. Studies from Peru (Guthmann et al., 2002), Bangladesh (Haque et al., 2010, 2011), Malawi (Chirombo et al., 2014), Gambia (Sonko et al., 2014), and India (Yadav et al., 2014) present conflicting evidence regarding the age and gender distributions of malaria.

5.2.4 Modelling techniques

To predict the malaria status of individual i living in district j, a Bayesian generalised linear model (GLM) with random effects is fitted (Gelman and Hill, 2007; Kruschke,

2010). We assume, for the data model, that a person's malaria status arises from the realisation of a Bernoulli trial with the probability of success (malaria-positive) π_{ij} as shown in equation 4.1. In the process model (equations 4.2 and 4.3), we take the logit of π_{ij} and model it as a linear combination of observed individual (x_{ij}) and contextual (x_j) characteristics with parameter vector β plus an unobserved district-specific effect ξ_j. The ξ_j can be intuitively understood as random intercepts indicating how much the risk of contracting malaria in each district varies from the island's average (β_0) after accounting for the effects of all observed covariates ($\sum_{p=1}^{P} \beta_p x_{pij}$). This district-specific effect is further decomposed additively into its spatially structured (u_j) and unstructured (v_j) components, which, in combination, are capable of incorporating the dependency structure of spatially correlated multilevel data (Figure 5.3) into the modelling process (Lawson, 2013).

$$y_{ij} \sim \text{Bernoulli}(\pi_{ij}) \tag{5.1}$$

$$\text{logit}(\pi_{ij}) = X_{ij}\beta + \xi_j \tag{5.2}$$

$$\log\left[\frac{\pi_{ij}}{1 - \pi_{ij}}\right] = \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij} + u_j + v_j \tag{5.3}$$

$$\beta \sim \text{Normal}(0, 10^{-4}) \tag{5.4}$$

$$u_j | u_k, j \neq k, \tau_u \sim \text{Normal}\left(\sum_{j \sim k} \frac{u_k}{\mathcal{N}_j}, \frac{1}{\mathcal{N}_j \tau_u}\right) \tag{5.5}$$

$$v_j \sim \text{Normal}(0, \tau_v) \tag{5.6}$$

$$\tau_u \sim \text{Gamma}(10^{-3}, 10^{-3}) \tag{5.7}$$

$$\tau_v \sim \text{Gamma}(10^{-3}, 10^{-3}) \tag{5.8}$$

The body of epidemiology and parasitology research (Basáñez et al., 2004; Clements et al., 2006; Soares Magalhães et al., 2011; Thomson et al., 1999) suggests that either ignoring *spatial heterogeneity* (vertical dependency) induced by the clustering of individuals within areas of residence or omitting *spatial autocorrelation* (horizontal dependency) among adjacent areas could result in severely underestimated uncertainty with respect to the estimation of regression parameters; in some cases, it could even result in biased estimates (see Goldstein, 2011; Jones, n.d.; Snijders and Bosker, 2012 for elaboration in general context). Chirombo et al. (2014) suggest that, technicalities aside, the spatially structured random effect u_j plays a crucial role in capturing the unmeasured between-area variation in access to health facilities and interventions, while the unstructured component v_j is useful for absorbing the unobserved level of immunity to malaria that varies randomly across the locations. In general, one may view this random effects specification as a method of incorporating the effects of

unmeasurable natural and social features that transcend political borders.

Prior distributions for the unknown random parameters are specified as follows. The regression parameter β, which determines how the risk of malaria is distributed across socio-demographic/economic strata in Papua, is assigned a diffuse normal prior with mean zero and extremely low precision (equation 5.4; henceforth all normally distributed prior distributions are defined in terms of mean and *precision*, not variance). The spatially structured random effect u_j is given a conditional autoregressive (CAR) prior (Besag et al., 1991) whose mean and precision depend on the structure as well as the number ($\mathcal{N}_j$) of the adjacent first-order neighbours ($j \sim k$) of each district (equation 5.5). The binary adjacency matrix for this prior is constructed using queen contiguity criteria (Bivand et al., 2008); the implied neighbourhood graph is shown in the top panel of Figure 5.2. This Markov random field (MRF) approach to spatial modelling has been recently applied to analyses of malaria in Malawi (Chirombo et al., 2014; Kazembe, 2007), antenatal care in Kenya (O'Meara et al., 2013), and childhood health outcomes in Tanzania, Malawi and Zambia (Kandala et al., 2009; Kazembe, 2013), among others. Best et al. (2005) and Kauermann et al. (2012) report the relatively good performance of the MRF model in comparison to other spatial-statistical and spatial-econometrics models. For the spatially unstructured random effect v_j, a typical normal prior with an exchangeable structure is assumed (equation 5.6). We then choose Gamma$(0.001, 0.001)$, a proper approximation of a Jeffreys non-informative prior (Lunn et al., 2012), as the default prior for the precisions of u_j and v_j (equations 5.7 and 5.8) although later (in Figures 5.6 and 5.8), we also conduct sensitivity analysis using alternative Gamma(a, b) hyperpriors that are widely used in disease mapping literature (Pascutto et al., 2000). Bernardinelli et al. (1995) and Eberly and Carlin (2000) proposed a method for eliciting proper priors for the precision of u_j and v_j that are based on the assumption that excess variability is shared equally by the spatially structured and the unstructured random effects. They showed that their method may help with the identification of each random effects, but we do not implement their proposal because it was formulated without the presence of hyperprior distributions.

Marginal posterior distributions of model parameters are obtained using integrated nested Laplace approximation (INLA), which is not only a valid but also an efficient alternative to the commonly used Markov Chain Monte Carlo (MCMC) simulation method (Bivand et al., 2015; Blangiardo and Cameletti, 2015; Martino and Rue, 2010; Schrödle and Held, 2011). INLA's efficiency makes Bayesian hierarchical modelling of large datasets feasible and allows for robustness analysis to be carried out quickly using several prior distributions. To facilitate interpretation, we derive quantities that

are of particular interest to policy-makers, such as the odds ratio ($\exp[\beta]$, $\exp[\xi_j]$), the probability of excess risk ($\Pr[\exp\{\xi_j\} > 1|\mathbf{y}] = \Pr[\xi_j > 0|\mathbf{y}]$), the baseline probability of malaria infection ($\text{logit}^{-1}[\beta_0 + \xi_j] = \text{logit}^{-1}[\beta_{0j}]$), as well as the fraction of district-level variance attributed to spatial autocorrelation ($\phi = \sigma_u^2/[\sigma_u^2 + \sigma_v^2]$). A deviance information criterion (DIC; Spiegelhalter et al., 2002) is used to evaluate the performance of the full model against the null. Where a density curve is not shown, we summarise the posterior distribution of a model parameter using its mean, accompanied by the 95% credible interval.

5.3 Results

5.3.1 Descriptive and bivariate analysis

The second column in Table 5.1 shows the univariate description of the sample. Confirming the official tabulation released by the Ministry of Health (Kemenkes, 2008), about one-fifth of study participants (21.06%) reported they had been infected with malaria. In the sample, sex appears to be distributed equally; about 60% of study participants are of working age (≥ 15 years old); the vast majority (78%) of them are ITN non-users or rural dwellers. With respect to elevation, it appears that only 6 out of 27 districts (22.22%) are categorised as highland districts ($\geq 1,200$ metre). It turns out that about half ($\hat{p} = 0.52$; SD = 0.24) of Papuan population live in the vicinity of forest; and assuming a historical 1 US Dollar (USD) to 10,000 Indonesian Rupiah (IDR) exchange rate, the district median per capita daily consumption expenditure is around USD 1.30 (SD = 0.50).

The magnitude of bivariate associations between an individual's malaria status and its predictors is presented in the rightmost column of Table 5.1. Confirming conventional wisdom, the analysis suggests that Papuans living in rural area or in poor, densely forested, lowland districts are at a relatively higher risk of contracting malaria than their counterparts in urban or highland settings. Age and gender do not seem to explain much of the between-individual variability in disease prevalence, although there appears to be a weak indication for the presence of a threshold effect in the relationship between age and malaria status. Contradicting a priori expectation, the analysis shows that the odds of being malaria-positive increase with respondents' use of ITN on the night prior to data collection. Of course, it is prudent to note that this paradoxical finding could arise from our application of a simple GLM that neither

Table 5.1: Descriptive and bivariate analysis

Variable	Summary statistic	Unadjusted odds ratio [95% CI]
Individual characteristics (N = 22,643)		
Malaria status:		
No	78.94%	
Yes	21.06%	
Sex:		
Male	49.62%	1.00
Female	50.38%	0.96 [0.90, 1.02]
Age group:		
0–4 (Infant)	12.39%	1.00
5–14	26.84%	0.93 [0.83, 1.04]
15–24	14.36%	0.90 [0.80, 1.02]
25–34	16.41%	1.02 [0.90, 1.15]
35–44	14.98%	1.01 [0.90, 1.15]
45–54	9.40%	1.03 [0.90, 1.18]
55+	5.62%	1.15 [0.98, 1.35]
Sleep under ITN:		
No	78.62%	1.00
Yes	21.38%	1.15 [1.07, 1.25]
Residential location:		
Urban	22.14%	1.00
Rural	77.86%	1.43 [1.31, 1.55]
District characteristics (N = 27)		
Median household elevation:		
Highland ($\geq 1,200$ metre)	22.22%	1.00
Lowland (≤ 200 metre)	77.78%	1.65 [1.51, 1.79]
Proportion living in or near forest	0.52 ± 0.24	1.07 [1.05, 1.08]
Median income:		
Quintile 1 (Poorest)	22.22%	1.00
Quintile 2	18.52%	1.41 [1.27, 1.57]
Quintile 3	22.22%	0.95 [0.87, 1.04]
Quintile 4	18.52%	0.91 [0.82, 1.01]
Quintile 5 (Richest)	18.52%	0.72 [0.66, 0.80]

adjusts for confounding nor accounts for the complex dependency structure of the data. Whether this unexpected ITN effect is simply a statistical artefact is to be tested in the multivariate analysis presented next.

5.3.2 Multivariate analysis

Figure 5.5 displays the confounding-adjusted odds ratios (diamond) along with their 80% (bold line) and 95% (fine line) credible intervals. The most striking feature of the analysis is that the odds of contracting malaria for individuals living in lowland districts versus those in highland districts have doubled from 1.65 (95% CI: 1.51–1.79) in the simple bivariate model to 2.99 (95% CI: 1.84–4.59) in the multivariate model.

Figure 5.5: Posterior means of adjusted odds ratio and their 80% and 95% credible intervals

Intercept	0.07 [0.04, 0.13]
Female vs Male	0.94 [0.88, 1.00]
5–14 vs Infant	0.94 [0.83, 1.05]
15–24 vs Infant	0.94 [0.82, 1.07]
25–34 vs Infant	1.09 [0.96, 1.23]
35–44 vs Infant	1.07 [0.93, 1.21]
45–54 vs Infant	1.07 [0.92, 1.23]
55+ vs Infant	1.18 [1.00, 1.39]
Sleep under ITN	1.25 [1.14, 1.38]
Rural vs Urban	1.43 [1.29, 1.57]
Lowland vs Highland	2.99 [1.84, 4.59]
Proportion living in or near forest	1.08 [1.00, 1.17]
Q2 vs Poorest District	0.98 [0.52, 1.66]
Q3 vs Poorest District	0.72 [0.41, 1.17]
Q4 vs Poorest District	0.71 [0.39, 1.19]
Richest vs Poorest District	0.62 [0.35, 1.01]

Living in a rural area (OR = 1.43, 95% CI: 1.29–1.57) and in a densely forested district (OR = 1.08, 95% CI: 1.00–1.17) are both associated with higher odds of being malaria-positive; their posterior means (credible intervals) do not, however, vary much from those obtained from the previous bivariate model. The analysis also makes the socio-economic gradient of malaria prevalence in Papua much clearer. The odds of being infected with malaria seem to follow a non-linear, monotonically decreasing function of district median income such that individuals living in the richest 20% of districts have 38% lower odds of being malaria-positive, holding all other factors constant. The multivariate analysis also presents evidence of the existence of a threshold effect in the relationship between age and malaria status, because only the elderly (55+ age group) seem to have a distinctively elevated risk of malaria. In addition, the sex difference is now more precisely estimated, with female individuals having 4% lower odds than their male counterparts. After controlling for all of these, however, we still find an unexpected positive ITN effect, with study participants who slept under a bed-net estimated to have 25% higher odds of contracting malaria. We discuss plausible explanations for this in the discussion section.

Table 5.2 compares the performance of the fully specified model against the null model. Clearly, the full fits better than the null, as its improvement in terms of model deviance ($\bar{D}$) far outweighs the increased model complexity (pD), leading to a 94.36 point

Figure 5.6: Posterior density of fixed effects coefficients (β) under some alternative hyperprior specifications, logit scale
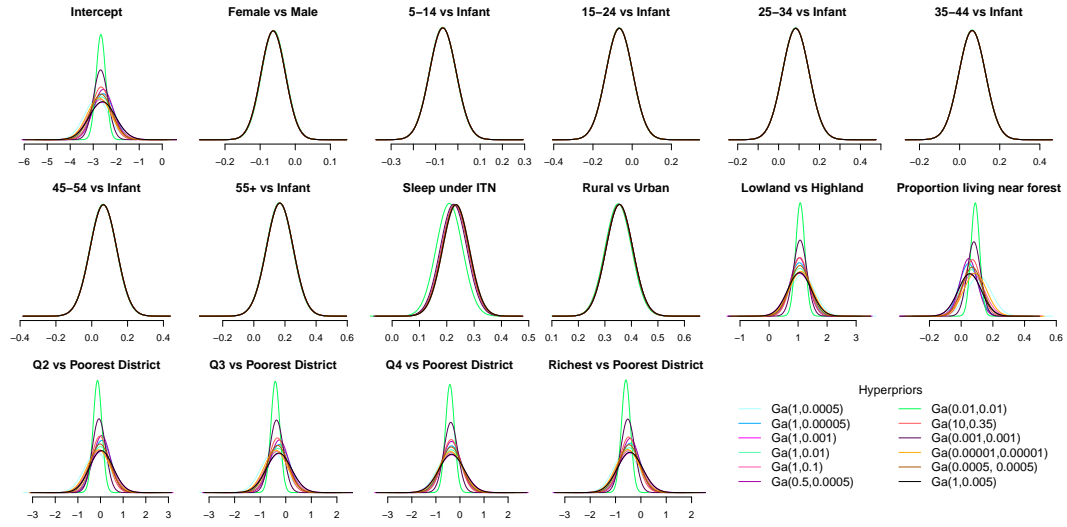
smaller DIC statistic. The covariates seem to have a strong explanatory power; their inclusion into the model leads to a 71% reduction in the between-district variability of malaria prevalence ($\sigma_u^2 + \sigma_v^2$). These covariates account for a disproportionately larger proportion of the spatially unstructured between-district variability (σ_v^2) than the structured one (σ_u^2), which, in turn, inflate the proportion of variance attributed to spatial autocorrelation (ϕ) from just 4% in the empty model to 32% in the full model. We should, however, note that ϕ is useful only if both components are well-identified (Eberly and Carlin, 2000; Lawson, 2013).

In Figure 5.6, we test the sensitivity of regression parameters with respect to the specification of alternative Gamma hyperpriors. Results show that the posteriors are robust to the choice of commonly suggested hyperpriors, albeit with some degree of variation around the width of the credible intervals of the intercept and contextual determinants. Nevertheless, since their means, medians

Table 5.2: Summary of model fit

Statistic	Null model	Full model
$\bar{D}$	20656.42	20553.76
pD	26.59	34.89
DIC	20683.01	20588.65
$\sigma_u^2 + \sigma_v^2$	0.76	0.22
ϕ	0.04	0.32

and modes are all very close, the interpretation above remains. This finding confirms the results of Bernardinelli et al. (1995) regarding the relative insensitivity to the choice of prior distribution of the fully Bayesian approach to disease mapping.

Having investigated the social and environmental correlates of malaria in Papua, we now turn our attention to Figure 5.7, which shows the spatial distribution of the disease. The raw odds ratio ($\exp[\xi_j]$) displayed in the top-left panel shows where in

Figure 5.7: Estimated malaria risk in each district
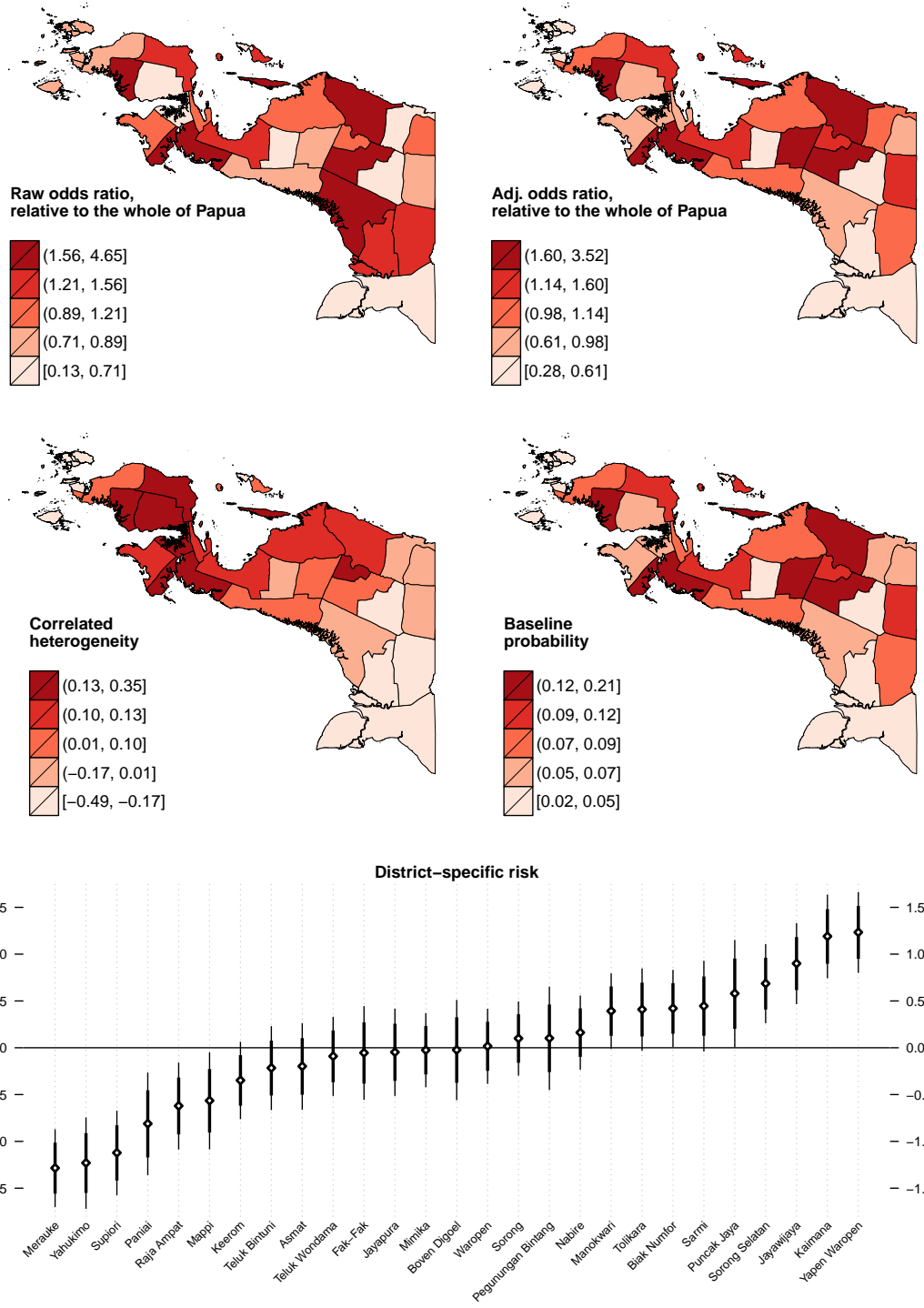
Table 5.3: Risk category, based on Richardson et al. (2004)

Positively significant[a]	Negatively significant[b]	Not significant[c]
Yapen Waropen	Merauke	Teluk Wondama
Kaimana	Yahukimo	Fak-fak
Jayawijaya	Supiori	Jayapura
Sorong Selatan	Paniai	Mimika
Biak Numfor	Raja Ampat	Boven Digoel
Puncak Jaya	Mappi	Waropen
Manokwari	Keerom	Pegunungan Bintang
Tolikara	Teluk Bintuni	Sorong
Sarmi	Asmat	Nabire

[a] $\Pr(\xi_j > 0|\mathbf{y}) \geq 0.80$
[b] $\Pr(\xi_j > 0|\mathbf{y}) \leq 0.20$
[c] $0.20 < \Pr(\xi_j > 0|\mathbf{y}) < 0.80$

Figure 5.8: Posterior density of district-specific effects (ξ_j) under some alternative hyperprior specifications, logit scale

109
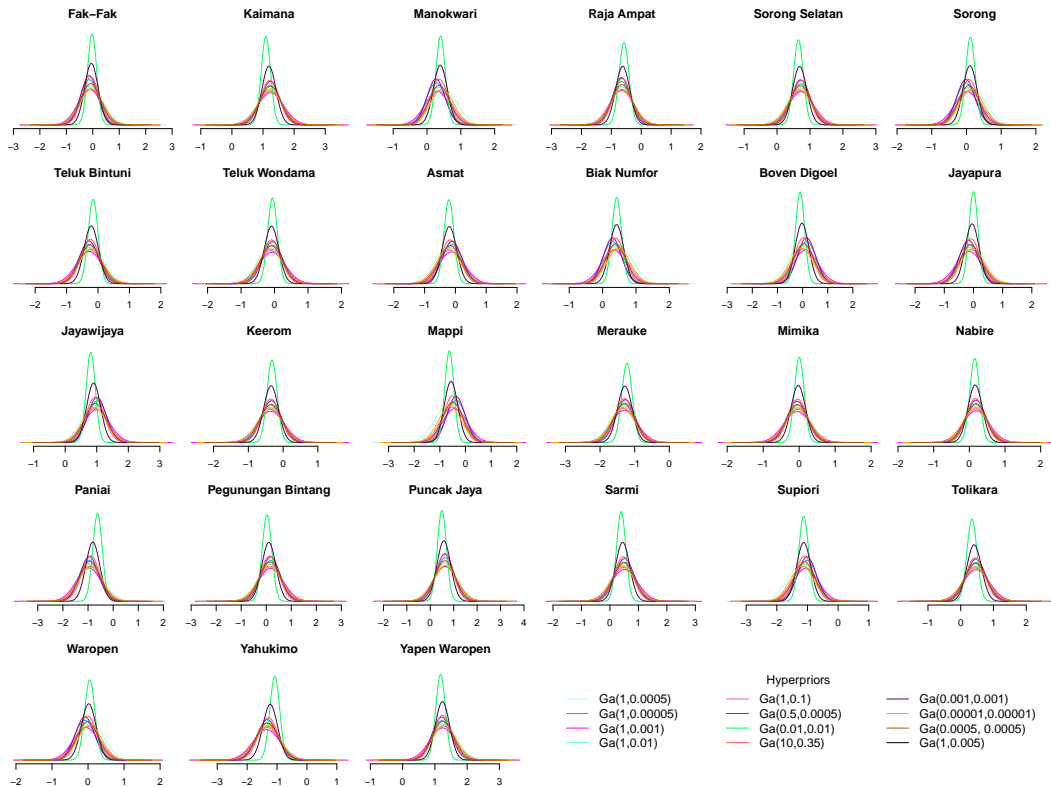
Papua malaria is more prevalent (Null model), whereas the adjusted odds ratio shown in the top-right panel indicates which district has higher than expected prevalence *after* accounting for the effect of predictor variables in the Full model. It should be appreciated that, although the spatial patterning of malaria does not seem to vary that much between the two models, its variability is clearly reduced after the inclusion of the covariates. Apparent in the middle-left panel is the gradient of spatially correlated heterogeneity (u_j, in logit scale) that varies smoothly from the north-western side (high risk) to the south-eastern side (low risk) of the island. The middle-right panel re-expresses the estimated risk in terms of how likely, in the probability scale, the reference individuals are to be infected with malaria in each district ($\text{logit}^{-1}[\beta_0 + \xi_j]$). Finally, in the bottommost panel of the same figure, we rank the district-specific risk estimates (ξ_j, in logit scale) along with their 80% (bold line) and 95% (fine line) credible intervals. It is evident from these plots that, net of differentials in observable characteristics, four districts have unambiguously higher-than-expected malaria risks (Yapen Waropen, Kaimana, Jayawijaya, and Sorong Selatan). However, it is only when we apply Richardson's criterion (Richardson et al., 2004) to the posterior probability distributions ($\Pr[\exp\{\xi_j\} > 1|\mathbf{y}]$) that we become aware of nine districts whose risks are deemed to be positively significant (Table 5.3). According to this criterion, clusters of elevated malaria risks are identified in north-central Papua, near the islands of Biak and Yapen, and around the north-western part of Papua. Figure 5.8 shows that this risk ranking exercise is robust to the assumption of hyperprior distributions.

5.4   Discussion and conclusion

Analysing a subset of the largest public health data ever collected in Indonesia (National Basic Health Research 2007; N = 987,205), this study quantifies the district-specific risk of malaria in Papua and investigates how the disease is distributed across socio-demographic/economic strata. We predict the malaria status of 21,740 Papuans living in 27 districts using a Bayesian logistic regression model that accounts for the clustering of individuals within their areas of residence and the spatial autocorrelation among these locations. Both individual (age, sex, bed-net use, urban/rural) and contextual characteristics (elevation, forest density, median income) are included as predictors in the model.

In the analysis, a spatial gradient that varies smoothly from the north-western (higher risk) to the south-eastern (lower risk) areas of the island is identified; after taking this patterning into account, we then calculate, rank and map malaria risk in each

district. We find that, even within this hyper-endemic island, the extent of spatial variation is *not* negligible. The model estimates that, while the baseline probability of malaria infection is about 2–5% in the healthiest 20% of districts, the figure can be as high as 12–21% in the least healthy ones. This means that a typical male Papuan infant would have a 4–5 times higher probability of suffering from malaria if he were born in high-risk districts instead of in low-risk districts. Whether or not this inequality is acceptable within the current climate of Papua's special autonomy (Resosudarmo et al., 2014) and Indonesia's political decentralisation (Hill, 2014) is, of course, open to public debate.

Our risk mapping exercise further reveals three clusters of statistically significant high-risk districts located in north-central Papua (Sarmi, Tolikara, Puncak Jaya and Jayawijaya), near the islands of Biak and Yapen (Biak Numfor and Yapen Waropen), and around the north-western part of Papua (Kaimana, Sorong Selatan and Manokwari). Because this risk ranking is independent of common socio-demographic/economic differentials and does seem to be robust to prior assumptions, health policy-makers or planners may, therefore, want to conduct further epidemiological studies in these areas to unravel the possible social and environmental drivers of this excess risk. Furthermore, should there ever emerge an urgent need for allocating limited funds or human-capital resources in order to help local autonomous Papuan administrators achieve the country's 2030 malaria elimination target (Kemenkes, 2009), the Indonesian government could now consider utilising risk estimates and probabilistic maps presented in this study as a tool for prioritising development projects or other forms of intervention that may be funded by transfers from central to local governments. Such risk mapping activity is of high relevance for policy-makers because the success of malaria control in many under-resourced countries often depends on targeted development of much-needed healthcare facilities in remote and sparsely populated areas (Elyazar et al., 2011a,b).

Independent of the aforementioned spatial effect, an elevated malaria risk is associated with living in rural areas, in densely forested districts, and in lowlands. This can be explained by the biology of the disease, as we have noted earlier. The literature suggests that these places provide not only a conducive environment for successful completion of the *plasmodium* growth-cycle (Alegana et al., 2014; CDC, 2015; Manh et al., 2011; Sachs and Malaney, 2002) but also a suitable breeding site and feeding ground for the *anopheles* vector (Ernst et al., 2009; Lowe et al., 2014; Mmbando et al., 2011; Stresman, 2010). Small increases in malaria risk are also associated with being male and with being over age 55. These differentials may be driven by social norms

with regards to gender roles and risk-exposure preferences (Chirombo et al., 2014; Haque et al., 2011; Mendis et al., 2001; Ricci, 2012)—for instance, women (children) should stay safe at home while men (adults) have to work outside to provide for the family—although we ought to note that these effects may yet be confounded by the respondent's immigration status. That non-native Papuans are more likely to seek malaria treatment and that they have lower acquired immunity to malaria due to their lack of exposure to infection are well-established in the literature (Baird et al., 2003; Barcus et al., 2007; Elyazar et al., 2011b; Karyana et al., 2008; Tjitra et al., 2008); unfortunately, information on individuals' immigration status is unavailable in this particular survey data.

We further find that, even after adjusting for all these conventional risk factors, the risk of malaria in Papua remains far from evenly distributed by income level. Papuans living in the richest districts are estimated to have 38% lower odds of having the disease than their peers in the poorest districts. So, if our reference infant were born in one of the richest districts, his estimated probability of being malaria-positive would be just 4% instead of 6%. This demonstrates that an income gradient in malaria prevalence indeed exists, even in Indonesia's most deprived island group (recall Figure 5.4). This finding is therefore consistent with the hypothesis that poverty creates conditions (poor housing, lack of knowledge, negative health behaviours) that favour the spread of infectious diseases and restrict access to prevention and treatment (Haque et al., 2011; Ingstad et al., 2012; Teklehaimanot and Meija, 2008; WHO, 2012a).

Contrary to conventional wisdom, our analysis reveals that respondents' use of ITN has a positive association with being malaria-positive. Initially, we suspected that this might be attributable to confounding or to an unaccounted data dependency structure in our naïve bivariate analysis. However, after fitting the fully specified multivariate model, the association persists. One plausible explanation is that a systematic bias due to differential item functioning (DIF) (Hanandita and Tampubolon, 2016a; Sen, 2002) is at play, meaning that ITN users may over-report their illnesses simply because they are more aware of malaria symptoms than their non-user peers (Opeskin, 2009; Somi et al., 2008; Sonko et al., 2014). An equally plausible explanation is that this counter-intuitive result is actually an artefact of the targeted distribution of ITN to the less healthy sub-population, such that individuals who use ITN are actually those who have already been infected (Pitt et al., 1995). Alternatively, it could simply be a mask for the unaccounted effect of non-native Papuans who tend to use the net because they have less immunity to malaria than the native. Another explanation, as documented in one ethnographic study from Malawi (Ingstad et al., 2012), is that

economically disadvantaged individuals may use the net to enhance their outdoor income-generating activities (such as fishing), which could in turn, lead to increased risk exposure. Understanding which of these scenarios fits the reality in Papua is, indeed, a good motivation for future investigations.

The present study is not without limitations. One is that, due to a lack of data, we are unable to investigate how the prevalence of malaria varies by individual income and immigration status. Secondly, we are unable to estimate malaria risks in *Kota Sorong* and *Kota Jayapura* municipalities because their spatial polygons are not available. A more serious limitation, however, pertains to our use of clinical malaria data, which are fraught with measurement error. In the presence of DIF, clinical data could overestimate the true prevalence of malaria; but in hyper-endemic areas, they may just as easily underestimate the true prevalence because of the presumably high incidence of asymptomatic malaria (Lowe et al., 2014; Sonko et al., 2014). Somi et al. (2008) point out that such measurement error, among other things, is often responsible for the attenuated estimates of socio-economic gradient in malaria prevalence (attenuation bias).

Despite these limitations, the present study still contributes to the literature in several ways. First, to the best of our knowledge, this study is among the first to provide a probabilistic characterisation of how malaria is distributed spatially *and* socially within Indonesian Papua. The Bayesian hierarchical modelling framework we adopt in this paper has proven to be useful and feasible for the purpose; policy-makers could, therefore, consider employing it more routinely in the planning and evaluation of malaria elimination efforts in the country. The study is further distinguished in its use of randomly sampled population data, which have helped us contain to a large extent the threat of confounding and selection biases that limit the generalisability of existing community or facility studies (Worrall et al., 2005). Finally, the present study shows that in addition to being geography-dependent, malaria in Indonesian Papua is also a disease of poverty. A comprehensive malaria elimination programme in this region should therefore consider not only proximal factors impacting the biology of the *plasmodium* parasite and the *anopheles* vector but also distal socio-economic conditions facilitating malaria transmission (Allotey et al., 2010; Lowe et al., 2014; Teklehaimanot and Meija, 2008; Tusting et al., 2013). This means that classical health interventions via bed-net distribution, insecticide residual spraying, curative medication, and environmental controls should ideally be implemented alongside development programmes in the forms of job-creation, investment in education, income redistribution, and provision of affordable and accessible healthcare facilities

(de Castro and Fisher, 2012; Ingstad et al., 2012). Unless the socio-economic factors that modulate the risk of infection are addressed, malaria elimination efforts in Papua will not be as effective as they are intended to be.

❀ ❀ ❀

Chapter 6

Multidimensional poverty in Indonesia: Trend over the last decade (2003–2013)

Abstract: The notion of poverty as an experience of multiple deprivation has been widely acknowledged. In Indonesia, however, poverty assessment has almost exclusively been conducted within the monetary space; even when multidimensionality is admitted, it has always been computed using variants of marginal method that are indifferent to joint deprivation. Applying a novel measurement method that is sensitive to both the incidence and the intensity of multiple deprivation to data from the National Socio-economic Survey (Susenas), this paper investigates the extent and the patterns of multidimensional poverty in Indonesia from 2003 to 2013 (N = 7,148,964). An Indonesian version of the Multidimensional Poverty Index is constructed by augmenting the existing consumption poverty measure with information on health and education. Results suggest that there was an unambiguous poverty reduction over the last decade at both national and sub-national levels. The data also reveal that progress has been inclusive across population subgroups, although spatial variation remains notable. The new poverty measurement method proves to be easily adaptable to the Indonesian context and could complement the methods currently employed by the Indonesian Statistical Bureau.

Keywords: poverty assessment, multidimensional poverty index, Indonesia, Susenas, Alkire-Foster method

6.1 Introduction

Income, or consumption poverty measures such as the World Bank's dollar-a-day headcount ratio (Ravallion et al., 2009), is still the most prevalent measure of poverty used across the globe. However, from Asia to Africa (Batana, 2013; Klasen, 2000; Santos, 2013; Ranis and Stewart, 2012; Yu, 2013), and across Europe to Latin America (Battison et al., 2013; Brandolini and D'Alessio, 1998; Whelan et al., 2004), scholars have consistently documented that the lack of money is not always an accurate proxy for deprivations that society cares about. It has been argued that money metrics do not tell the whole story of human suffering, because poverty is not only about one's inability to spend on essential goods and services. More than that, it is about one's inability to enjoy valuable beings and doings (Sen, 1985). Indeed, what is now generally accepted is a notion of poverty (or well-being for that matter) as an intrinsically multidimensional construct that encompasses the whole range of ways in which an individual can participate effectively in society.

Since the seminal works of Townsend (1979) and Sen (1985), different multidimensional poverty measures have been developed. Yet, as noted by Santos and Ura (2008: 1), 'some of the proposed measures seem to have incorporated a multidimensional perspective at the cost of giving up the simplicity and intuition that characterise the unidimensional measures'. Statistical approaches to multidimensional poverty measurement (Filmer and Pritchett, 2001; Sahn and Stifel, 2003), for instance, rely on multivariate or latent-variable techniques to the extent that parameters are completely data-driven, leaving evaluators with limited control over the measure. Some axiomatic alternatives such as Bourguignon and Chakravarty (2003), on the other hand, satisfy a number of useful measurement properties but do strictly necessitate the availability of cardinal data; in reality, vital social indicators such as literacy and completion of primary school are usually ordinal in nature.

In an attempt to address these problems, Alkire and Foster (2011a; henceforth AF) proposed a new sort of multidimensional poverty measure: one that is not only simple to construct, but also retains many of the properties of the well-known Foster-Greer-Thorbecke (FGT) measures (Foster et al., 1984) of unidimensional poverty measurement. The AF method combines the FGT with the counting approach (Atkinson, 2003), which is easy to understand and has a long history in sociology. The method deals with ordinal data in a straightforward manner by dichotomising individuals' achievement into deprived and non-deprived states. Aggregation is then performed, first across deprivations experienced by each individual, and then across

individuals, yielding a measure that is intuitively interpretable as the share of deprivations that poor individuals experience out of the total deprivations that the society could possibly experience.

As a generalisation of the classical FGT, the AF family of multidimensional poverty measures satisfies an array of desirable axioms (Alkire and Foster, 2011a). Foremost among them are their 'subgroup decomposition' and 'dimensional breakdown' properties, which allow the overall poverty measure to be broken down into its social, geographical or dimensional constituents in a way that is both conceptually and technically defensible. A thorough characterisation of joint deprivations is further made possible by the availability of partial indices that capture the incidence as well as the intensity of poverty. The methodology is also transparent in the sense that all parameters are under the control of the evaluator, allowing normative decisions with regard to the selection of indicators, dimensional and poverty cut-offs, and weighting schemes to be easily incorporated into the analysis. In fact, acknowledging these novelties, in 2010 the United Nations Development Programme (UNDP) replaced its Human Poverty Index (HPI; first published in 1997) with the new Multidimensional Poverty Index (MPI), based on the AF family of multidimensional measure (UNDP, 2014, 2010: 95).

Applying the AF method to the National Socio-economic Survey data from Indonesia, this paper seeks to estimate the extent and to investigate the regional as well as the temporal patterns of multidimensional poverty in Indonesia for 11 consecutive years spanning from 2003 to 2013. The aim of this study is *not to replace* the official consumption poverty estimate with a new one, but rather *to augment* the conventional poverty measure with additional information on health and education using the same data source that has historically been used to estimate the official consumption poverty figure in Indonesia. This version of an Indonesian multidimensional poverty index (MPI) is constructed in a way that income poor individuals are 'automatically' multidimensionally poor, but not the converse. The present study considers the following questions: taking into account income, health and education dimensions, how many Indonesians are poor overall? Are urban areas always better off? Which island of the archipelago is the most deprived? Did recent progress, if any, benefit the poorest of the poor? And what happened to gender and spatial inequities during the last decade?

Indonesia, the world's largest archipelagic state and the third-most populous developing country, is known for its exemplary achievement in terms of income poverty reduction and overall human development (Ranis and Stewart, 2012). However, there

is little research attempting to understand the nature of *simultaneous deprivations* experienced by its people. The majority of recent poverty evaluations have been conducted exclusively within the monetary space (Ilmma and Wai-Poi, 2014; Strauss et al., 2004; Sumarto et al., 2014); even when multidimensionality is sought, it has always been computed using variants of marginal method (BPS, 2015b; BPS et al., 2004) that are blind to joint deprivation (Alkire, 2011: 503–504).

To date, only two studies attempted to measure the extent of simultaneous deprivations in Indonesia. Alkire and Foster (2011a), in the earliest showcase of their methodology, provided a national poverty estimate for the year 2007 using the Indonesia Family Life Survey data (IFLS; Thomas et al., 2012). But it is known that the IFLS sampling frame is not entirely representative of the population (RAND, 2007); it neglects individuals living in the eastern islands of the archipelago (RAND, 2014), yielding a sample that favours the relatively well-developed areas in western Indonesia. Alkire and Santos (2014) carried out further study on Indonesia using the Demographic Health Survey (ICF International, 2012) data as a part of a grand endeavour to construct a globally comparable MPI (UNDP, 2010). While they are completely representative of the population, the DHS data do not, however, provide household consumption expenditure information, preventing a useful comparison with the official measure of consumption poverty.

The contribution of the present study to the existing literature is threefold. Firstly, in estimating the extent of multidimensional poverty in Indonesia, this study uses large and nationally representative data that have been regarded as the primary source of information among Indonesian policy-makers as well as international observers. While concurring with Alkire and Santos (2014: 266) who stress that data availability has been the major bottleneck in the development of an internationally comparable MPI, we would like to demonstrate that even when using an existing data source, the construction of an Indonesian MPI is not only technically feasible but also substantively meaningful. The collection of better well-being data is of course desirable, but Indonesians do not have to wait until the 'perfect' data becomes available to have their progress assessed. Secondly, the inclusion of consumption expenditure information makes this version of Indonesian MPI not only comparable to the official poverty measure, but also sensitive to economic fluctuations (Ravallion, 2010: 11). Lastly, by providing an annual analysis of the trend of multidimensional poverty in the last 11 years, this study presents a richer picture compared to one that analyses only selected points in time over the same period.

The remainder of the paper is structured as follows. The next section describes the AF

method, the data and the dimensions. It then investigates the degree to which income poverty correlates with non-income deprivations. Section 6.3 presents the results. Initially, unidimensional deprivations are investigated using the marginal dashboard approach. Then, MPI estimates at national and sub-national levels are presented along with robustness checks. Finally, changes in the distribution of deprivations among the poor are studied. Section 6.4 concludes.

6.2 Methods

6.2.1 The Alkire-Foster Method

This section describes the Alkire-Foster method for multidimensional poverty measurement (Alkire and Foster, 2011a). For brevity, we focus only on those aspects of the methodology that are directly relevant to the present study. We also limit our attention to the general case, where the social indicators being considered might not have cardinal meaning. Further in-depth expositions are available in Alkire and Foster (2011a), Alkire and Foster (2011b), Alkire and Santos (2013) and Seth and Alkire (2014).

Setup

Before describing the identification and the aggregation steps of the Alkire-Foster method, it is necessary to outline some preliminary setups. First of all, let us consider an $n \times d$ dimensional achievement matrix $x = [x_{ij}]$, in which the row $i = 1, \ldots, n$ indexes the individuals under study and the column $j = 1, \ldots, j$ indexes indicators for every dimension that the society cares about. No restriction is placed on the cardinality of indicators entered into the matrix; ordinal variables are acceptable. In this matrix, individuals' achievements are recorded in the row vectors ($x_{i\cdot}$), while the marginal distribution of achievements is reflected in the column vectors ($x_{\cdot j}$). We also define a deprivation cut-off vector $z = (z_1, \ldots, z_j)$ indicating the minimum level of achievement in every social indicator that should be attained by each individual in the society. The relative importance (trade-off) of each achievement indicator in the achievement matrix x is governed by a vector of weight $w = (w_1, \ldots, w_d)$ such that $\sum_{j=1}^{d} w_j = d$ or $\sum_{j=1}^{d} w_j = 1$. Of course, the choice of indicator, deprivation cut-off, and weighting scheme is largely contingent upon the specific context of study (what

the society values, the aim and scope of the study, or data availability) and is open to public debate (Alkire, 2011).

Identification

In the AF framework, identification begins with the construction of a deprivation matrix $g^0 = \left[g_{ij}^0 \right]$ whose element is defined as $g_{ij}^0 = w_j$ if $x_{ij} < z_j$ and $g_{ij}^0 = 0$ if otherwise. This deprivation matrix contains information about 'who is deprived in which indicator and how much weight the indicators carry' (Alkire and Santos, 2013: 242). From g^0 matrix, a deprivation count vector $c = (c_i, \ldots, c_n)$ whose element is $c_i = \sum_{j=1}^{d} g_{ij}^0$ is constructed. This column vector stores the sum of weighted deprivations experienced by each individual under study. The Alkire-Foster identification function $\rho_k(x_i; z)$ is such that $\rho_k(x_i; z) = 1$ if $c_i \geq k$ and $\rho_k(x_i; z) = 0$ if otherwise, where k is the poverty cut-off denoting the minimum sum of weighted deprivations required to be multidimensionally poor (Alkire and Santos, 2013). The plausible choice of poverty cut-off is $k \in \left[min(w_j), d \right]$ and like other parameters in the AF framework, its value may be subjected to sensitivity analysis. Naturally, one would expect that the larger the k, the smaller the number of individuals identified as multidimensionally poor, and vice versa. When $k = min(w_j)$, a *union* identification criterion is obtained, but when $k = d$, an *intersection* identification criterion is reached. In practice, however, an *intermediate* criterion ($k = 0.33–0.50$) is usually preferred (Alkire, 2011).

Having assessed how deprived each individual is and identified who the poor are, the next step is to construct a censored deprivation matrix $g^0(k) = \left[g_{ij}^0(k) \right]$ whose element is defined as $g_{ij}^0(k) = g_{ij}^0$ if $c_i \geq k$ and $g_{ij}^0(k) = 0$ if otherwise. Likewise, a censored deprivation count vector is constructed such that $c_i(k) = c_i$ if $c_i \geq k$ and $c_i(k) = 0$ if otherwise. This censoring mechanism allows analysts to focus only on those individuals who are identified as multidimensionally poor, guaranteeing that the aggregate poverty measure is insensitive to the achievement of non-poor individuals.

Aggregation

The main aggregation method in the AF family of multidimensional poverty measure is the *adjusted headcount ratio* or M_0, which is 'the proportion of weighted deprivations that the poor experience in a society out of all the total potential deprivations that the society could experience' (Santos, 2013: 261). It is obtained by taking the arithmetic

mean of the censored deprivation matrix $g^0(k)$:

$$M_0(x;z) \quad = \quad \frac{1}{nd} \sum_{i=1}^{n} \sum_{j=1}^{d} g_{ij}^0(k) \tag{6.1}$$

$$= \quad \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left[\frac{1}{d} \sum_{j=1}^{d} g_{ij}^0(k) \right]}_{\text{individual poverty}} \tag{6.2}$$

$$= \quad \frac{1}{d} \sum_{j=1}^{d} \underbrace{\left[\frac{1}{n} \sum_{i=1}^{n} g_{ij}^0(k) \right]}_{\text{censored deprivations}} \tag{6.3}$$

$$= \quad \underbrace{\frac{1}{n} q(k)}_{\text{incidence (H)}} \underbrace{\left[\frac{1}{q(k)} \sum_{i=1}^{q(k)} \frac{c_i(k)}{d} \right]}_{\text{intensity(A)}} \quad \text{where} \quad q(k) = \underbrace{\sum_{i=1}^{n} \rho_k(x_i;z)}_{\text{headcount}} \tag{6.4}$$

Intuitively, M_0 can also be understood either as the weighted sum of individual poverty (equation 6.2), the weighted sum of censored deprivations by indicators (equation 6.3), or the intensity-adjusted poverty incidence (equation 6.4: $M_0 = H \times A$). The measure is, as its name implies, simultaneously sensitive to both the prevalence (incidence) and the scope (average deprivation among the poor, or intensity) of poverty. By definition, it is expected that as k increases, H will get smaller and A will get larger, and vice versa.

Decomposition

Because the adjusted headcount ratio can be expressed as the weighted sum of individual poverty (equation 6.2), the measure is decomposable by population subgroups. It follows that overall poverty can be expressed as the weighted sum of poverty measures in l number of population subgroups:

$$M_0 = \sum_{s=1}^{l} \frac{n_s}{n} M_0^{(s)} \tag{6.5}$$

and the contribution of a population subgroup s to the overall poverty M_0 is:

$$C_s = \frac{n_s}{n} \times \frac{M_0^{(s)}}{M_0} \quad \text{for} \quad s = 1, \dots, l \tag{6.6}$$

where $\frac{n_s}{n}$ and $M_0^{(s)}$ are the population share and the adjusted headcount ratio of subgroup s, respectively. Such a decomposition enables the assessment of the extent of inequality among subgroups by comparing each subgroup's contribution to overall poverty relative to its population share (Alkire and Santos, 2013: 245). A severe deviation from $C_s/\left(\frac{n_s}{n}\right) = 1$ is indicative of the fact that a particular subgroup bears a disproportionately large (or small) share of poverty.

Similarly, the adjusted headcount ratio can also be broken down by its indicators, because the measure is expressible as the weighted sum of the censored deprivations by indicators (equation 6.3). The overall poverty can thus be expressed as:

$$M_0 = \sum_{j=1}^{d} \left(\frac{w_j}{d}\right) h_j(k) \qquad (6.7)$$

and the contribution of a social indicator j to overall poverty M_0 is:

$$C_j = \frac{w_j}{d} \times \frac{h_j(k)}{M_0} \quad \text{for} \quad j = 1, \ldots, d \qquad (6.8)$$

where $h_j(k)$ is the censored headcount ratio of indicator j. From this, we know that whenever $C_j/\left(\frac{w_j}{d}\right)$ deviates severely from unity, then there is a relatively high (or low) deprivation in an indicator (Alkire and Santos, 2013: 245). Dimensional contribution is obtainable simply by adding up C_j within a particular dimension.

Robustness analysis

In the AF framework, robustness is established through sensitivity analysis employing different sets of indicators, deprivation cut-off, weight, or poverty cut-off (Alkire and Santos, 2014). In this study, we apply poverty cut-off dominance analysis, confidence intervals overlap testing, and rank correlation testing, which constitute the standard robustness toolbox for the AF family of multidimensional poverty measures.

Rate of change

Once some degree of robustness has been established, the rate of inter-temporal change in aggregate poverty can be calculated (Alkire and Vaz, 2014). The absolute

(ΔM_0) and relative (δM_0) rates of change are defined as follows:

$$\Delta M_0 = M_0^{(t2)} - M_0^{(t1)} \tag{6.9}$$

$$\delta M_0 = \frac{M_0^{(t2)} - M_0^{(t1)}}{M_0^{(t1)}} \times 100 \tag{6.10}$$

where $t2$ and $t1$ denote the later and the initial time points, respectively. When the two time points span over a number of years, it is sometimes useful to express the changes in their annualised values:

$$\bar{\Delta} M_0 = \frac{M_0^{(t2)} - M_0^{(t1)}}{t2 - t1} \tag{6.11}$$

$$\bar{\delta} M_0 = \left[\left(\frac{M_0^{(t2)}}{M_0^{(t1)}} \right)^{\frac{1}{t2-t1}} - 1 \right] \times 100 \tag{6.12}$$

which give us the average absolute (or relative) change during the period of observation.

Inequality among the poor and across subgroups

Finally, after assessing the incidence and intensity of multidimensional poverty, it is only natural to ask whether poverty reduction, if any, has been inclusive among the poor and uniform across population subgroups (the 'triple I' of poverty: incidence, intensity, inequality; Sen, 1976). Finding the right inequality measure for such a purpose has been proven to be non-trivial in the multidimensional setting; Seth and Alkire (2014) recently proposed a decomposable inequality measure based on the *positive-multiple of variance* to overcome the obstacles stemming mainly from the use of non-cardinal indicator variables in the construction of M_0. Following their proposal, inequality (I) among poor individuals (I^q) and across subgroups (I^s) can be expressed as:

$$I^q = \tilde{\beta} \times \frac{1}{q(k)} \sum_{i=1}^{q(k)} [c_i(k) - A]^2 \tag{6.13}$$

$$I^s = \tilde{\beta} \times \sum_{s=1}^{l} \frac{n^s}{n} \left[M_0^{(s)} - M_0 \right]^2 \tag{6.14}$$

where $\tilde{\beta}$ is a normalisation factor that must be chosen such that $I = [0,1]$, respecting the properties of any standard inequality index. Because it is known that 'the

maximum possible value that variance takes is one fourth of the range of the deprivation score vector, which is attained when half of the population have the lowest deprivation scores and the other half have the highest deprivation scores' (Seth and Alkire, 2014: 16), $\tilde{\beta}$ in the between-poor-individual equation equals the inverse of $\frac{1}{4}\left\{max\left[c_i(k)\right] - min\left[c_i(k)\right]\right\}^2$. Accordingly, as $M_0 = [0,1]$ then it is obvious that $\tilde{\beta} = 4$ in the between-subgroup equation.

6.2.2   Data

We analyse data from the *Survei Sosial Ekonomi Nasional* (National Socio-economic Survey; henceforth Susenas), an annual cross-sectional household survey administered by the Indonesian Statistical Bureau (*Badan Pusat Statistik*, BPS). Initiated in 1963, Susenas is a large and nationally representative survey, which has for decades served as the main source of information not only for the government of Indonesia but also for many international bodies, including the World Bank PovcalNet (see also Surbakti, 1995; van de Walle, 1988). The survey consists of a yearly core module (health, education, employment, household consumption expenditure, housing, fertility, contraception, and communication) and one of three alternating modules on (1) culture and education, (2) housing and health, and (3) household consumption expenditure, each administered once every three years.

Compared to other Indonesian household survey data available to date, the strength of Susenas lies in (1) its comprehensive information on consumption expenditure (more than 300 food and non-food items in 2013), education and literacy; and in (2) its large sample and periodicity which permit precise annual inferences to be made at low levels of administration. However, it should be noted that the information on health available in Susenas is neither as comprehensive as that available in the Demographic and Health Survey (DHS) nor in the Indonesia Family Life Survey (IFLS) that were analysed previously by Alkire and Santos (2014) and Alkire and Foster (2011a). While Susenas records information on the number of disabled days and morbidity for each individual, it does not provide any anthropometric measure. Of course, a household survey that was first designed nearly four decades ago is by no means ideal for the contemporary purpose of multidimensional poverty measurement, but aside from this, the consumption expenditure data available in Susenas provide us with the opportunity to address the concern about the unresponsiveness to economic fluctuations (Ravallion, 2010: 11) of the living standard indicators used in the current version of UNDP's Multidimensional Poverty Index (Alkire and Santos, 2014; UNDP,

2010).

Since the Susenas sample is drawn using a multi-stage stratified random sampling design (urban/rural stratification, census blocks as the primary sampling unit, households within each block as the secondary sampling unit), the survey design along with the sampling weight is always incorporated into analysis. Our exploration indicates that ignoring the unequal sampling probability underestimates the proportion of individuals living on Java island severely (30% instead of 60%), leading to a potentially biased estimate of the population parameter.

We analyse eleven consecutive years of Susenas data, from 2003 (just before the enactment of Law 32/2004 on Regional Government that marked the decentralisation era; 346 districts) to 2013 (the latest available Susenas; 499 districts). We regroup all districts that split during the period of observation into their original 2003 districts. The unit of analysis is an individual aged 18 and older. Children are excluded from the analysis because the relevant dimensions of their well-being depend on their age (Roche, 2013), and information on those dimensions are missing from Susenas. We believe that children deserve special consideration that takes into account their own specificities (see Trani et al., 2013 and the cited works therein). Only complete cases are used in the analysis: individuals that have any social indicator (presented next) missing are dropped from the sample. This yields a total complete-case sample size of 7,148,964 individuals ($N \approx 650,000$ per year) with each survey wave having a final sample size of 91–100% of the original sample size.

6.2.3   Dimensions, indicators, cut-offs and weights

In an ideal world, the choice of dimensions, indicators, cut-offs and weights for the measurement of multidimensional poverty would be guided by the revealed preferences of the poor (what the poor think of being poor, what deprivations matter the most, and what trade-offs the poor assign between deprivations). Yet, with the notable exceptions of Mexico's *Voices of the Poor* study (Székely, 2003) and Bhutan's *Gross National Happiness* survey (Santos and Ura, 2008; Ura et al., 2012), large-scale participatory exercises are rare. In contrast to these countries, the official conceptualisation of poverty in Indonesia is still the traditional consumption (or income) poverty measurement, defined as the failure to attain the consumption level required for the fulfilment of a basket of basic food and non-food needs (BPS, 2015c). The idea of poverty as an experience of simultaneous deprivations has rarely penetrated the

Table 6.1: Dimensions, indicators, deprivation cut-offs and relative weights

Dimension	Indicator variable	Deprivation cut-off	Weight		
		'An individual is deprived if ...'	$w1$	$w2$	$w3$
Income[1]	Per capita daily consumption	< $1.51 PPP	$1/3$	$1/3$	$1/3$
Health[2]	Illness episode	> 4 days	$1/6$	$1/6$	$1/3$
	Morbidity	> 3 diseases	$1/6$	$1/6$	0
Education[3]	Schooling	Has not completed primary school	$1/6$	$1/3$	$1/3$
	Literacy	Cannot read and write Latin characters	$1/6$	0	0

[1] The first MDGs. The fourth paragraph of preamble, article 27(2) and 28C(1) of the Constitution.
[2] The fourth, fifth and sixth MDGs. Article 28H(1) and 34(3) of the Constitution.
[3] The second MDGs. The fourth paragraph of preamble, article 28C(1), 31(1) and 31(2) of the Constitution.

nation's discourse of development (for example see Hill, 2014 or Strauss et al., 2004). In this light, we base our elicitation of dimensions, indicators, cut-offs and weights on the existing Human Development Index (HDI; UNDP, 2010), Multidimensional Poverty Index (MPI; UNDP, 2010), Millennium Development Goals (MDGs), and the 1945 Constitution of the Republic of Indonesia (MPR RI, 2011), subject to constraints imposed by data availability in Susenas survey.

Three dimensions are included in our version of Indonesian MPI: (1) income, (2) health and (3) education, mimicking the UNDP's latest HDI and MPI. Alkire and Santos (2014: 253) note that these dimensions are not only instrumental to many other vital outcomes but also intrinsically valuable in themselves. Furthermore, they argue that having only three dimensions simplifies communication and interpretability because 'the contribution of the chosen dimensions is widely recognized across political and ideological divides'. As shown in Table 6.1, these dimensions are clearly related to the values of the Constitution, not to mention the MDGs.

Income is operationalised using per capita daily consumption, which is obtained by deflating total daily household expenditure by household size. The figure is measured in international dollar (an expression of purchasing power parity, or PPP; UNSD, 2014) and adjusted for spatial cost-of-living differences using the provincial urban-rural adjustment factors derived from the relative differences between the national and the local poverty lines (Ilmma and Wai-Poi, 2014: 132). Fixed adjustment factors (from 2008) are used for the entire 2003–2013 series because these data are not available prior to 2007 (BPS, 2015a; see also Alkire et al., 2013: 2–4 for a similar approximation method). We consider an individual to be deprived in the income domain if his or her daily consumption is less than the Asia-specific poverty line of $1.51 (ADB, 2014). This cut-off is more stringent than both Indonesia's national poverty line ($1.43) and the World Bank's extreme poverty line ($1.25).

Health status is assessed using two indicators: illness episode (number of days disabled

within the last month) and morbidity (number of illnesses within the last month), which, admittedly, may not be as informative as the body mass index (BMI) indicator used in Alkire and Foster (2011a) and in Alkire and Santos (2014). However, these are the best available health measures in the Susenas survey, and since the disabling burden of poor health may lead not only to potentially missed income-generating opportunities (Schultz and Tansel, 1997) but also, ultimately, to an unfulfilled life, we consider that these indicators make a good representation of the health domain. The inclusion of the illness episode indicator variable into a multidimensional poverty index is not new; such a measure has been used previously in a version of Bhutanese MPI (Santos, 2013). In addition, the measure is often employed to operationalise Grossman's model of health production function (Grossman, 1972) in the health economics literature. An individual is deprived in health if he or she was ill for more than 4 days or caught more than 3 diseases within the last month (Table 6.1). These, we believe, are reasonable cut-offs considering the high prevalence (60–70%) of informal self- and seasonal employment in Indonesia (Nazara, 2010).

Like health, education is also operationalised using two indicators: the completion of primary school (schooling) and the ability to read and write Latin characters (literacy). An individual is deprived if he or she has not completed primary education or is illiterate (Table 6.1). Relative to other indicators described above, primary schooling and literacy are perhaps the most universally accepted social indicators. They are highly valued worldwide: their presence in the HDI, MPI, MDGs and even the 1945 Constitution of the Republic of Indonesia testifies to this.

Having chosen the social indicators to be included in the multidimensional poverty index, we now define their weights, which are necessary for identification purposes (Alkire, 2011: 14–16). This weight assignment, which means making the trade-offs between social indicators explicit, clearly entails value judgements (Decancq and Lugo, 2013). In this study, as in many other applications of the Alkire-Foster method, we use a normative weight because of the unavailability of preferences data. For the proposed Indonesian MPI, an equal-nested weighting scheme, which assigns an equal relative weight $\left(\frac{1}{3}\right)$ to each dimension and also an equal weight to all indicators within a dimension, is used (weight $w1$ in Table 6.1). We then set the poverty cut-off to $k = \frac{1}{3}$ so that an income-poor individual is 'automatically' multidimensionally poor, but not the converse. This choice of parameters reflects the beliefs that (1) income, health and education are equally important for human development and (2) income still holds a special position in poverty measurement 'given its fungibility and its key role in facilitating other capabilities' (Foster, 2007: 9). Notwithstanding the preference

Table 6.2: Spearman correlation matrix of deprivations (2003–2013 maximum value; unweighted sample)

	Income	Illness episode	Morbidity	Schooling	Literacy
Income	1.00				
Illness episode	0.02	1.00			
Morbidity	0.01	0.23	1.00		
Schooling	0.14	0.10	0.06	1.00	
Literacy	0.13	0.10	0.05	0.34	1.00

for this setting, we still conduct sensitivity analysis employing alternative weighting schemes (weights $w2$ and $w3$ in Table 6.1) and/or poverty cut-offs for $k \in \left[\frac{1}{6}, 1\right]$.

At this point, critics may contend that an index of multidimensional poverty is unnecessary because the social indicators included in its construct are, presumably, highly correlated either to income or to each other, representing a 'double counting'. Our Indonesian data prove that this is not the case. As shown in Table 6.2, the correlation between income poverty and other deprivations in health and education during the last decade is never larger than 0.14; the figure among indicators of health and education is always less than 0.35. This mismatch between income poverty and deprivations in other social indicators conforms to the general finding in the literature (see Battison et al., 2013 on Latin American countries; Batana, 2013 and Klasen, 2000 on Africa; Brandolini and D'Alessio, 1998 on Italy; Ranis and Stewart, 2012 on Bangladesh, Chile, Indonesia, Kazakhstan, Laos and Zambia; Santos, 2013 on Bhutan; Whelan et al., 2004 on Europe; Yu, 2013 on China), providing 'good empirical basis to support a multidimensional approach to poverty measurement, which goes beyond income and asset ownership' (Santos, 2013: 267).

6.3 Results

6.3.1 Unidimensional deprivations

We begin by describing the trend of unidimensional deprivations in Indonesia during the 2003–2013 period. The top-left panel of Figure 6.1 shows a strong 83% income poverty reduction at the national level. Nearly half (46%) of adult Indonesians were income poor in 2003, but a decade later, the figure improved significantly to just 8%, registering an absolute 0.38 point reduction. It is apparent from the trend-line that income poverty reduction is characterised by two quinquennial regimes. Reduction

Figure 6.1: Income poverty
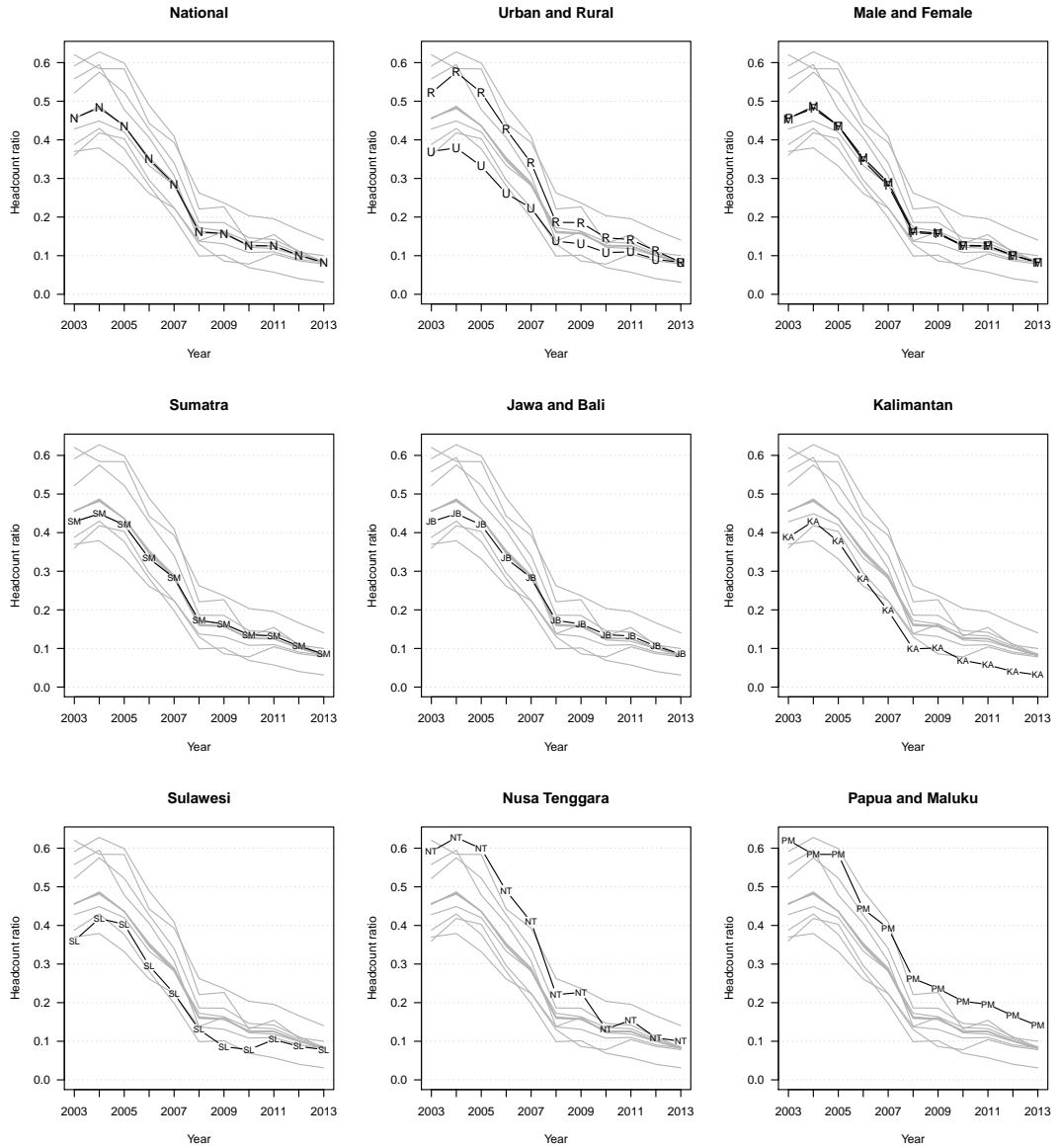
was faster in the 2003–2008 period (65%) than in the 2008–2013 period (50%), a phenomenon that is consistent with the fact that the global economy was relatively more buoyant in the former period (high real growth rate and high commodity prices; Bourguignon et al., 2008: 12–13) than in the latter (the 2008 financial crisis and the ensuing drop in commodity prices; Battison et al., 2013: 308). At the same time, this discontinuity could reflect the differing efficacy in governance between the first (2004–2009) and the second United Indonesia Cabinets (2009–2014).

The top-middle panel of Figure 6.1 shows that income poverty reduction has been accompanied by a substantial improvement in the urban/rural disparity. The rural-to-urban poverty ratio fell sharply from 1.41 in 2003 to only 1.01 in 2013. The gender gap had not been of serious concern over the 11 years of observation, as the female-to-male poverty ratio has hardly ever deviated from the 1.00–1.01 range (the top-left panel of Figure 6.1). However, we should keep in mind that this figure is obtained from household expenditure data rather than from individual income data. The trend in regional disparity seems to be similar to that of the urban/rural one. When we analyse each island-group separately (the middle to bottom panels of Figure 6.1), the pattern of two-regime poverty reduction pattern still holds, while the between-island variance shrank by 90% from 0.013 in 2003 to 0.001 in 2013, indicating a converging regional poverty. Despite this tremendous progress, it should be noted that Papua and Maluku, whose headcount ratio has remained constant at 1.5 times higher than the national average since 2008, seem to be left behind. Moreover, it is noticeable that Sulawesi has failed to register any significant improvement after 2008. This suggests that while Indonesia has enjoyed substantial progress in terms of sharply reduced income poverty and gradually converging urban/rural as well as regional disparities during the last decade, the East-West divide remains (see also poverty maps in Figure 6.8).

Having described the state of income poverty, we now investigate trends in health and education (Figure 6.2). Since non-income achievements are usually represented by stock rather than flow variables, and are therefore unlikely to change in the short run (Battison et al., 2013: 296), it is expected that their trend-lines will be relatively more stable than that of income.

Some clear patterns emerge from the two plots in the top panel of Figure 6.2, which displays the evolution of health deprivations over the decade. First, while the reduction of income poverty was at its fastest rate (2003–2008), the nation's illness episode deprivation increased by 7% year-on-year (13% for morbidity) before eventually peaking in 2007 and then gradually returning to its initial level in 2010/2011. This

Figure 6.2: Non-income deprivations
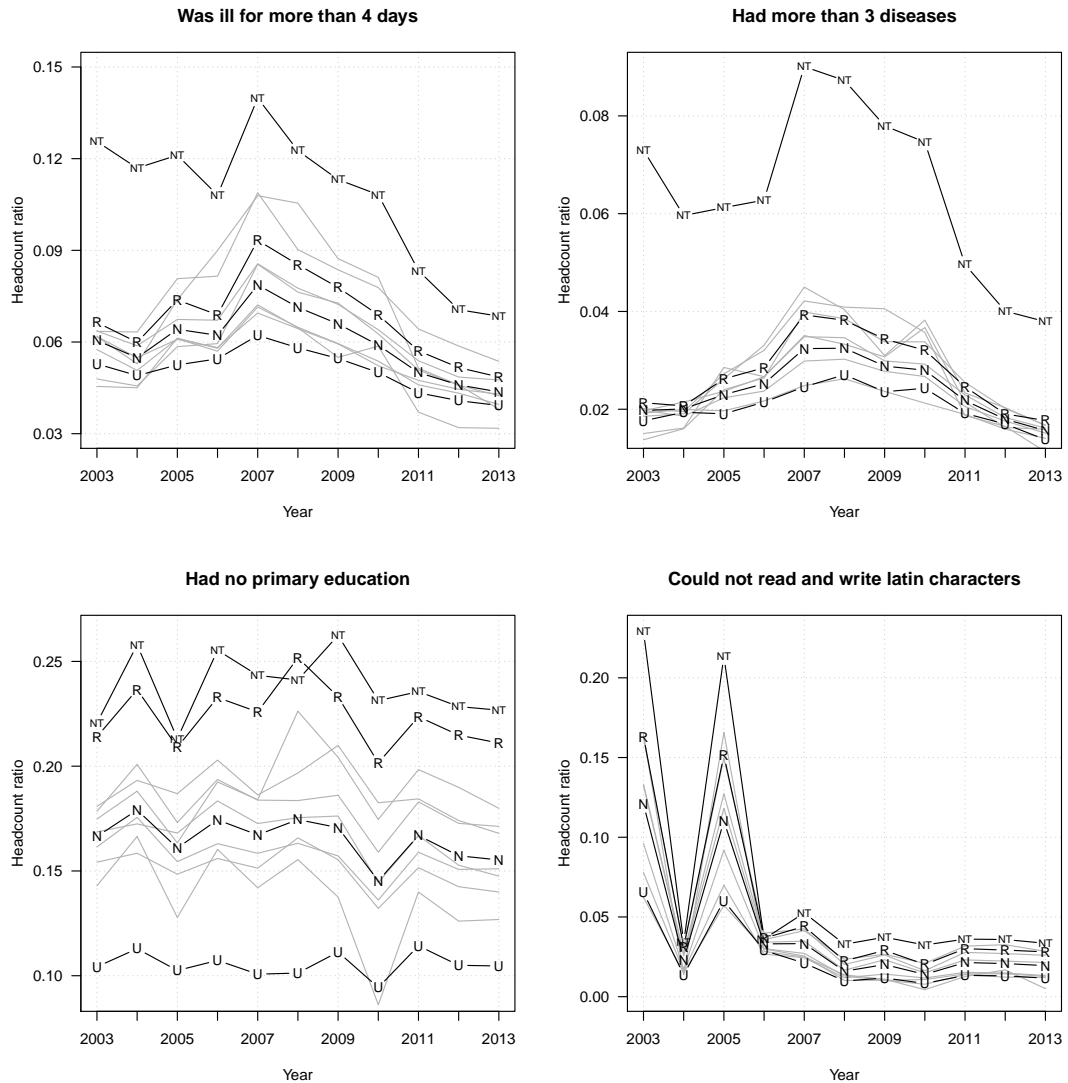
inverted U-shape trend-line suggests that there may have been a short-term surge of negative health-related behaviour that followed the rising income level. In addition, it might also capture the health cost of both natural and man-made disasters (the Indian ocean earthquake and tsunami, the Java earthquake, the Sumatra flood and earthquake, the Sulawesi flood and landslide, and the Sidoarjo mud flow, to name only a few) that occurred relentlessly during the 2004–2007 period. Second, the trajectory of rural-to-urban health deprivation ratio also follows this inverted-U shape. The figure was about 1.20, 1.50 and 1.20 in 2003, 2007 and 2013, respectively, suggesting little to no improvement in terms of urban/rural health equality. The plots also present evidence regarding the disturbingly weak health status in Nusa Tenggara islands. Illness episode deprivation in Nusa Tenggara was 1.53–2.13 times greater than the national average and the figure for morbidity was in the range of 2.23–3.70

times greater. Moreover, it is evident that Nusa Tenggara exhibits a very distinctive trend-line compared to the rest of Indonesia, thereby exerting undue influence over the between-island variability. In contrast to the patterns of urban/rural and regional disparities, the female-to-male health deprivation ratio has always been stable in the range of 0.80–0.90, indicating that Indonesian women seem to be slightly healthier than their male counterparts (trend-lines greyed out).

The two plots at the bottom panel of Figure 6.2 show the trends of schooling and literacy deprivations. With the exception of two irregularly spiking literacy deprivations in 2003 and 2005, which appear to be a data quality problem, the trends seem to be stable over the window of observation. The rural-to-urban education deprivation ratio was constant in the range of 2.00–2.25 for both indicators. The female-to-male deprivation ratio was at about 1.20 for schooling and 1.80 for literacy; and the between-island variability has barely changed over the decade. It is evident that nearly a quarter of Indonesian adults living in rural areas failed to complete primary school, despite the substantial reduction in income poverty and the constitutional mandate for the provision of universal primary schooling. Again, by studying all four plots shown in Figure 6.2, one can see immidiately that Nusa Tenggara islands are doubly burdened by both poor health and education outcomes.

These findings demonstrate that Indonesia's laudatory income poverty reduction over the last decade has not been complemented by equivalently strong improvements in non-income dimensions. The findings also suggest that different population sub-groups (urban/rural, men/women, island-groups) performed differently in different dimensions of well-being.

We now ask some follow-up questions. Taking those social indicators altogether, how many Indonesians are poor overall? Are urban areas always better off? Which island within the archipelago is the most deprived? What happened to gender and spatial inequalities? The marginal dashboard approach (Ravallion, 2010, 2011) that we have just applied throughout this section is incapable of answering these questions because it only allows us to look at the *marginal distribution* of deprivations (Figure 6.3), while our inquiries demand a characterisation of the *joint distribution* of multiple deprivations (Alkire et al., 2011). In other words, in order to be able to answer these questions, the poverty measure has to take into account the extent of simultaneous deprivation experienced by individuals in the society. The multidimensional poverty measure to be reported next does just that.

Figure 6.3: Hypothetical situation in which marginal method fails to identify the multiply deprived individual (row = individual, column = social indicator)

Society A

$$
g^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow H = \frac{1}{4}
$$

Every individual is deprived on 1 dimension; marginal headcount ratio for each dimension is 0.25.

Society B

$$
g^0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \cdots \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow H = \frac{1}{4}
$$

An individual is deprived on all 4 dimensions; marginal headcount ratio for each dimension is also 0.25.

6.3.2 Multidimensional poverty at the national level

Figure 6.4 presents the trend of multidimensional poverty at the national level. The top-left panel shows that overall poverty (M_0) has declined at an annual rate of 14% over the 2003–2013 period, owing much to the sharp reduction in the proportion of individuals identified as multidimensionally poor (poverty incidence, H), but less to the improvement in the average deprivations experienced by the poor (poverty intensity, A). In 2003, 48% of Indonesian adults were multidimensionally poor and, collectively, they experienced about one-fifth (0.19) of the total possible deprivations that all adults could experience. A decade later, only 11% of adults were poor: the overall poverty figure was just 0.04, indicating a substantial 78% improvement. As in the case of income poverty, reduction in multidimensional poverty was also faster in the first six years ($\delta M_0 = 60\%$) than in the second five ($\delta M_0 = 44\%$).

Despite of a steady 2% annual decline in the contribution to overall poverty, income remains the main contributor to multidimensional poverty (the top-right panel of Figure 6.4). Its contribution to overall poverty in 2013 was still about two times larger than its relative weight (64%). On the other hand, health and education contributed less than their relative weights, suggesting that there were relatively low deprivations in these dimensions. It is noteworthy, however, that as time passes, the contribution of non-income dimensions to overall poverty increases steadily.

The middle panel of Figure 6.4 displays the results of poverty cut-off dominance analysis. On the left, we plot the estimated adjusted headcount ratio (M_0) for the year 2003, 2008 and 2013, along with their analytical 95% confidence intervals, against all possible poverty cut-offs spanning from the union ($k = \frac{1}{6}$) to the intersection ($k = 1$) identification criterion. It turns out that the curves never cross, nor do their confidence

Figure 6.4: Multidimensional poverty at the national level
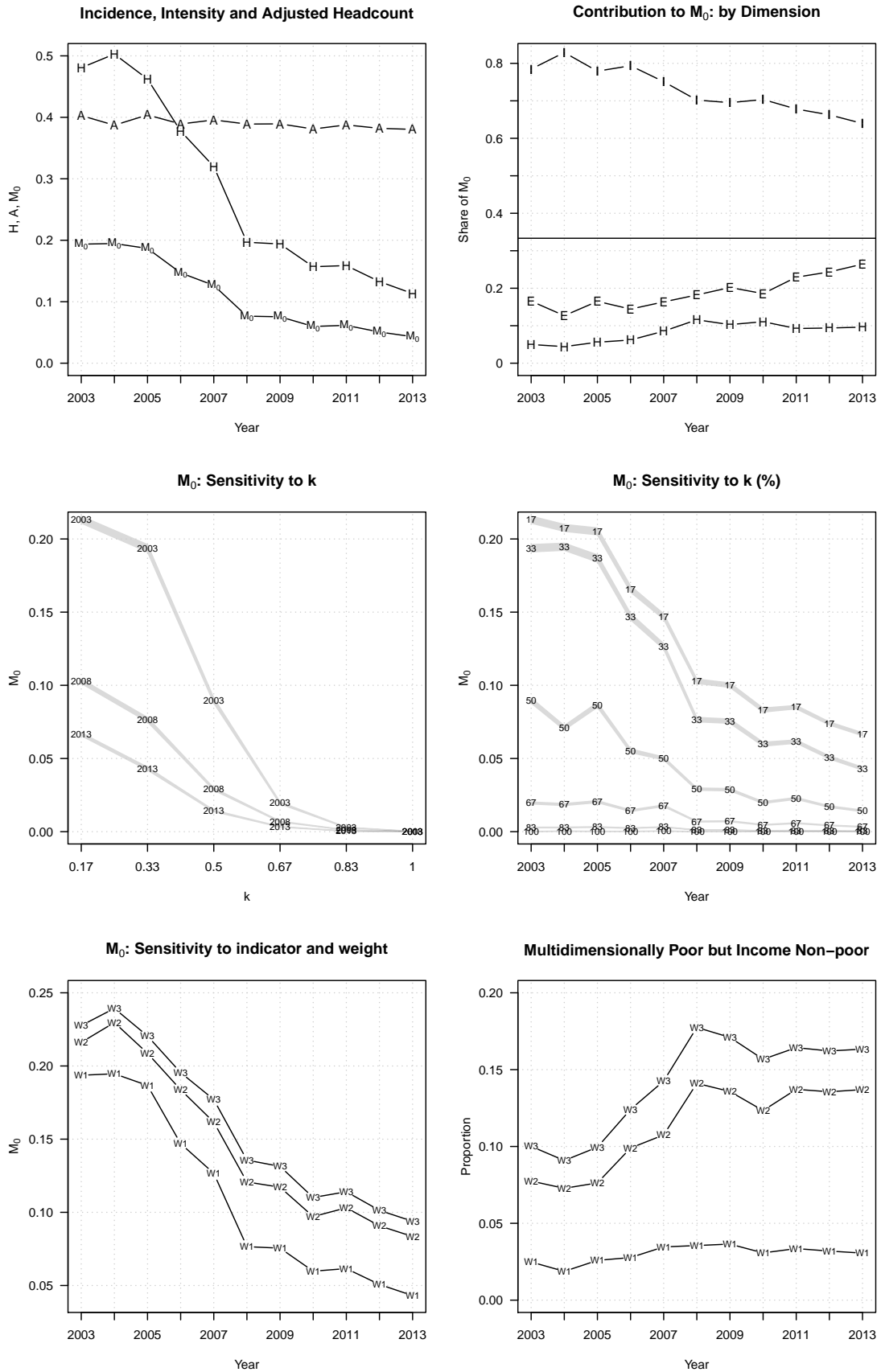
Table 6.3: Correlation matrix of rank orderings across different weights

Ranking pair[†]	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
National:											
$\tau_B(w1, w2)$	1*										
$\tau_B(w1, w3)$	1*										
$\tau_B(w2, w3)$	1*										
Urban/rural:											
$\tau_B(w1, w2)$	1	1	1	1	1	1	1	1	1	1	1
$\tau_B(w1, w3)$	1	1	1	1	1	1	1	1	1	1	1
$\tau_B(w2, w3)$	1	1	1	1	1	1	1	1	1	1	1
Gender:											
$\tau_B(w1, w2)$	1	1	1	1	1	-1	1	1	1	1	1
$\tau_B(w1, w3)$	1	1	1	1	1	-1	1	1	1	1	1
$\tau_B(w2, w3)$	1	1	1	1	1	1	1	1	1	1	1
Island:											
$\tau_B(w1, w2)$	1	0.87	1	0.87	1	0.87	0.73	0.87	0.73	0.73	1
$\tau_B(w1, w3)$	1	1	0.87	0.87	0.87	0.87	0.87	1	0.47	0.73	0.87
$\tau_B(w2, w3)$	1	0.87	0.87	1	0.87	1	0.87	0.87	0.73	1	0.87

[†] Reported are Kendall's τ_B correlation coefficient for pairs of ranking while holding $k = \frac{1}{3}$.
* Correlation coefficient for pairs of year ranking at the national level (2003–2013).

intervals overlap, meaning that there was an unambiguous poverty reduction over the 2003–2013 period. On the right, we present similar analysis, plotting the M_0 estimates for each poverty cut-off against the year. Results suggest that the shape of the trend-line is robust to any poverty cut-off for $k \in \left[\frac{1}{6}, \frac{1}{2}\right]$. Furthermore, the data reveal that even when alternative weights are specified, the shape of the downward-sloping trend (the bottom-left panel of Figure 6.4) as well as the ordering of year ranking (the first five rows of Table 6.3) remain largely unaltered.

In the bottom-right panel of Figure 6.4, we calculate the extent of mismatch when targeting the poor using the conventional measure of income poverty versus using the proposed multidimensional poverty index ($k = \frac{1}{3}$). We found that the mismatch is about 3% for the baseline weight ($w1$), which equals approximately 4.5 million adult Indonesians in 2013. The discrepancy increases significantly as we specified alternative weights that assign more importance to the schooling indicator ($w2$; 7–14%), or to both schooling and illness episode indicators ($w3$; 9–17%). This finding, in combination with that of the steadily increasing contribution of non-income deprivations to overall poverty (the top-right panel of Figure 6.4), underlines the growing relevance of the multidimensional conceptualisation of poverty in Indonesia.

6.3.3 Multidimensional poverty across population subgroups

Having described the trend of multidimensional poverty at the national level, we now decompose the national MPI into contextually relevant subgroups, which, in Indonesia, means measuring poverty by urban/rural, gender and island-group separately.

The goals of this exercise are to understand whether the pattern of poverty reduction has been uniform across subgroups and to identify any especially disadvantaged segment of Indonesian society.

Figure 6.5 displays the trend of urban/rural poverty. In both urban and rural areas, multidimensional poverty has declined significantly over the 11 years of observation (the middle-left panel of Figure 6.5). Progress was faster in rural than in urban areas (δM_0 = 79% versus 74%), resulting in a progressively narrowing rural-to-urban poverty ratio (1.53 in 2003 versus 1.25 in 2013). Two distinct patterns emerge as to how poverty reduction was achieved. As shown in the top panel of Figure 6.5, poverty reduction in rural areas was driven by improvement in *both* poverty incidence (δH = 77%) and intensity (δA = 7.4%) whereas in urban areas, where amelioration in intensity was minuscule (δA = 1.7%), poverty reduction was chiefly attributable to the diminishing proportion of individuals experiencing multiple deprivations (δH = 74%). Results of poverty cut-off dominance analysis (the bottom four plots in Figure 6.5) and a rank correlation test (Table 6.3) suggest that multidimensional poverty was unambiguously higher in rural than in urban areas for *each* year in the 2003–2013 period.

Next, we examine multidimensional poverty by gender. The data reveal that the female-to-male poverty ratio has never oscillated outside the 0.99 (2008) to 1.10 (2013) range (the middle-left panel of Figure 6.6). In fact, the proportion of Indonesian women experiencing simultaneous deprivations over the last decade did not differ much compared to that of their male counterparts (the top-left panel of Figure 6.6). Poverty intensity was slightly higher (1–4%) among women than men during the 2003–2007 period, but since 2008, there has been hardly any difference with regard to the average deprivations experienced by the poor of both genders (the top-right panel of Figure 6.6). A rank correlation test in Table 6.3 indicates that the gender ranking is generally robust to the choice of weight, but the results of poverty cut-off dominance analysis presented in the middle and bottom panels of Figure 6.6 suggest that women are not unambiguously more deprived than men. This result, however, should be interpreted with caution because we still cannot disentangle the precise income (consumption) of men and women that live in the same household using household expenditure data.

We now turn to investigating the trend of regional poverty (Figure 6.7). In general, the last decade saw a substantial poverty reduction in all six island-groups that make up the Indonesian archipelago (the middle-left panel of Figure 6.7). With the exception of Nusa Tenggara islands, improvement was principally attributable to diminishing

Figure 6.5: Multidimensional poverty by urban/rural
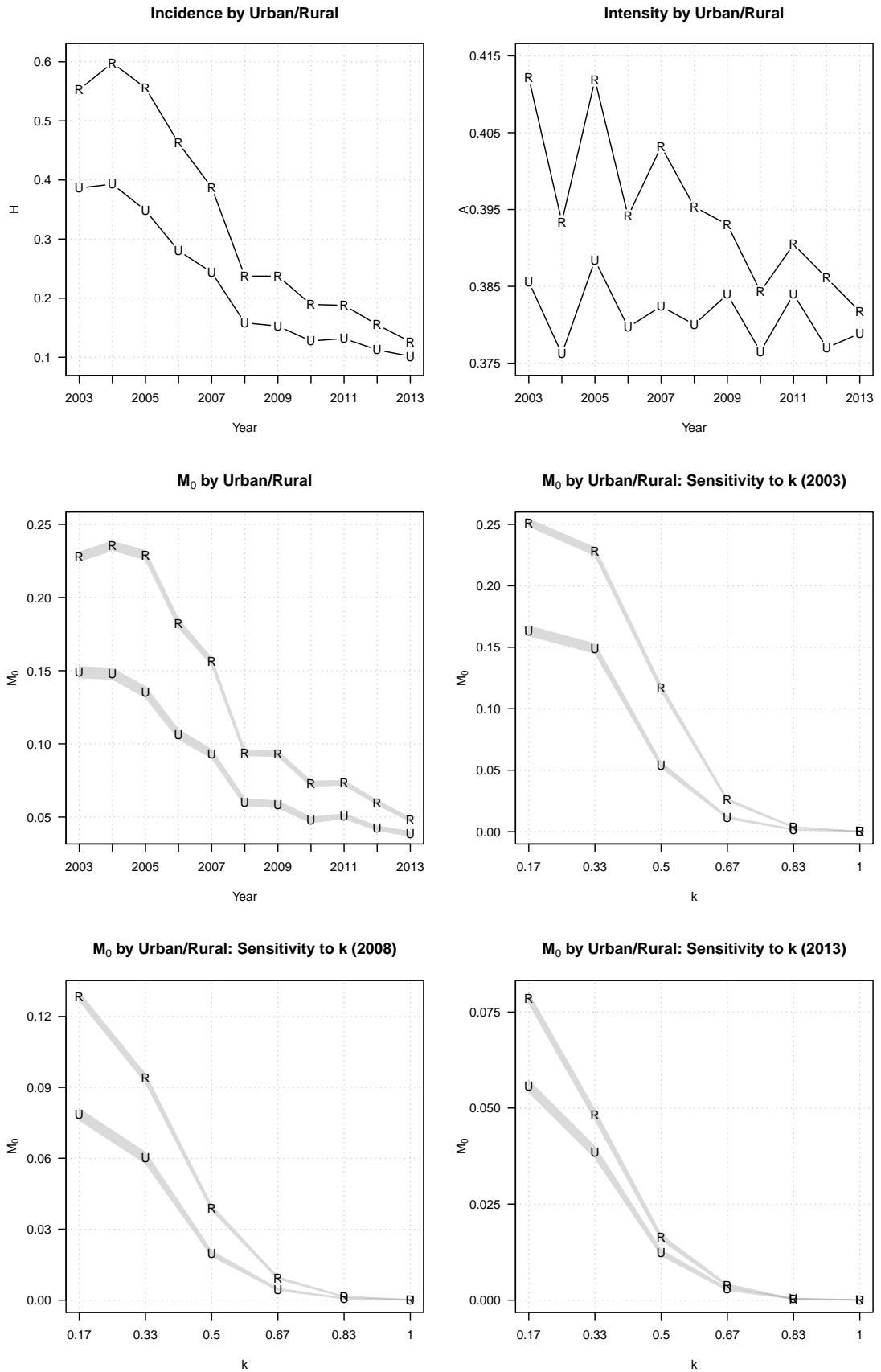
Figure 6.6: Multidimensional poverty by gender
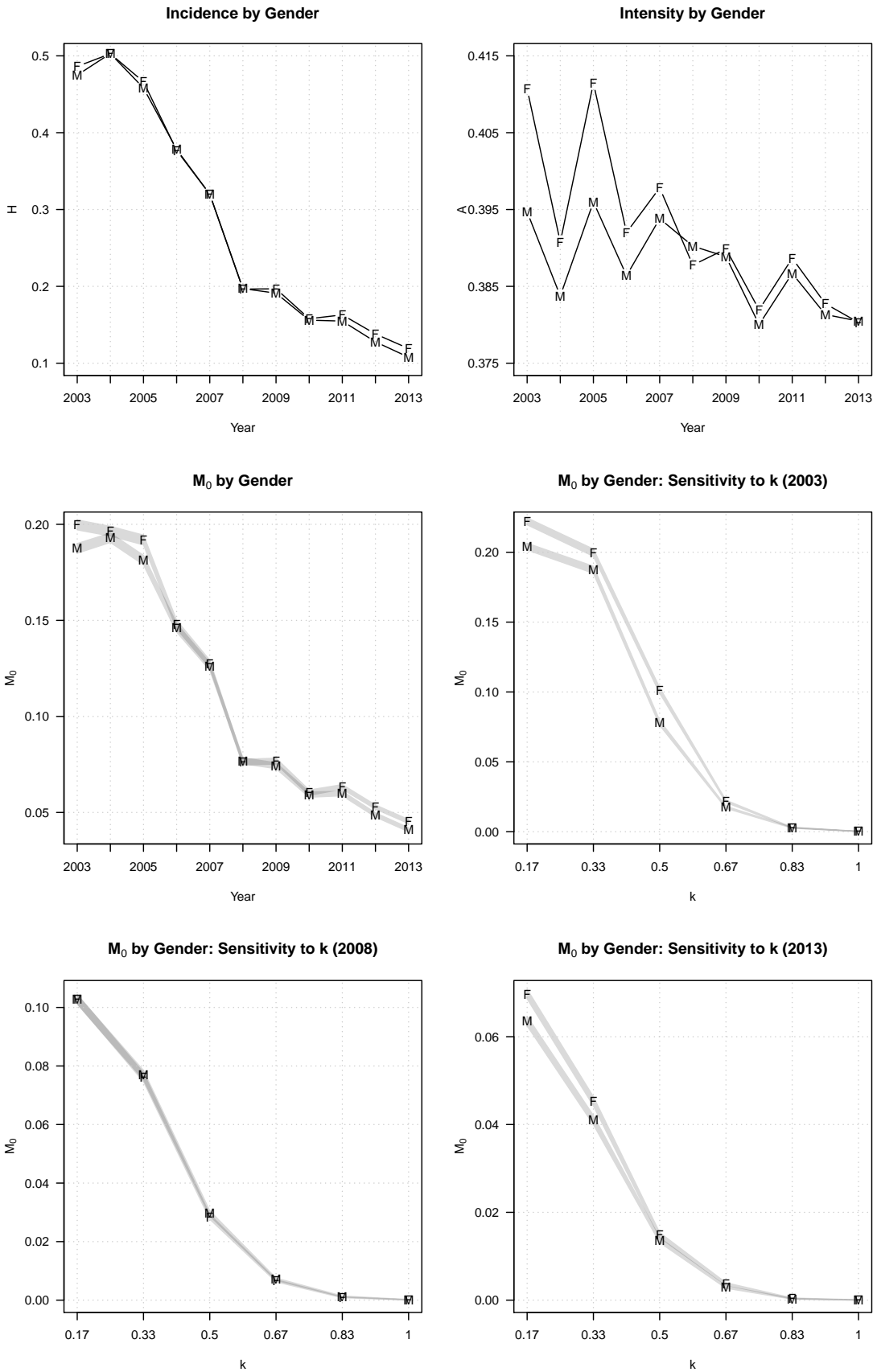
Figure 6.7: Multidimensional poverty by island
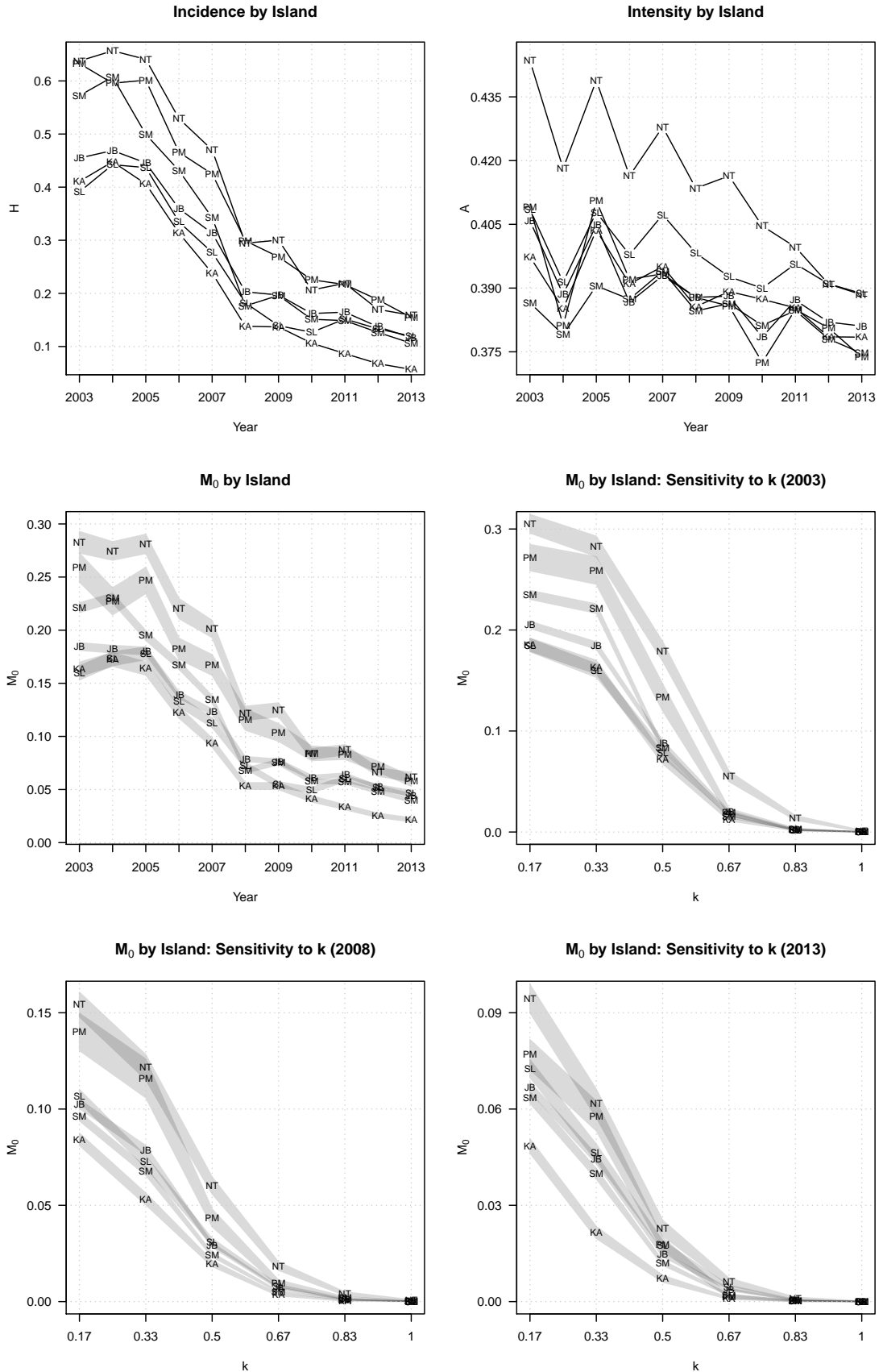
Figure 6.8: Incidence, intensity and adjusted headcount ratio at the district level (2013)
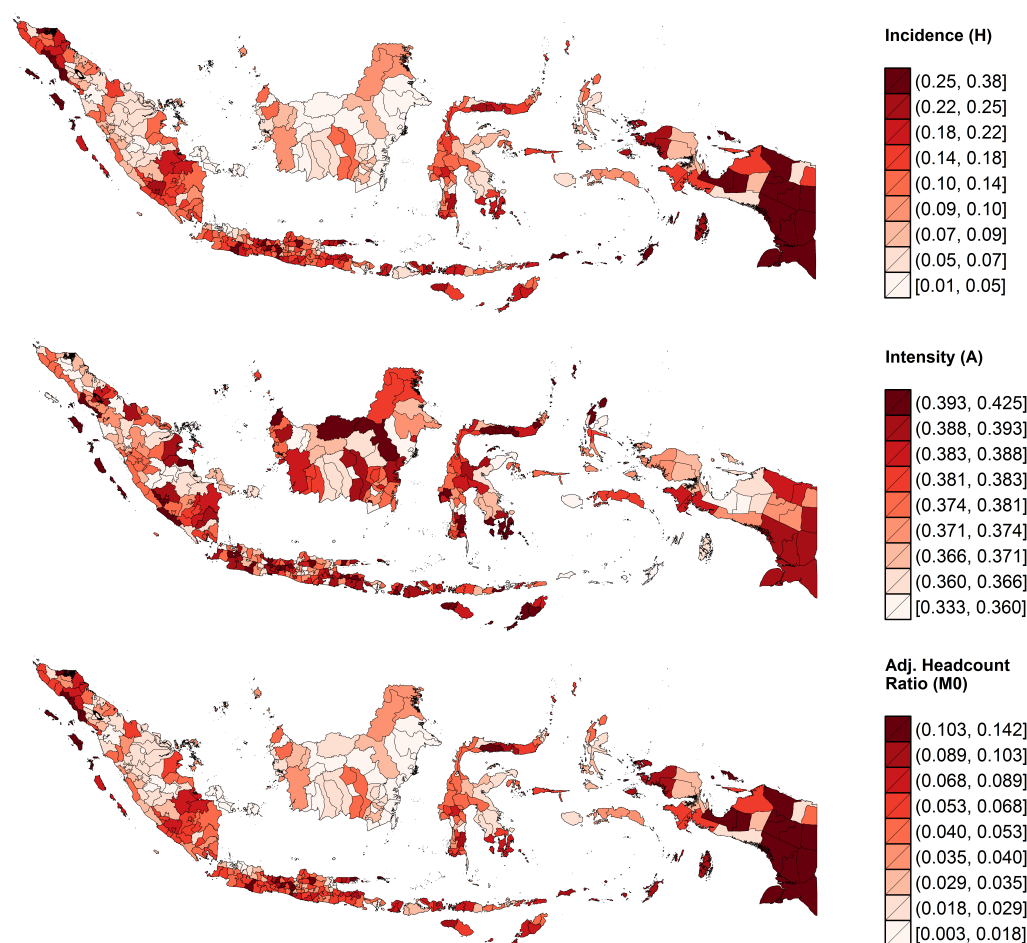
poverty incidence, with only minimal progress in terms of poverty intensity (the top panel of Figure 6.7). Poverty reduction was faster in the 2003–2008 period ($\delta M_0 = 55$–67%) than in 2008–2013 ($\delta M_0 = 36$–60%). The fastest progress was observed in Kalimantan ($\bar{\delta} M_0 = 18\%$), whereas the slowest was in Sulawesi ($\bar{\delta} M_0 = 12\%$).

Robustness tests for the between-island comparison fail to yield clear-cut results. On the one hand, the rank correlation test in Table 6.3 shows that the ranking of regional poverty is relatively robust to the selection of weight. On the other hand, poverty cut-off dominance analysis presented in the middle to the bottom panels of Figure 6.7 reveals that there are only limited dominances between the islands *and* over the 11-year period. Yet, within the limited scope of statistically meaningful comparisons that can be drawn, we can still at least deduce that from 2010 onwards, Papua and Maluku, along with Nusa Tenggara, have always been the poorest islands of the country, whereas Kalimantan is the least poor; between these two extremes are Sumatra, Java, Bali and Sulawesi, whose level of multidimensional poverty has always

Figure 6.9: Within-island variation and districts' incidence and intensity (2013)
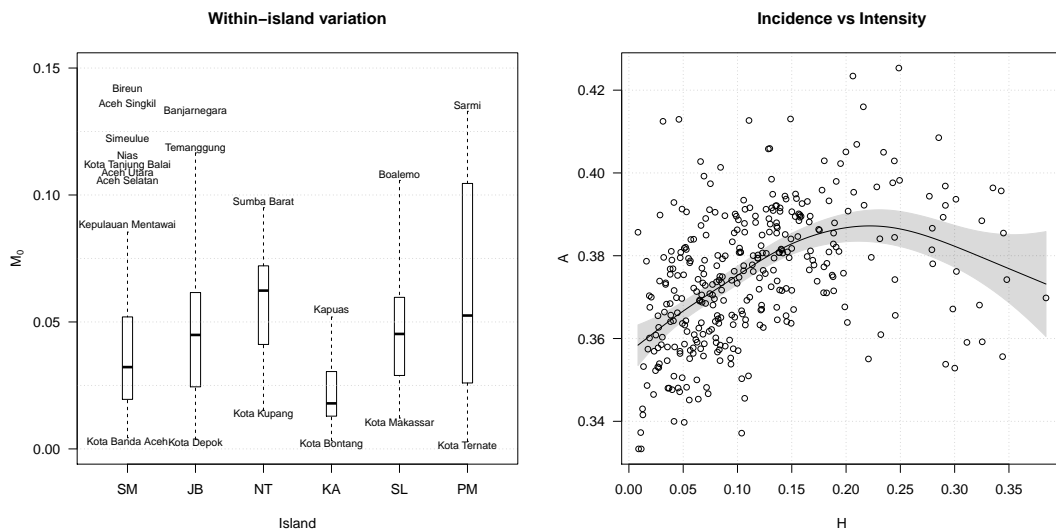
been close to the national average.

Now, what happens if multidimensional poverty is measured at the lowest level of autonomous administrative areas instead of across island-groups? Figure 6.8 presents poverty maps displaying the extent of incidence, intensity, and overall poverty in 346 districts in Indonesia for the year 2013. It appears that the island ranking observed above still holds generally, but such comparison tends to conceal a large amount of variation between the districts on those islands. As we dig deeper (Table 6.4), it turns out that while five out of the ten poorest districts in 2013 are indeed located in Papua, three of them are in Sumatra (Aceh) and two are, perhaps surprisingly, in Central Java. There is unmistakable within-island variation as well (the left panel of Figure 6.9). The island of Java, for example, is paradoxically home to one of the most (Banjarnegara, M_0 = 0.133) *and* the least (Kota Depok, M_0 = 0.004) deprived districts in Indonesia. Furthermore, even within a single province, variation can be immense. It is somewhat disconcerting to note that overall poverty in *Kabupaten* Bangkalan (M_0 = 0.07) can be seven times higher than in *Kota* Surabaya (M_0 = 0.01), even though they belong to the same East Java province and are separated by no more than a 90-minute drive.

The maps in Figure 6.8 further reveal that the spatial patterning of poverty incidence tends not to match that of poverty intensity (Pearson's ρ = 0.40). In contrast to a cross-national pattern reported in the Human Development Report 2010 (UNDP, 2010: 98), the incidence of poverty among districts in Indonesia does not seem to be linearly correlated with its intensity (the right panel of Figure 6.9). Finally, also evident from this district-level analysis is the fact that urban areas tend to dominate

Table 6.4: The top 10 poorest and the least poor districts in 2013

	Incidence			Intensity			Adjusted Headcount Ratio	
Island	District	H	Island	District	A	Island	District	M_0
SM	Bireun	0.384	SL	Boalemo	0.425	SM	Bireuen	0.142
PM	Jayawijaya	0.348	SL	Bantaeng	0.423	SM	Aceh Singkil	0.136
PM	Sarmi	0.345	SL	Gorontalo	0.416	PM	Sarmi	0.133
SM	Simeulue	0.344	SL	Kepulauan Sangihe	0.413	JB	Banjarnegara	0.133
SM	Aceh Singkil	0.343	KA	Kapuas Hulu	0.413	PM	Jayawijaya	0.133
JB	Banjarnegara	0.335	KA	Sambas	0.413	PM	Pegunungan Bintang	0.126
PM	Paniai	0.325	KA	Kota Banjar Baru	0.412	SM	Simeulue	0.122
PM	Pegunungan Bintang	0.325	JB	Blora	0.409	PM	Biak Numfor	0.119
PM	Biak Numfor	0.323	SM	Kepulauan Mentawai	0.407	JB	Temanggung	0.118
SM	Kota Tanjung Balai	0.311	PM	Halmahera Utara	0.406	PM	Paniai	0.117
SM	Karo	0.018	JB	Kota Yogyakarta	0.345	SM	Karo	0.006
JB	Badung	0.017	PM	Kota Sorong	0.345	JB	Sidoarjo	0.006
JB	Sidoarjo	0.016	KA	Kota Samarinda	0.343	JB	Badung	0.006
JB	Kota Denpasar	0.013	SM	Kota Banda Aceh	0.342	JB	Kota Denpasar	0.005
SM	Kota Banda Aceh	0.013	SM	Kota Sabang	0.340	SM	Kota Banda Aceh	0.004
KA	Kota Samarinda	0.013	SM	Kota Pematang Siantar	0.340	KA	Kota Samarinda	0.004
KA	Kota Balikpapan	0.011	KA	Kota Depok	0.337	KA	Kota Balikpapan	0.004
JB	Kota Depok	0.011	JB	Kota Tebing Tinggi	0.337	JB	Kota Depok	0.004
PM	Kota Ternate	0.009	JB	Kota Ternate	0.333	KA	Kota Bontang	0.003
KA	Kota Bontang	0.008	KA	Kota Balikpapan	0.333	PM	Kota Ternate	0.003

rural areas. It is obvious that most of the top 10 least-deprived districts listed in Table 6.4 are municipalities (notice the *Kota* prefix) that, in general, have a higher level of urbanicity than ordinary districts (the *Kabupaten*). This is consistent with the result obtained earlier in Figure 6.5.

6.3.4 Inequality among the poor and across subgroups

Thus far, we have analysed the trend of multidimensional poverty in Indonesia by looking at the overall (M_0) and partial indices (H, A) as well as by decomposing the indices into relevant geographical or social subgroups. We found that there was an unambiguous poverty reduction between 2003 and 2013, both nationally and subnationally. But, with such an improvement, questions of distribution arise. Did the progress benefit the poorest of the poor? Has poverty reduction over the last decade been shared uniformly across population subgroups that make up Indonesian society?

In order to evaluate these, in Table 6.5 we calculate an inequality index (I) using the method proposed by Seth and Alkire (2014). This index is bounded between zero and one, capturing a state of complete equality up to that of complete inequality. It turns out that the reduction of multidimensional poverty in Indonesia within the last 11 years has been accompanied by an amelioration of the distribution

Table 6.5: Measures of inequality

Inequality across ...	$I(2003)$	$I(2013)$	ΔI^*
Individual	0.098	0.076	0.022
Urban/rural	0.006	0.000	0.006
Gender	0.000	0.000	0.000
Island	0.003	0.000	0.003
Province	0.011	0.001	0.010
District	0.023	0.003	0.020

* All changes are statistically significant at 5% level.

of deprivations among the poor. The among-the-poor inequality decreased statistically significantly from 0.098 in 2003 to 0.076 in 2013, indicating inclusive progress. Similarly, the data show that there has been a convergence in poverty, meaning that poorer subgroups improved faster than the less poor. The disparity across subgroups has gone down for all relevant groupings (urban/rural, gender, island, province and district) over the 2003–2013 period. Finally, also evident from Table 6.5 is the fact that, in Indonesia, spatial inequality seems to matter more than gender or urban/rural inequality.

6.4 Discussion and conclusion

Applying the Alkire-Foster method of multidimensional poverty measurement to the National Socio-economic Survey (Susenas) data of Indonesia, this study estimates the extent and investigates the regional as well as the temporal patterns of multidimensional poverty in Indonesia from 2003 to 2013. An Indonesian version of Multidimensional Poverty Index (MPI) is developed through an augmentation of the existing consumption poverty measure with information on health and education that are represented by indicators of illness episode, morbidity, completion of primary school, and literacy.

It is found that, irrespective of the poverty cut-offs or weights specified, there was an unambiguous multidimensional poverty reduction over the last decade at both national and sub-national levels. About half (48%) of Indonesian adults were multidimensionally poor in 2003 and, collectively, they experienced about one-fifth (0.19) of the total possible deprivations that the society could experience. In 2013, the situation was unmistakably better: only one in ten adults (11%) was identified as multidimensionally poor, while the overall poverty figure fell to 0.04 (78% reduction). The data suggest that the rate of poverty reduction was faster in the 2003–2008 period (60%) than in 2008–2013 (44%).

With the exceptions of rural areas and the Nusa Tenggara islands, there was minimal improvement with regard to the average deprivations experienced by the poor (intensity); overall poverty reduction was driven mainly by the decline in poverty incidence. It is further found that, when the overall measure is broken down into its dimensional constituents, income deprivation remains the main contributor to multidimensional poverty (60–70%), albeit with a 2% rate of decrease annually. Also estimated in the national-level analysis is the mismatch between income and multidimensional poverty identification. Results show that approximately 3% of adult Indonesians (4.5 million individuals in 2013) would be classified as non-poor if poverty identification did not take into account deprivations in health and education. This figure could be as high as 7–17% (11–26 million), depending on how much importance is assigned to schooling and/or illness episode indicators.

In an attempt to gain a more complete understanding of joint deprivation, the overall poverty measure is broken down by relevant population sub-groups. The data show that for each year from 2003 to 2013, multidimensional poverty was unambiguously higher in rural than in urban areas, but the gap between them has been progressively

narrowing thanks to substantial improvement in both the incidence *and* the intensity of poverty in rural areas (rural-to-urban poverty ratio was 1.53 in 2003 versus 1.25 in 2013). The data further reveal that Indonesian women are not unambiguously more deprived than men, although they appeared to have slightly more deprivations on average in the 2003–2007 period. Nevertheless, we cannot ascertain whether Indonesia has fared well in terms of gender equality because we cannot disentangle fully the information of women's income using household expenditure data available at present.

In contrast to the clear trend seen in urban/rural and gender decompositions, we found only faint dominance in between-island comparisons over the 11-year period. It is only from 2010 onwards that it can be asserted with statistical confidence that poverty is unambiguously higher in Papua, Maluku and Nusa Tenggara (or lower in Kalimantan) than anywhere else in the archipelago. Even so, it is still important to note that such between-island comparisons mask a substantial amount of within-island and between-district variations, echoing both Ilmma and Wai-Poi (2014) and Sumarto et al. (2014). While five out of the ten poorest districts in 2013 are indeed located in Papua, three of them are in Sumatra and the other two are in Java, neither of which are thought of as places with extreme poverty. Analysis at the district level further reveals that, departing from the pattern observed in a cross-national study (UNDP, 2010: 98), the intensity of poverty among districts in Indonesia does not seem to be related in a linear way to its incidence.

When the distribution of deprivations among the poor is studied, it is found that between 2003 and 2013, there were statistically significant improvements in terms of inequality among the poor and disparity across subgroups. The data show that poorer subgroups progress faster than the less poor, irrespective of the social or geographical groupings considered (converging subgroup poverty level). This finding indicates that the progress achieved within the last 11 years is relatively inclusive, although it should be noted that the between-district inequality within the Indonesian archipelago remains striking.

Overall, these trends are comparable to those obtained from recent consumption poverty evaluations conducted by Ilmma and Wai-Poi (2014) and Sumarto et al. (2014), highlighting the fact that, even a decade after a 'big-bang' decentralisation (Hill, 2014) was initiated, spatial inequity remains a serious challenge for Indonesia. It has been argued that the immense variation in poverty levels across districts reflects heterogeneity in the 'capacity and resources of local governments to develop and implement poverty reduction strategies, and to quickly provide good public services'

(Sumarto et al., 2014: 310). Only competent local government can formulate sound development plans, allocate budgets efficiently, and deliver public services effectively. Therefore, there is plenty of room for local administrators to learn lessons from the top-performing districts (Maharani and Tampubolon, 2014).

While this study has presented a thorough investigation into the state of multidimensional poverty in Indonesia over the last decade, it is inevitably bound by several limitations. Firstly, the present study is unable to include children and adolescents younger than 18 years old in the analysis because information on the relevant dimensions of their well-being (Trani et al., 2013) are not available in Susenas survey. Secondly, the health indicators used in this study (illness episode and morbidity) are weak and by no means comparable to the indicators stipulated in the Millennium Development Goals (malnutrition). Thirdly, with the absence of preference data obtained from large-scale participatory study, the trade-offs between social indicators used in this study are entirely normative. In addition, although the measurement of chronic multidimensional poverty under the Alkire-Foster methodology has recently become feasible (Alkire et al., 2014), this study was unable to make use of it due to the cross-sectional nature of Susenas survey. It is indeed indisputable that future poverty evaluations would benefit from the availability of more comprehensive micro-data.

Even with these limitations, the study still contributes to the literature in at least three ways. First, using nationally representative survey data from Indonesia, the present study shows that the conventional measure of income poverty is not comprehensive. The Indonesian data reveal that income poverty only weakly correlates with deprivations in the domains of health and education, confirming the findings documented in other Asian (Ranis and Stewart, 2012; Santos, 2013; Yu, 2013), African (Batana, 2013; Klasen, 2000), European (Brandolini and D'Alessio, 1998; Whelan et al., 2004) and Latin American (Battison et al., 2013) countries. This may motivate future assessment of multidimensional poverty in other parts of the world.

Second, in using consumption expenditure data as the indicator of income, this study allows the poverty measure to become more sensitive to economic fluctuations than the current version of the international MPI (UNDP, 2010), which uses asset ownership as a proxy for deprivation in living standards. This not only addresses one of the criticisms of the MPI (Ravallion, 2010: 11), but also makes the MPI more comprehensible to Indonesians, who have for decades been accustomed to the conceptualisation of poverty as a consumption shortfall in essential goods and services.

Finally and most importantly, the present study demonstrates the feasibility of adapting the Alkire-Foster methodology to the Indonesian context using an existing official

data source that has been in production since the 1960s (Surbakti, 1995). Because the data are readily available, and the proposed multidimensional poverty measure presented here makes identification of multiply-deprived Indonesians possible, the MPI could nicely complement the existing indices that are routinely reported by the Indonesian Statistical Bureau (BPS). The new measure is suitable as a tool for monitoring the progress of national development, and could also be used as a device for prioritising investment projects or other forms of intervention that are funded by transfers from central to local governments (see Salazar et al., 2013 for a recent proposal in Colombia). With the demonstrated novelty, feasibility and utility of the Alkire-Foster method, policy-makers should now more than ever want to incorporate the idea of poverty as an experience of multiple deprivations into the discourse of national development.

<div align="center">❈ ❈ ❈</div>

Chapter 7

Conclusion

Observations of socio-economic inequalities in health are among the most enduring in social and health sciences research. Empirical studies in developing countries, however, are handicapped by the lack of experimental and longitudinal data. Moreover, the utilisation of cross-sectional data in existing studies remains sub-optimal as geographical information and other valuable material contained in special survey modules that have been collected at enormous cost are often ignored in the standard analyses of these data (Kandala and Ghilagaber, 2014).

Set against this backdrop, the objective of this PhD research is to demonstrate the feasibility and utility of employing advanced analytic techniques to cross-sectional survey data from Indonesia in order to deal with some technical challenges that are often encountered either in the estimation of social gradient in health or in the monitoring and evaluation of well-being as a multidimensional construct.

To this end, five empirical essays in the domain of physical and mental health as well as self-rated health are presented. The first two essays of the thesis (Chapters 2 and 3) study the effect of bias in the empirical estimation of socio-economic gradient in health using observational cross-sectional data. The following two essays (Chapters 4 and 5) demonstrate how to effectively incorporate geographical information into a joint estimation of social and spatial distributions of health so that stakeholders can use the resulting model to inform policy targeting. Finally, the fifth essay (Chapter 6) suggests a way to monitor the impact of policy interventions on the overall level of well-being using a novel multidimensional poverty measurement method that is sensitive to both the incidence and intensity of multiple deprivations.

This final chapter will summarise the major findings of these essays and conclude the thesis.

7.1 Summary

Chapter 2 seeks to estimate the causal effect of poverty on mental health by exploiting seasonal precipitation anomalies as a form of natural experiment that randomly modulates individual income in the Indonesian archipelago. The motivation for this work comes from the idea that the widely documented association between socio-economic status and mental health may be conflated with endogeneity bias that is induced by simultaneous causation, unobserved confounding, or measurement error. In the presence of the former two, it is likely that the observed income gradient in mental health is actually an overestimation of the true gradient. On the contrary, if measurement error is more dominant, then the true gradient may be underestimated. The theoretical consequences of these problems were worked out decades ago in both statistics and econometrics literature, but in the realm of applied empirical research, it is difficult if not impossible to predict the potential bias resulting from their presence *a priori* because these problems might be present simultaneously.

Employing linear and non-linear instrumental variable as well as control function estimation techniques (Cameron and Trivedi, 2005, 2010), Chapter 2 attempts to isolate the exogenous variation of individual income and derive a consistent estimate of income gradient in mental health under the assumed presence of endogeneity. The estimation results suggest that low income does cause poor mental health in Indonesia. The chapter also finds that income gradient in mental health is five times stronger than is conventionally estimated, suggesting that measurement error, rather than confounding or reverse causality, was the main source of bias. Furthermore, as this estimate is robust to varying distributional assumptions, model specifications, estimation techniques and sample stratification, it becomes apparent that income is indeed a strong determinant of mental health, regardless of whether it is treated as an exogenous or an endogenous variable. Recently, the innocuous hypothesis that income influences the mental well-being of people in developing countries has been questioned by a number of economists at the World Bank (Das et al., 2007, 2009). The results obtained in Chapter 2 should therefore put this question to rest. Overall, this chapter contributes to the advancement of the mental health literature in the developing world by shifting the focus of research from the study of association

using small facility or community samples to the study of causal effect using large observational data.

Chapter 3 investigates the extent to which the estimates of demographic and socio-economic inequalities in self-rated health are biased by survey respondents' differential reporting behaviour. This work is motivated by the fact that the use of self-rated health questionnaires in the study of socio-economic inequalities in health could potentially be fraught with problems when individuals have different expectations of what constitutes good health. As Sen (2002) has shown using cases from India and the United States, poor individuals could paradoxically report better health status than their richer counterparts simply because the poor have a much higher tolerance for health problems than the rich. Stated differently, socially disadvantaged individuals might be less able to diagnose and perceive their own illness than privileged ones.

In order to study such interpersonal heterogeneity in reporting style, a parametric anchoring vignette methodology is applied (King et al., 2004). Survey respondents were asked to rate a number of hypothetical scenarios that describe varying levels of health status in six health domains (mobility, pain, cognition, sleep, depression and breathing) using the same ordinal response scale that is applied to the self-report health questionnaire. After obtaining this additional information that can be used to calibrate each respondent's response scale, a compound hierarchical ordered probit model is fitted to obtain health differences by demographic and socio-economic status. The obtained regression coefficients are then compared to those of the standard ordered probit model that assumes no reporting heterogeneity.

The results suggest that there is some evidence for the existence of differential reporting behaviour by socio-economic status, which is proven to be insensitive to the choice of hypothetical scenarios presented to survey respondents. In particular, it is found that Indonesians with more education tend to rate a given health status more negatively than their less-educated counterparts to the extent that had such differentials in health standards not been allowed for in the modelling process, then the salutary effect of education on health would have been severely underestimated. This finding has a significant implications for future research that measures health inequalities in low- and middle-income countries using self-rated health questionnaires. In essence, it warns policy-makers that they cannot rely only on people's perception of their own health when attempting to measure the extent of health inequalities in a society.

From these two chapters, it can be concluded that the threat of bias in the empirical estimation of social gradient in health is indeed a legitimate one. However, arriving at

an unbiased estimate of health inequality by social group is not the only issue that every policy-maker must consider. In practice, the limited resources available for health intervention measures often dictate the need to prioritise resources by geographical area. In other words, as it is known that health and social interventions are most cost-effective when they are delivered to communities rather than to individuals, identifying the areas of greatest need becomes indispensable. The formulation of sound evidence- and need-based intervention measures, therefore, requires understanding not only *who* gets particular diseases but also *where* these diseases strike.

Techniques that allow for such joint estimation of social and spatial distributions of health have been developed by statisticians and epidemiologists over the last two decades. Yet, up until now, empirical applications in public health and social sciences research have remained scarce. Aiming to promote the wider use of these cutting-edge techniques, the third and fourth essays of this thesis demonstrate how policy-makers could simultaneously learn about the extent of social and spatial inequalities in health using variants of generalised linear mixed model (GLMM) under both the frequentist and the Bayesian interpretations (Gelman and Hill, 2007).

Chapter 4 investigates the coexistence and determinants of under- and overnutrition problems (the double burden of malnutrition) in Indonesia using a formal model-based inferential approach. The motivation for this work comes from the fact that while the notion of the double burden of malnutrition is undeniably tied to the concept of population and place, there has been surprisingly little research that explicitly incorporates geographical information into the determination of the coexistence of under- and overweight in a population. The majority of existing studies, as noted by Corsi et al. (2011: 1), have thus far 'been based on the prevalence of these conditions, making their interpretation problematic without an appropriate reference by which to determine the occurrence of a double burden'.

Following the pioneering lead of Subramanian and Smith (2006), Ackerson et al. (2008), and Corsi et al. (2011), Chapter 4 examines the existence and determinants of the double burden of malnutrition in 440 districts in Indonesia using a multilevel multinomial regression technique (Goldstein, 2011). A formal determination of the double burden is achieved by allowing for the possible correlation in the random effects at district level across different contrasts (underweight vs. normal and overweight vs. normal). Areas of greatest need are identified by means of mapping the empirical Bayes modes of the random effects. Stratified analyses by survey respondents' sex and urban/rural location are performed in order to assess the robustness of the results.

Despite its popularity among social and health researchers, the results suggest that there is little support for the double burden hypothesis in Indonesia. Far from finding the existence of double malnutrition, the estimation results instead reveal that under- and overnutrition problems are spatially segregated into non-overlapping clusters in the Indonesian archipelago, irrespective of varying model specification and sample stratification. Further examination of this result suggests that Indonesian policy-makers should view islands in eastern Indonesia, especially Nusa Tenggara and Sulawesi, as the top priority because of their higher relative vulnerability to under- and overnutrition problems, respectively. With regard to determinants, the chapter finds that while education, employment and income do protect Indonesians from undernutrition, they also increase their probability of being overweight. This indicates that undernutrition in Indonesia remains a disease of poverty, while overnutrition is one of the affluent. In light of this finding, it can be learned that increasing the population's material living standards does not appear to constitute an adequate policy measure in promoting healthy nutritional status in countries experiencing rapid economic and epidemiologic transitions.

Chapter 5 demonstrates a more advanced application of a generalised linear mixed modelling framework for the joint estimation of social and spatial inequalities in health. Using a Bayesian hierarchical logistic regression model that accommodates not only the clustering of individuals within administrative units but also the spatial autocorrelation among these locations (Lawson, 2013), this chapter attempts to simultaneously map the social and spatial distributions of malaria in Indonesian Papua. This work is motivated by the fact that while Papua is one of the most malaria-affected regions in the world, there remains little malaria-cartographic activity or population-based epidemiological studies in this region. In the analysis, quantities that are of particular interest to policy-makers, such as the odds-ratio, the probability of excess risk, and the baseline probability of malaria infection, are derived; then, a malaria risk category for each area is determined using the Richardson's criterion (Richardson et al., 2004). In addition, sensitivity analyses with respect to the specification of hyperprior distributions are presented.

The estimation results reveal that even within this hyper-endemic island, the extent of spatial variation in malaria prevalence is not negligible. It is estimated that, holding all other observed individual and environmental risk factors constant, a typical male Papuan infant living in the healthiest 20% of districts would have a 2–5% probability of suffering from malaria, while the figure for his peer in the least-healthy quintile could be as high as 12–25%. The analysis also reveals three clusters of high-risk districts

located in north-central Papua, near the islands of Biak and Yapen, and around the north-western part of Papua. Since the analysis shows that the identification of these clusters is robust to the assumption of hyperprior distributions, these areas should therefore be the top priority for need-based intervention measures. This chapter ultimately shows that even after adjusting for all observable risk factors, the risk of malaria remains far from evenly distributed by income level. This means that in addition to being geography-dependent, malaria in Indonesian Papua is also a disease of poverty. A comprehensive malaria elimination programme in this region should therefore consider not only proximal factors impacting the biology of the parasite and the vector but also distal socio-economic conditions facilitating malaria transmission.

Having studied the effect of bias in masking the true extent of socio-economic inequalities in health (Chapters 2 and 3) and after demonstrating how to guide efficient and equitable allocation of limited resources available for public health intervention using a formal model-based inferential approach (Chapters 4 and 5), the final part of the thesis (Chapter 6) deals with the way to monitor the impact of policy interventions on the overall well-being of the population. As it is known that the effectiveness of public health interventions in low- and middle-income countries is dependent upon the impact of other development efforts, such as education and poverty alleviation, in a web of complex interactions and mutual causation, there emerges a need for 'creating an evaluation paradigm that establishes an index of "poverty" as the primary outcome measure, yet which can be further disaggregated based on the contributions of its defined dimensions such as health, education, and living standard' (Victor et al., 2014: 2). Ideally, this evaluation paradigm should enable policy-makers to summarise the interactions and co-dependencies of relevant dimensions of well-being, while at the same time allowing for the independent assessment of each dimension.

One family of multidimensional poverty measurement method that satisfies such requirements is the Alkire-Foster (AF) method (Alkire and Foster, 2011a; Alkire and Santos, 2013). This novel poverty measurement method combines the classical Foster-Greer-Thorbecke (FGT) measures of unidimensional poverty (Foster et al., 1984) with the counting approach (Atkinson, 2003) that is relatively easy to understand and has a long history in sociology. Unlike other axiomatic methods, the AF method deals with ordinal data in a straightforward manner by dichotomising social achievement into deprived and non-deprived states. Unlike data-driven multivariate and latent-variable approaches to multidimensional poverty measurement, the AF method grants the investigator complete control over all parameters, allowing normative decisions with regard to the selection of indicators, dimensional and poverty cut-offs, and weighting

scheme to be easily incorporated into the analysis. On top of that, another key feature of the AF method is that the resulting aggregate multidimensional poverty index can be meaningfully broken down into its dimensional and geographical constituents.

Analysing National Socio-economic Survey data that have been in production in Indonesia since the 1960s, Chapter 6 evaluates the extent and the patterns of multidimensional poverty in Indonesia from 2003 to 2013. Rather than replacing the existing consumption poverty measure, however, the chapter augments the classical measure with information on health and education. The results suggest that there was an unambiguous multidimensional poverty reduction over the last decade at both national and sub-national levels. It is found that progress has been relatively inclusive across population subgroups, although spatial variation remains notable. Because the new poverty measure is capable of identifying multiply-deprived individuals, it is further suggested that the measure should be used not only to monitor the progress of national development, but also to guide geographical prioritisation of development projects that are funded by transfers from central to local governments. This chapter contributes to the literature by addressing concerns about the unresponsiveness to economic fluctuations of the living standard indicators used in the current version of the UNDP's Multidimensional Poverty Index (MPI) (UNDP, 2010), attributable to its use of consumption expenditure data from Indonesia. The chapter also shows that income poverty is only weakly correlated with deprivations in the domains of health and education, suggesting that the conventional measure of income poverty is not a comprehensive measure of well-being.

7.2   Concluding remarks

To sum up, this thesis has demonstrated ways to deal with some technical challenges that are typically encountered either in the empirical estimation of social gradients in health or in the monitoring and evaluation of well-being in low- and middle-income countries. It does so by bringing recent methodological advances in statistics and econometrics to bear on health inequality research in Indonesia, the world's fourth most populous country, whose health and medical experiences are one of the least-discussed in the global health conversation.

Despite analysing only cross-sectional data, the thesis has shown that the application of advanced analytic techniques to these data could significantly enhance policy-makers' ability to justify intervention measures, prioritise limited resources, and monitor the

impact of public policies on the overall well-being of the population. Of course, these tasks would have been much easier to accomplish had experimental and longitudinal data been available to the investigator. It would be interesting, therefore, for future research to revisit the five studies presented in this thesis using richer datasets, since the methods could flexibly be adapted to experimental or longitudinal settings.

However, given the limited research funding and the politics of data in many developing countries, it is unlikely that such rich data will be widely available in the very near future. In the meantime, applied health and social scientists may want to maximise the depth of insights that can be obtained from existing cross-sectional data by using the latest available analytic techniques, which are capable of delivering greater 'value for money' than standard approaches. This PhD thesis did just that.

❁ ❁ ❁

Bibliography

Ackerson, L. K., Kawachi, I., Barbeau, E. M., Subramanian, S. V., 2008. Geography of underweight and overweight among women in India: A multilevel analysis of 3204 neighborhoods in 26 states. Economics & Human Biology 6 (2), 264–280.

ADB, 2014. Key Indicators for Asia and the Pacific 2014. Asian Development Bank, Manila, Philippines, Ch. Poverty in Asia: A Deeper Look.

Adler, N. E., Ostrove, J. M., 1999. Socioeconomic status and health: What we know and what we don't. Annals of the New York Academy of Sciences 896 (1), 3–15.

Alegana, V. A., Wright, J. A., Nahzat, S. M., Butt, W., Sediqi, A. W., Habib, N., Snow, R. W., Atkinson, P. M., Noor, A. M., 2014. Modelling the incidence of Plasmodium vivax and Plasmodium falciparum malaria in Afghanistan 2006–2009. PLoS ONE 9 (7), e102304.

Alkire, S., 2011. Multidimensional poverty and its discontents. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 46.
URL http://www.ophi.org.uk/multidimensional-poverty-and-its-discontents/

Alkire, S., Apablaza, M., Chakravarty, S. R., Yalonetzky, G., 2014. Measuring chronic multidimensional poverty: A counting approach. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 75.
URL http://www.ophi.org.uk/measuring-chronic-multidimensional-poverty-a-counting-approach/

Alkire, S., Foster, J., 2011a. Counting and multidimensional poverty measurement. Journal of Public Economics 95 (7–8), 476–487.

Alkire, S., Foster, J., 2011b. Understandings and misunderstandings of multidimensional poverty measurement. Journal of Economic Inequality 9 (2), 289–314.

Alkire, S., Foster, J., Santos, M. E., 2011. Where did identification go? Journal of Economic Inequality 9 (3), 501–505.

Alkire, S., Roché, J. M., Summer, A., 2013. Where do the world's multidimensionally poor people live? Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 61.
URL `http://www.ophi.org.uk/where-do-the-worlds-multidimensionally-poor-people-live/`

Alkire, S., Santos, M. E., 2013. A multidimensional approach: Poverty measurement & beyond. Social Indicators Research 112 (2), 239–257.

Alkire, S., Santos, M. E., 2014. Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index. World Development 59, 251–274.

Alkire, S., Vaz, A., 2014. Multidimensional poverty dynamics with time series and panel data. OPHI Summer School on Multidimensional Poverty Analysis, 11–23 August 2014.
URL `http://www.ophi.org.uk/wp-content/uploads/3-SS14-Analysis_Over_Time.pdf`

Allotey, P., Reidpath, D. D., Pokhrel, S., 2010. Social sciences research in neglected tropical diseases 1: the ongoing neglect in the neglected tropical diseases. Health Research Policy and Systems 8, 32.

Alonge, O., Peters, D. H., 2015. Utility and limitations of measures of health inequities: a theoretical perspective. Global Health Action 8, 27591.

Angrist, J. D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, New Jersey.

Araya, R., Lewis, G., Rojas, G., Fritsch, R., 2003. Education and income: which is more important for mental health? Journal of Epidemiology and Community Health 57 (7), 501–505.

Arcaya, M. C., Arcaya, A. L., Subramanian, S. V., 2015. Inequalities in health: definitions, concepts, and theories. Global Health Action 8, 27106.

Atkinson, A. B., 2003. Multidimensional deprivation: Contrasting social welfare and counting approaches. Journal of Economic Inequality 1 (1), 51–65.

Au, N., Lorgelly, P. K., 2014. Anchoring vignettes for health comparisons: an analysis of response consistency. Quality of Life Research 23 (6), 1721–1731.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., van Doorslaer, E., 2011a. Education-related inequity in healthcare with heterogeneous reporting of health. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 639–664.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., van Doorslaer, E., 2011b. Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. Journal of Human Resources 46 (4), 875–906.

Bago d'Uva, T., O'Donnell, O., van Doorslaer, E., 2008a. Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. International Journal of Epidemiology 37 (6), 1375–1383.

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., O'Donnell, O., 2008b. Does reporting heterogeneity bias the measurement of health disparities? Health Economics 17 (3), 351–375.

Baird, J. K., Basri, H., Weina, P., Maguire, J. D., Barcus, M. J., Picarema, H., Elyazar, I. R. F., Ayomi, E., Sekartuti, 2003. Adult Javanese migrants to Indonesian Papua at high risk of severe disease caused by malaria. Epidemiology and Infection 131, 791–797.

Barcus, M. J., Basri, H., Picarima, H., Manyakori, C., Sekartuti, Elyazar, I., Bangs, M. J., Maguire, J. D., Baird, J. K., 2007. Demographic risk factors for severe and fatal vivax and falciparum malaria among hospital admissions in northeastern Indonesian Papua. American Journal of Tropical Medicine and Hygiene 77 (5), 984–991.

Barker, D. J. P., 1997. Maternal nutrition, fetal nutrition, and disease in later life. Nutrition 13 (9), 807–813.

Basáñez, M.-G., Marshall, C., Carabin, H., Gyorkos, T., Joseph, L., 2004. Bayesian statistics for parasitologists. Trends in Parasitology 20 (2), 85–91.

Batana, Y. M., 2013. Multidimensional measurement of poverty among women in Sub-Saharan Africa. Social Indicators Research 112 (2), 337–362.

Battison, D., Cruces, G., Lopez-Calva, L. F., Lugo, M. A., Santos, M. E., 2013. Income and beyond: Multidimensional poverty in six Latin American countries. Social Indicators Research 112 (2), 291–314.

Baum, C. F., Schaffer, M., Stillman, S., 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. Stata Journal 7 (4), 465–506.

Benzeval, M., Bond, L., Campbell, M., Egan, M., Lorenc, T., Petticrew, M., Popham, F., 2014. How does money influence health? Joseph Rowntree Foundation Report. URL `http://www.jrf.org.uk/publications/how-does-money-influence-health`

Berkman, L. F., Kawachi, I., Glymour, M. M. (Eds.), 2014. Social Epidemiology, 2nd Edition. Oxford University Press, New York.

Bernardinelli, L., Clayton, D., Montomoli, C., 1995. Bayesian estimates of disease maps: How important are priors? Statistics in Medicine 12, 2411–2431.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43 (1), 1–20.

Best, N., Richardson, S., Thomson, A., 2005. A comparison of Bayesian spatial models for disease mapping. Statistical Methods in Medical Research 14, 35–59.

Beusenberg, M., Orley, J., 1994. A User's Guide for the Self Reporting Questionnaire (SRQ). Division of Mental Health, World Health Organization. URL `http://apps.who.int/iris/handle/10665/61113`

Bivand, R. S., Gómez-Rubio, V., Rue, H., 2015. Spatial data analysis with R-INLA with some extensions. Journal of Statistical Softwares 63 (20).

Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with R. Use R! Springer, New York.

Björkman-Nyqvist, M., 2013. Income shocks and gender gaps in education: Evidence from Uganda. Journal of Development Economics 105, 237–253.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons, Chichester.

Block, S. A., Kiess, L., Webb, P., Kosen, S., Moench-Pfanner, R., Bloem, M. W., Peter Timmer, C., 2004. Macro shocks and micro outcomes: child nutrition during Indonesia's crisis. Economics & Human Biology 2 (1), 21–44.

Bolt, D. M., Lu, Y., Kim, J.-S., 2014. Measurement and control of response styles using anchoring vignettes: A model-based approach. Psychological Methods 19 (4), 528–541.

Bound, J., Jaeger, D. A., Baker, R. M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. Journal of the American Statistical Association 90 (430), 443–450.

Bourguignon, F., Bénassy-Quéré, A., Dercon, S., Estache, A., Gunning, J. W., Kanbur, R., Klasen, S., Maxwell, S., Platteau, J.-P., Spadaro, A., 2008. Millennium development goals at midpoint: Where do we stand and where do we need to go? European Commission Report.
URL `http://ec.europa.eu/development/icenter/repository/mdg_paper_final_20080916_en.pdf`

Bourguignon, F., Chakravarty, S. R., 2003. The measurement of multidimensional poverty. Journal of Economic Inequality 1 (1), 25–49.

Boutayeb, A., 2006. The double burden of communicable and non-communicable diseases in developing countries. Transactions of the Royal Society of Tropical Medicine and Hygiene 100 (3), 191–199.

BPS, 2010. Sensus Penduduk 2010. Badan Pusat Statistik [online, accessed 18 November 2015].
URL `http://sp2010.bps.go.id/index.php/site/index`

BPS, 2015a. Jumlah dan Persentase Penduduk Miskin, Garis Kemiskinan, Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2) Menurut Provinsi, 2007–2009 (Maret), 2010–2011, 2012 (Maret dan September). Badan Pusat Statistik [online, accessed 27 February 2015].
URL `http://www.bps.go.id/webbeta/frontend/linkTabelStatis/view/id/1489`

BPS, 2015b. Konsep Indeks Pembangunan Manusia. Badan Pusat Statistik [online, accessed 27 February 2015].
URL `http://www.bps.go.id/webbeta/frontend/Subjek/view/id/26#subjekViewTab1`

BPS, 2015c. Konsep Kemiskinan. Badan Pusat Statistik [online, accessed 27 February 2015].
URL `http://www.bps.go.id/webbeta/frontend/Subjek/view/id/23#subjekViewTab1`

BPS, Bappenas, UNDP, 2004. Ekonomi dari Demokrasi: Membiayai Pembangunan Manusia Indonesia. Laporan Pembangunan Manusia Indonesia 2004.

BPS Papua, 2015. Rata-rata Suhu Udara, Kelembaban, Tekanan Udara, Kecepatan Angin, Curah Hujan, Hari Hujan, dan Penyinaran Matahari Menurut Stasiun di Provinsi Papua 2010–2013. Badan Pusat Statistik Provinsi Papua [online, accessed 18 November 2015].
URL `http://papua.bps.go.id/linkTabelStatis/view/id/14`

BPS Papua Barat, 2015. Iklim. Badan Pusat Statistik Provinsi Papua Barat [online, accessed 18 November 2015].
URL `http://irjabar.bps.go.id/Subjek/view/id/151#subjekViewTab3|accordion-daftar-subjek1`

Brandolini, A., D'Alessio, G., 1998. Measuring well-being in the functioning space. Research note.
URL `http://www.iberoamericana.edu.mx/humanismocristiano/seminario_capability/pdf/3.pdf`

Brown, P. J., Konner, M., 1987. An anthropological perspective on obesity. Annals of the New York Academy of Sciences 499 (1), 29–46.

Buckley, J., 2008. Survey context effects in anchoring vignettes. Research note.
URL `http://polmeth.wustl.edu/media/Paper/surveyartifacts.pdf`

Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Cameron, A. C., Trivedi, P. K., 2010. Microeconometrics Using Stata, revised Edition. Stata Press, College Station, Texas.

Cawley, J., Meyerhoefer, C., 2012. The medical care costs of obesity: An instrumental variables approach. Journal of Health Economics 31 (1), 219–230.

CDC, 2010. Where malaria occurs. Centers for Disease Control and Prevention [online, accessed 18 November 2015].
URL `http://www.cdc.gov/malaria/about/distribution.html`

CDC, 2015. Ecology of malaria. Centers for Disease Control and Prevention [online, accessed 18 November 2015].
URL `http://www.cdc.gov/malaria/about/biology/ecology.html`

Chadwick, E., 1842. Report on the Sanitary Condition of the Labouring population of Great Britain. Poor Law Commission, London.

Chen, S., Zhao, G., Li, L., Wang, Y., Chiu, H., Caine, E., 2009. Psychometric properties of the Chinese version of the Self-Reporting Questionnaire 20 (SRQ-20) in community settings. International Journal of Social Psychiatry 55 (6), 538–547.

Chin, B., 2010. Income, health, and well-being in rural Malawi. Demographic Research 23 (35), 997–1030.

Chirombo, J., Lowe, R., Kazembe, L., 2014. Using structured additive regression models to estimate risk factors of malaria: Analysis of 2010 Malawi Malaria Indicator Survey data. PLoS ONE 9 (7), e101116.

Clements, A. C. A., Lwambo, N. J. S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A., Brooker, S., 2006. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. Tropical Medicine & International Health 11 (4), 490–503.

Cockerham, W. C., 2013. Social Causes of Health and Disease, 2nd Edition. Polity Press, Cambridge.

Cohen, B., 2006. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. Technology in Society 28 (1–2), 63–80.

Cooper, S., Lund, C., Kakuma, R., 2012. The measurement of poverty in psychiatric epidemiology in LMICs: critical review and recommendations. Social Psychiatry and Psychiatric Epidemiology 47 (9), 1499–1516.

Corsi, D. J., Finlay, J. E., Subramanian, S. V., 2011. Global burden of double malnutrition: Has anyone seen it? PLoS ONE 6 (9), e25120.

Cutler, D. M., Katz, L. F., 1992. Rising inequality? Changes in the distribution of income and consumption in the 1980's. American Economic Review 82 (2), 546–551.

Das, J., Do, Q.-T., Friedman, J., McKenzie, D., 2009. Mental health patterns and consequences: Results from survey data in five developing countries. World Bank Economic Review 23 (1), 31–55.

Das, J., Do, Q.-T., Friedman, J., McKenzie, D., Scott, K., 2007. Mental health and poverty in developing countries: Revisiting the relationship. Social Science & Medicine 65 (3), 467–480.

Datta Gupta, N., Kristensen, N., Pozzoli, D., 2010. External validation of the use of vignettes in cross-country health studies. Economic Modelling 27 (4), 854–865.

de Castro, M. C., Fisher, M. G., 2012. Is malaria illness among young children a cause or a consequence of low socioeconomic status? evidence from the United Republic of Tanzania. Malaria Journal 11, 161.

Deaton, A., Zaidi, S., 2002. Guidelines for constructing consumption aggregates for welfare analysis. Living Standards Measurement Study Working Paper No. 135, World Bank.
URL https://openknowledge.worldbank.org/handle/10986/14101

Decancq, K., Lugo, M. A., 2013. Weights in multidimensional indices of wellbeing: An overview. Econometric Reviews 32 (1), 7–34.

Delisle, H. F., 2008. Poverty: The double burden of malnutrition in mothers and the intergenerational impact. Annals of the New York Academy of Sciences 1136 (1), 172–184.

Doak, C. M., Adair, L. S., Bentley, M., Monteiro, C., Popkin, B. M., 2005. The dual burden household and the nutrition transition paradox. International Journal of Obesity 29 (1), 129–136.

Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B. G., Skodol, A. E., Stueve, A., 1992. Socioeconomic status and psychiatric disorders: the causation-selection issue. Science 255 (5047), 946–952.

Douglas, N. M., Pontororing, G. J., Lampah, D. A., Yeo, T. W., Kenangalem, E., Poespoprodjo, J. R., Ralph, A. P., Bangs, M. J., Sugiarto, P., Anstey, N. M., Price, R. N., 2014. Mortality attributable to Plasmodium vivax malaria: a clinical audit from Papua, Indonesia. BMC Medicine 12, 217.

Dzator, J., 2013. Hard times and common mental health disorders in developing countries: Insights from urban Ghana. Journal of Behavioral Health Services & Research 40 (1), 71–87.

Eberly, L. E., Carlin, B. P., 2000. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Statistics in Medicine 19, 2279–2294.

Ecob, R., Smith, G. D., 1999. Income and health: what is the nature of the relationship? Social Science & Medicine 48 (5), 693–705.

Elyazar, I. R. F., Gething, P. W., Patil, A. P., Rogayah, H., Kusriastuti, R., Wismarini, D. M., Tarmizi, S. N., Baird, J. K., Hay, S. I., 2011a. Plasmodium falciparum malaria endemicity in Indonesia in 2010. PLoS ONE 6 (6), e21315.

Elyazar, I. R. F., Gething, P. W., Patil, A. P., Rogayah, H., Sariwati, E., Palupi, N. W., Tarmizi, S. N., Kusriastuti, R., Baird, J. K., Hay, S. I., 2012. Plasmodium vivax malaria endemicity in Indonesia in 2010. PLoS ONE 7 (5), e37325.

Elyazar, I. R. F., Hay, S. I., Baird, J. K., 2011b. Malaria distribution, prevalence, drug resistance and control in Indonesia. Advances in Parasitology 74, 41–175.

Ernst, K. C., Lindblade, K. A., Koech, D., Sumba, P. O., Kuwuor, D. O., John, C. C., Wilson, M. L., 2009. Environmental, socio-demographic and behavioural determinants of malaria risk in the western Kenyan highlands: a case-control study. Tropical medicine & International Health 14 (10), 1258–1265.

Ettner, S. L., 1996. New evidence on the relationship between income and health. Journal of Health Economics 15 (1), 67–85.

Faden, R. R., Powers, M., 2008. Health inequities and social justice: The moral foundations of public health. Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 51 (2), 151–157.

Ferraro, K. F., 1980. Self-ratings of health among the old and the old-old. Journal of Health and Social Behavior 21 (4), 377–383.

Filho, A. D. P. C., Kawachi, I., Wang, Y. P., Viana, M. C., Andrade, L. H. S. G., 2013. Does income inequality get under the skin? A multilevel analysis of depression, anxiety and mental disorders in São Paulo, Brazil. Journal of Epidemiology and Community Health 67 (11), 966–972.

Filmer, D., Pritchett, L. H., 2001. Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. Demography 38 (1), 115–132.

Firpo, S., Fortin, N. M., Lemieux, T., 2009. Unconditional quantile regressions. Econometrica 77 (3), 953–973.

Foster, J., 2007. A report on Mexican multidimensional poverty measurement. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 40.
URL `http://www.ophi.org.uk/a-report-on-mexican-multidimensional-poverty-measurement/`

Foster, J., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 52 (3), 761–766.

Freedman, D., 2001. Ecological Inference. In: Smelser, N. J., Baltes, P. B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Pergamon, Oxford, pp. 4027–4030.

Freedman, D., 2009. Statistical Models: Theory and Practice, Revised Edition. Cambridge University Press, New York.

Freedman, D., 2010. Statistical Models and Causal Inference: A Dialogue with the Social Sciences. Cambridge University Press, New York.

Frijters, P., Haisken-DeNew, J. P., Shields, M. A., 2005. The causal effect of income on health: Evidence from German reunification. Journal of Health Economics 24 (5), 997–1017.

Frongillo, E. A., Bégin, F., 1993. Gender bias in food intake favors male preschool Guatemalan children. Journal of Nutrition 123 (2), 189–196.

Gallup, J. L., Sachs, J. D., 2001. The economic burden of malaria. American Journal of Tropical Medicine and Hygiene 64 (1S), 85–96.

Ganihartono, I., 1996. Psychiatric morbidity among patients attending the Bangetayu community health centre in Indonesia. Buletin Penelitian Kesehatan 24 (4), 42–51.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Analytical Methods for Social Research. Cambridge University Press, New York.

Geraci, M., 2014. Linear quantile mixed models: The `lqmm` package for laplace quantile regression. Journal of Statistical Software 57 (13).

Geraci, M., Bottai, M., 2014. Linear quantile mixed models. Statistics and Computing 24 (3), 461–479.

Ghubash, R., Daradkeh, T., El-Rufaie, O. F., Abou-Saleh, M. T., 2001. A comparison of the validity of two psychiatric screening questionnaires: the Arabic General Health Questionnaire (AGHQ) and Self-Reporting Questionnaire (SRQ-20) in UAE, using receiver operating characteristic (ROC) analysis. European Psychiatry 16 (2), 122–126.

Giang, K. B., Allebeck, P., Kullgren, G., van Tuan, N., 2006. The Vietnamese version of the Self Reporting Questionnaire 20 (SRQ-20) in detecting mental disorders in rural Vietnam: A validation study. International Journal of Social Psychiatry 52 (2), 175–184.

Gillespie, S., Haddad, L. J., 2003. The Double Burden of Malnutrition in Asia: Causes, Consequences, and Solutions. SAGE, New Delhi.

Glewwe, P., King, E. M., 2001. The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter? World Bank Economic Review 15 (1), 81–113.

Glymour, M. M., Avendano, M., Kawachi, I., 2014. Socioeconomic Status and Health. In: Berkman, L. F., Kawachi, I., Glymour, M. M. (Eds.), Social Epidemiology, 2nd Edition. Oxford University Press, New York, pp. 17–62.

Goldberg, E. M., Morrison, S. L., 1963. Schizophrenia and social class. British Journal of Psychiatry 109 (463), 785–802.

Goldstein, H., 2011. Multilevel Statistical Models, 4th Edition. Wiley Series in Probability and Statistics. John Wiley & Sons, Sussex.

Gould, W., 2011. Use poisson rather than regress; tell a friend. Stata Blog, 22 August 2011 [online, accessed 29 October 2013].
URL `http://blog.stata.com/2011/08/22/use-poisson-rather-than-regress-tell-a-friend/`

Graunt, J., 1662. Natural and political observations mentioned in a following index, and made upon the bills of mortality. [online, accessed 9 May 2016].
URL `http://www.edstephan.org/Graunt/bills.html`

Greene, W. A., Hensher, D. A., 2010. Modeling Ordered Choices: A Primer. Cambridge University Press, Cambridge, UK.

Greenland, S., 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. International Journal of Epidemiology 30 (6), 1343–1350.

Grilli, L., Rampichini, C., 2006. A multilevel multinomial logit model for the analysis of graduates' skills. Statistical Methods and Applications 16 (3), 381–393.

Groce, N., Kembhavi, G., Wirz, S., Lang, R., Trani, J.-F., Kett, M., 2011. Poverty and disability: A critical review of the literature in low and middle-income countries.

Leonard Cheshire Disability and Inclusive Development Centre Working Paper Series No. 16 [online, accessed 9 May 2016].
URL `http://www.ucl.ac.uk/lc-ccr/centrepublications/workingpapers/WP16_Poverty_and_Disability_review.pdf`

Groot, W., 2000. Adaptation and scale of reference bias in self-assessments of quality of life. Journal of Health Economics 19 (3), 403–420.

Grossman, M., 1972. On the concept of health capital and the demand for health. Journal of Political Economy 80 (2), 223–255.

Grundy, E., Holt, G., 2001. The socioeconomic status of older adults: How should we measure it in studies of health inequalities? Journal of Epidemiology and Community Health 55 (12), 895–904.

Gurrici, S., Hartriyanti, Y., Hautvast, J. G. A. J., Deurenberg, P., 1998. Relationship between body fat and body mass index: differences between Indonesians and Dutch Caucasians. European Journal of Clinical Nutrition 52 (11), 779–783.

Guthmann, J. P., Llanos-Cuentas, A., Palacios, A., Hal, A. J., 2002. Environmental factors as determinants of malaria risk. A descriptive study on the northern coast of Peru. Tropical Medicine & International Health 7 (6), 518–525.

Guyatt, H. L., Snow, R. W., 2001. The epidemiology and burden of Plasmodium falciparum-related anemia among pregnant women in sub-Saharan Africa. American Journal of Tropical Medicine and Hygiene 64 (1S), 36–44.

Ha, D. T. P., Feskens, E. J. M., Deurenberg, P., Mai, L. B., Khan, N. C., Kok, F. J., 2011. Nationwide shifts in the double burden of overweight and underweight in Vietnamese adults in 2000 and 2005: two national nutrition surveys. BMC Public Health 11 (1), 62.

Hamad, R., Fernald, L. C. H., Karlan, D. S., Zinman, J., 2008. Social and economic correlates of depressive symptoms and perceived stress in South African adults. Journal of Epidemiology and Community Health 62 (6), 538–544.

Hanandita, W., Tampubolon, G., 2016a. Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. Quality of Life Research 25 (5), 1137–1149.
URL `http://dx.doi.org/10.1007/s11136-015-1152-y`

Hanandita, W., Tampubolon, G., 2016b. Multidimensional poverty in Indonesia: Trend over the last decade (2003–2013). Social Indicators Research 128 (2), 559–587.
URL `http://dx.doi.org/10.1007/s11205-015-1044-0`

Haque, U., Soares Magalhães, R. J., Mitra, D., Kolivras, K. N., Schmidt, W.-P., Haque, R., Glass, G. E., 2011. The role of age, ethnicity and environmental factors in modulating malaria risk in Rajasthali, Bangladesh. Malaria Journal 10, 367.

Haque, U., Soares Magalhães, R. J., Reid, H. L., Clements, A. C. A., Ahmed, S. M., Islam, A., Yamamoto, T., Haque, R., Glass, G. E., 2010. Spatial prediction of malaria prevalence in an endemic area of Bangladesh. Malaria Journal 9, 120.

Harding, T. W., De Arango, V., Baltazar, J., Climent, C. E., Ibrahim, H. H. A., Ladrido-Ignacio, L., Wig, N. N., 1980. Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. Psychological Medicine 10 (2), 231–241.

Hensher, D. A., Rose, J. M., Greene, W. H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press, New York.

Hernán, M. A., Robins, J. M., 2006. Instruments for causal inference: An epidemiologist's dream? Epidemiology 17 (4), 360–372.

Hickson, M., 2006. Malnutrition and ageing. Postgraduate Medical Journal 82 (963), 2–8.

Hill, H. (Ed.), 2014. Regional Dynamics in a Decentralized Indonesia. ISEAS Publishing, Singapore.

Hirve, S., Gómez-Olivé, X., Oti, S., Debpuur, C., Juvekar, S., Tollman, S., Blomstedt, Y., Wall, S., Ng, N., 2013. Use of anchoring vignettes to evaluate health reporting behavior amongst adults aged 50 years and above in Africa and Asia - testing assumptions. Global Health Action 6, 21064.

Holding, P. A., Snow, R. W., 2001. Impact of Plasmodium falciparum malaria on performance and learning: review of the evidence. American Journal of Tropical Medicine and Hygiene 64 (1S), 68–75.

Hopkins, D. J., King, G., 2010. Improving anchoring vignettes designing surveys to correct interpersonal incomparability. Public Opinion Quarterly 74 (2), 201–222.

Horton, R., 2016. Offline: Indonesia—unravelling the mystery of a nation. Lancet 387 (10021), 830.

Hosseinpoor, A. R., Bergen, N., Kostanjsek, N., Kowal, P., Officer, A., Chatterji, S., 2015. Socio-demographic patterns of disability among older adult populations of low-income and middle-income countries: results from World Health Survey. International Journal of Public Health (Online First), 1–9.

Howe, L. D., Galobardes, B., Matijasevich, A., Gordon, D., Johnston, D., Onwu-jekwe, O., Patel, R., Webb, E. A., Lawlor, D. A., Hargreaves, J. R., 2012. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: a methods of measurement in epidemiology paper. International Journal of Epidemiology 41 (3), 871–886.

Howe, L. D., Hargreaves, J. R., Huttly, S. R. A., 2008. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. Emerging Themes in Epidemiology 5, 3.

Huijts, T., Perkins, J. M., Subramanian, S. V., 2010. Political regimes, political ideology, and self-rated health in Europe: A multilevel analysis. PLoS ONE 5 (7), e11711.

ICF International, 2012. Demographic and health surveys. [online, accessed 27 February 2015].
URL http://dhsprogram.com/data/

Idler, E. L., Benyamini, Y., 1997. Self-rated health and mortality: A review of twenty-seven community studies. Journal of Health and Social Behavior 38 (1), 21–37.

IHME, 2013. Global burden of disease data visualizations. Institute for Health Metrics and Evaluation [online, accessed 14 November 2013].
URL   http://www.healthmetricsandevaluation.org/tools/data-visualizations

Ilmma, A., Wai-Poi, M., 2014. Patterns of regional poverty in the new Indonesia. In: Hill, H. (Ed.), Regional Dynamics in a Decentralized Indonesia. ISEAS Publishing, Singapore, pp. 98–132.

Imbens, G. W., Wooldridge, J. M., 2007. What's new in econometrics - lecture notes. NBER mini-course in Cambridge, Massachussetts, July 31–August 1, 2007.
URL http://www.nber.org/WNE/WNEnotes.pdf

Ingstad, B., Munthali, A. C., Braathen, S. H., Grut, L., 2012. The evil circle of poverty: a qualitative study of malaria and disability. Malaria Journal 11, 15.

Irmansyah, I., Dharmono, S., Maramis, A., Minas, H., 2010. Determinants of psychological morbidity in survivors of the earthquake and tsunami in Aceh and Nias. International Journal of Mental Health Systems 4, 8.

Jehn, M., Brewis, A., 2009. Paradoxical malnutrition in mother–child pairs: Untangling the phenomenon of over- and under-nutrition in underdeveloped economies. Economics & Human Biology 7 (1), 28–35.

Jolliffe, D., 2011. Overweight and poor? On the relationship between income and the body mass index. Economics & Human Biology 9 (4), 342–355.

Jones, A. M., Rice, N., Bago d'Uva, T., Balia, S., 2013. Applied Health Economics, 2nd Edition. Routledge, Oxon, UK.

Jones, K., n.d. Do multilevel models ever give different results? Multilevel models: An introduction and FAQs, Centre for Multilevel Modelling, University of Bristol [online, accessed 18 November 2015].
URL http://www.bristol.ac.uk/cmm/learning/multilevel-models/

Kandala, N.-B., Fahrmeir, L., Klasen, S., Priebe, J., 2009. Geo-additive models of childhood undernutrition in three sub-Saharan African countries. Population, Space and Place 15 (5), 461–473.

Kandala, N.-B., Ghilagaber, G. (Eds.), 2014. Advanced Techniques for Modelling Maternal and Child Health in Africa. Springer, Dordrecht.

Kapteyn, A., Smith, J. P., van Soest, A., 2007. Vignettes and self-reports of work disability in the United States and the Netherlands. American Economic Review 97 (1), 461–473.

Karyana, M., Burdarm, L., Yeung, S., Kenangalem, E., Wariker, N., Maristela, R., Umana, K. G., Vemuri, R., Okoseray, M. J., Penttinen, P. M., Ebsworth, P., Sugiarto, P., Anstey, N. M., Tjitra, E., Price, R. N., 2008. Malaria morbidity in Papua Indonesia, an area with multidrug resistant Plasmodium vivax and Plasmodium falciparum. Malaria Journal 7, 148.

Kauermann, G., Haupt, H., Kaufmann, N., 2012. A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. Statistical Modelling 12 (5), 419–440.

Kazembe, L. N., 2007. Spatial modelling and risk factors of malaria incidence in northern Malawi. Acta Tropica 102 (2), 126–137.

Kazembe, L. N., 2013. An additive regression model for investigating the relationship between childhood health and socio-economic status. Spatial and Spatio-temporal Epidemiology 6, 71–84.

Kemenkes, 2008. Laporan Nasional Riskesdas 2007. Ministry of Health, The Republic of Indonesia [online, accessed 22 January 2014].
URL   http://labdata.litbang.depkes.go.id/menu-download/menu-download-laporan

Kemenkes, 2009. Keputusan Menteri Kesehatan Republik Indonesia Nomor 293/MENKES/SK/IV/2009 Tentang Eliminasi Malaria Di Indonesia. Ministry of Health, The Republic of Indonesia.

Kemenkes, 2013. Laporan Nasional Riskesdas 2013. Ministry of Health, The Republic of Indonesia [online, accessed 18 November 2015].
URL   http://labdata.litbang.depkes.go.id/menu-download/menu-download-laporan

King, G., Murray, C. J. L., Solomon, J. A., Tandon, A., 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. American Political Science Review 98 (1), 191–207.

King, G., Wand, J., 2007. Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. Political Analysis 15 (1), 46–66.

Kishore, K., Subbiah, A. R., Sribimawati, T., Diharto, S., Alimoeso, S., Rogers, P., Setiana, A., 2000. Indonesia Country Study. Asian Disaster Preparedness Center.
URL   http://archive.unu.edu/env/govern/ElNIno/CountryReports/pdf/indonesia.pdf

Klasen, S., 2000. Measuring poverty and deprivation in South Africa. Review of Income and Wealth 46 (1), 33–58.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, New York.

Koenker, R., Hallock, K. F., 2001. Quantile regression. Journal of Economic Perspectives 15 (4), 143–156.

Kruschke, J. K., 2010. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press, Oxford.

Kubzansky, L. D., Seeman, T. E., Glymour, M. M., 2014. Biological Pathways Linking Social Conditions and Health: Plausible Mechanisms and Emerging Puzzles. In:

Berkman, L. F., Kawachi, I., Glymour, M. M. (Eds.), Social Epidemiology, 2nd Edition. Oxford University Press, New York, pp. 512–561.

Lancet, 2014. Indonesia strides towards universal health care. Lancet 383 (9911), 2.

LaVeist, T. A., Gaskin, D., Richard, P., 2011. Estimating the economic burden of racial health inequalities in the United States. International Journal of Health Services 41 (2), 231–238.

Lawson, A. B., 2013. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, 2nd Edition. Interdisciplinary Statistics Series. CRC Press, Boca Raton, FL.

Lee, J., Houser, R. F., Must, A., de Fulladolsa, P., Bermudez, O. I., 2012. Socioeconomic disparities and the familial coexistence of child stunting and maternal overweight in Guatemala. Economics & Human Biology 10 (3), 232–241.

Leonhardt, D., 3/23 2010. In Health Bill, Obama Attacks Wealth Inequality. The New York Times.
URL http://www.nytimes.com/2010/03/24/business/24leonhardt.html

Levine, D., Yang, D., 2006. A note on the impact of local rainfall on rice output in Indonesian districts. Research note.
URL http://www-personal.umich.edu/~deanyang/papers/levineyang_ricerain.pdf

Lindeboom, M., van Doorslaer, E., 2004. Cut-point shift and index shift in self-reported health. Journal of Health Economics 23 (6), 1083–1099.

Link, B. G., Phelan, J., 1995. Social conditions as fundamental causes of disease. Journal of Health and Social Behavior Extra Issue, 80–94.

LNRI No. 135, 2001. Undang-Undang Republik Indonesia Nomor 21 Tahun 2001 Tentang Otonomi Khusus Bagi Provinsi Papua. Lembaran Negara Republik Indonesia Tahun 2001 Nomor 135.

Long, J. S., Freese, J., 2006. Regression Models for Categorical Dependent Variables Using Stata, 2nd Edition. Stata Press, College Station, Texas.

Lowe, R., Chirombo, J., Tompkins, A. M., 2014. Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi. Malaria Journal 12, 416.

Lund, C., Breen, A., Flisher, A. J., Kakuma, R., Corrigall, J., Joska, J. A., Swartz, L., Patel, V., 2010. Poverty and common mental disorders in low and middle income countries: A systematic review. Social Science & Medicine 71 (3), 517–528.

Lund, C., Myer, L., Stein, D. J., Williams, D. R., Flisher, A. J., 2013. Mental illness and lost income among adult South Africans. Social Psychiatry and Psychiatric Epidemiology 48 (5), 845–851.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. The BUGS Book: A Practical Introduction to Bayesian Analysis. CRC Press, Boca Raton, FL.

Machado, J. A. F., Parente, P. M. D. C., Santos Silva, J. M. C., 2014. QREG2: Stata module to perform quantile regression with robust and clustered standard errors. Statistical Software Components.
URL `http://ideas.repec.org/c/boc/bocode/s457369.html`

Maharani, A., Tampubolon, G., 2014. Has decentralisation affected child immunisation status in Indonesia? Global Health Action 7, 24913.

Manh, B. H., Clements, A. C. A., Thieu, N. Q., Hung, N. M., Hung, L. X., Hay, S. I., Hien, T. T., Wertheim, H. F. L., Snow, R. W., Horby, P., 2011. Social and environmental determinants of malaria in space and time in Viet Nam. International Journal for Parasitology 41 (1), 109–116.

Mani, S., 2012. Is there complete, partial, or no recovery from childhood malnutrition? Empirical evidence from Indonesia. Oxford Bulletin of Economics and Statistics 74 (5), 691–715.

Marmot, M., 2005. Social determinants of health inequalities. Lancet 365 (9464), 1099–1104.

Marmot, M., 2013. Review of social determinants and the health divide in the WHO European Region: final report. WHO Regional Office for Europe, Copenhagen, updated reprint 2014.

Marmot, M., 2015. The Health Gap: The challenge of an unequal world. Bloomsbury, London.

Marmot, M., Shipley, M. J., Rose, G., 1984. Inequalities in death-specific explanations of a general pattern? Lancet 323 (8384), 1003–1006.

Marmot, M., Smith, G. D., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., 1991. Health inequalities among British civil servants: the Whitehall II study. Lancet 337 (8754), 1387–1393.

Martino, S., Rue, H., 2010. Case studies in Bayesian computation using INLA. In: Mantovan, P., Secchi, P. (Eds.), Complex Data Modeling and Computationally Intensive Statistical Methods. Contributions to Statistics. Springer, Milan, pp. 99–114.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers of Econometrics. Academic Press, New York, pp. 105–142.

Meer, J., Miller, D. L., Rosen, H. S., 2003. Exploring the health–wealth nexus. Journal of Health Economics 22 (5), 713–730.

Mendis, K., Sina, B. J., Marchesini, P., Carter, R., 2001. The neglected burden of Plasmodium vivax malaria. American Journal of Tropical Medicine and Hygiene 64 (1S), 97–106.

Meyer-Christoffer, A., Becker, A., Finger, P., Rudolf, B., Schneider, U., Ziese, M., 2011. GPCC climatology version 2011 at 0.5°: Monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data.
URL http://dx.doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_025

Miguel, E., Satyanath, S., Sergenti, E., 2004. Economic shocks and civil conflict: An instrumental variables approach. Journal of Political Economy 112 (4), 725–753.

Milanovic, B., 1997. A simple way to calculate the Gini coefficient, and some implications. Economics Letters 56 (1), 45–49.

Mmbando, B. P., Kamugisha, M. L., Lusingu, J. P., Francis, F., Ishengoma, D. S., Theander, T. G., Lemnge, M. M., Scheike, T. H., 2011. Spatial variation and socio-economic determinants of Plasmodium falciparum infection in northeastern Tanzania. Malaria Journal 10, 145.

Molini, V., Nubé, M., 2007. Is the nutritional status of males and females equally affected by economic growth? Evidence from Vietnam in the 1990s. United Nations University - World Institute for Development Economics Research (UNU-WIDER) Research Paper No. 2007/54.

URL http://www.wider.unu.edu/publications/working-papers/research-papers/2007/en_GB/rp2007-54/_print/

Morabia, A., 2013. Epidemiology's 350th anniversary: 1662–2012. Epidemiology 24 (2), 179–183.

MPR RI, 2011. Undang-Undang Dasar Negara Republik Indonesia Tahun 1945. Majelis Permusyawaratan Rakyat Republik Indonesia.
URL https://www.mpr.go.id/pages/produk-mpr/uud-nri-tahun-1945

Mu, R., 2014. Regional disparities in self-reported health: Evidence from Chinese older adults. Health Economics 23 (5), 529–549.

Muntaner, C., Eaton, W. W., Miech, R., O'Campo, P., 2004. Socioeconomic position and major mental disorders. Epidemiologic Reviews 26 (1), 53–62.

Murray, C. J. L., Tandon, A., Solomon, J. A., Mathers, C. D., Sadana, R., 2002. Cross-population comparability of evidence for health policy. Global Programme on Evidence for Health Policy Discussion Paper No. 46, World Health Organization.

Myer, L., Stein, D. J., Grimsrud, A., Seedat, S., Williams, D. R., 2008. Social determinants of psychological distress in a nationally-representative sample of South African adults. Social Science & Medicine 66 (8), 1828–1840.

Nazara, S., 2010. The informal economy in Indonesia: size, composition and evolution. The International Labour Organization Working Paper.
URL http://www.ilo.org/jakarta/whatwedo/publications/WCMS_145781/lang--en/index.htm

Nichols, A., 2010. Regression for nonnegative skewed dependent variables. Bos10 Stata Conference.
URL http://repec.org/bost10/nichols_boston2010.pdf

Nuyens, Y., 2007. Setting priorities for health research: lessons from low- and middle-income countries. Bulletin of the World Health Organization 85 (4), 319–321.

Oddo, V. M., Rah, J. H., Semba, R. D., Sun, K., Akhter, N., Sari, M., de Pee, S., Moench-Pfanner, R., Bloem, M., Kraemer, K., 2012. Predictors of maternal and child double burden of malnutrition in rural Indonesia and Bangladesh. American Journal of Clinical Nutrition 95 (4), 951–958.

O'Meara, W. P., Platt, A., Naanyu, V., Cole, D., Ndege, S., 2013. Spatial autocorrelation in uptake of antenatal care and relationship to individual, household and village-level factors: results from a community-based survey of pregnant women in six districts in western Kenya. International Journal of Health Geographics 12, 55.

Opeskin, B., 2009. Malaria in Pacific populations: seen but not heard? Journal of Population Research 26 (2), 175–199.

Osypuk, T. L., Joshi, P., Geronimo, K., Acevedo-Garcia, D., 2014. Do social and economic policies influence health? A review. Current Epidemiology Reports 1 (3), 149–164.

Paccagnella, O., 2011. Anchoring vignettes with sample selection due to non-response. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 665–687.

Parmanto, B., Paramita, M. V., Sugiantara, W., Pramana, G., Scotch, M., Burke, D. S., 2008. Spatial and multidimensional visualization of Indonesia's village health statistics. International Journal of Health Geographics 7, 30.

Pascutto, C., Wakefield, J. C., Best, N. G., Richardson, S., Bernardinelli, L., Staines, A., Elliott, P., 2000. Statistical issues in the analysis of disease mapping data. Statistics in Medicine 19 (17–18), 2493–2519.

Patel, V., 2007. Mental health in low- and middle-income countries. British Medical Bulletin 81–82 (1), 81–96.

Paxson, C. H., 1992. Using weather variability to estimate the response of savings to transitory income in Thailand. American Economic Review 82 (1), 15–33.

Peracchi, F., Rosetti, C., 2012. Heterogeneity in health responses and anchoring vignettes. Empirical Economics 42 (2), 513–538.

Phelan, J., Link, B. G., 2013. Fundamental Cause Theory. In: Cockerham, W. C. (Ed.), Medical Sociology on the Move: New Directions in Theory. Springer, Dordrecht, pp. 105–125.

Phelan, J., Link, B. G., Tehranifar, P., 2010. Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. Journal of Health and Social Behavior 51 (S), 28–40.

Pitt, M. M., Rosenzweig, M. R., Gibbons, D. M., 1995. The determinants and consequences of the placement of government programs in Indonesia. In: van de Walle,

D., Nead, K. (Eds.), Public Spending and the Poor: Theory and Evidence. John Hopkins University Press, Maryland, pp. 114–149.

Pollock, J. I., Manaseki-Holland, S., Patel, V., 2006. Detection of depression in women of child-bearing age in non-western cultures: A comparison of the Edinburgh Postnatal Depression Scale and the Self-Reporting Questionnaire-20 in Mongolia. Journal of Affective Disorders 92 (2–3), 267–271.

Popkin, B. M., 1998. The nutrition transition and its health implications in lower-income countries. Public Health Nutrition 1 (1), 5–21.

Popkin, B. M., 1999. Urbanization, lifestyle changes and the nutrition transition. World Development 27 (11), 1905–1916.

Popkin, B. M., 2001. Nutrition in transition: The changing global nutrition challenge. Asia Pacific Journal of Clinical Nutrition 10 (S), 13–18.

Poterba, J. M., 1989. Lifetime incidence and the distributional burden of excise taxes. American Economic Review 79 (2), 325–330.

Prince, M., Acosta, D., Albanese, E., Arizaga, R., Ferri, C. P., Guerra, M., Huang, Y., Jacob, K., Jimenez-Velazquez, I. Z., Rodriguez, J. L., Salas, A., Sosa, A. L., Sousa, R., Uwakwe, R., Poel, R. V. D., Williams, J., Wortmann, M., 2008. Ageing and dementia in low and middle income countries – Using research to engage with public and policy makers. International Review of Psychiatry 20 (4), 332–343.

PubMed, 2016. `(indonesia[Title/Abstract]) AND "journal article"[Publication Type]`. PubMed Database [online, accessed 13 May 2016].
URL `http://www.ncbi.nlm.nih.gov/pubmed?term=(indonesia%5BTitle%2FAbstract%5D)%20AND%20%22journal%20article%22%5BPublication%20Type%5D`

Rabe-Hesketh, S., Skrondal, A., 2012. Multilevel and Longitudinal Modeling Using Stata: Categorical Responses, Counts, and Survival, 3rd Edition. Vol. 2. Stata Press, College Station, Texas.

Rahmanian, M., Kelishadi, R., Qorbani, M., Motlagh, M. E., Shafiee, G., Aminaee, T., Ardalan, G., Taslimi, M., Poursafa, P., Asayesh, H., Larijani, B., Heshmat, R., 2014. Dual burden of body weight among Iranian children and adolescents in 2003 and 2010: the CASPIAN-III study. Archives of Medical Science 10 (1), 96–103.

RAND, 2007. IFLS-4 Survey Description. [online, accessed 1 May 2014].
URL `http://www.rand.org/labor/FLS/IFLS/ifls4.html`

RAND, 2014. Indonesia Family Life Survey East (IFLS EAST). [online, accessed 27 February 2015].
URL `http://www.rand.org/labor/FLS/IFLS/ifls-east.html`

Ranis, G., Stewart, F., 2012. Success and failure in human development, 1970–2007. Journal of Human Development and Capabilities 13 (2), 167–195.

Ravallion, M., 2010. Mashup indices of development. World Bank Policy Research Working Paper No. 5432.
URL   `http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-5432`

Ravallion, M., 2011. On multidimensional indices of poverty. Journal of Economic Inequality 9 (2), 235–248.

Ravallion, M., Chen, S., Sangraula, P., 2009. Dollar a day revisited. World Bank Economic Review 23 (2), 163–184.

Resosudarmo, B. P., Mollet, J. A., Raya, U. R., Kaiwai, H., 2014. Development in Papua after special autonomy. In: Hill, H. (Ed.), Regional Dynamics in a Decentralized Indonesia. ISEAS Publishing, Singapore, pp. 433–459.

Ricci, F., 2012. Social implications of malaria and their relationships with poverty. Mediterranean Journal of Hematology and Infectious Diseases 4 (1), e2012048.

Rice, N., Robone, S., Smith, P., 2011. Analysis of the validity of the vignette approach to correct for heterogeneity in reporting health system responsiveness. European Journal of Health Economics 12 (2), 141–162.

Richardson, S., Thomson, A., Best, N., Elliott, P., 2004. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives 112 (9), 1016–1025.

Robinson, W. S., 1950. Ecological correlations and the behavior of individuals. American Sociological Review 15 (3), 351–357.

Roche, J. M., 2013. Monitoring progress in child poverty reduction: Methodological insights and illustration to the case study of Bangladesh. Social Indicators Research 112 (2), 363–390.

Roemling, C., Qaim, M., 2012. Obesity trends and determinants in Indonesia. Appetite 58 (3), 1005–1013.

Roemling, C., Qaim, M., 2013. Dual burden households and intra-household nutritional inequality in Indonesia. Economics & Human Biology 11 (4), 563–573.

Rokx, C., Giles, J., Satriawan, E., Marzoeki, P., Harimurti, P., Yavuz, E., 2010. New Insights into the Provision of Health Services in Indonesia: A Health Workforce Study. World Bank, Washington, DC.

Sachs, J., Malaney, P., 2002. The economic and social burden of malaria. Nature 415 (6872), 680–685.

Sahn, D. E., Stifel, D., 2003. Exploring alternative measures of welfare in the absence of expenditure data. Review of Income and Wealth 49 (4), 463–489.

Salazar, R. C. A., Diaz, B. Y., Pinzon, R. P., 2013. A counting multidimensional poverty index in public policy context: the case of Colombia. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 62.
URL `http://www.ophi.org.uk/a-counting-multidimensional-poverty-index-in-public-policy-context-the-case-of-colombia/`

Sandjaja, Poh, B. K., Rojroonwasinkul, N., Le Nyugen, B. K., Budiman, B., Ng, L. O., Soonthorndhada, K., Xuyen, H. T., Deurenberg, P, Parikh, P., 2013. Relationship between anthropometric indicators and cognitive performance in Southeast Asian school-aged children. British Journal of Nutrition 110 (S3), 57–64.

Santos, M. E., 2013. Tracking poverty reduction in Bhutan: Income deprivation alongside deprivation in other sources of happiness. Social Indicators Research 112 (2), 259–290.

Santos, M. E., Ura, K., 2008. Multidimensional poverty in Bhutan: Estimates and policy implications. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 14.
URL `http://www.ophi.org.uk/working-paper-number-14/`

Santos Silva, J. M. C., Parente, P. M. D. C., 2013. Quantile regression with clustered data. Economics Discussion Paper No. 728.
URL `http://ideas.repec.org/p/esx/essedp/728.html`

Santos Silva, J. M. C., Tenreyro, S., 2006. The log of gravity. Review of Economics and Statistics 88 (4), 641–658.

Saraceno, B., Levav, I., Kohn, R., 2005. The public mental health significance of research on socio-economic factors in schizophrenia and major depression. World Psychiatry 4 (3), 181–185.

Sarsons, H., 2015. Rainfall and conflict: A cautionary tale. Journal of Development Economics 115, 62–72.

Savage, D., Fichera, E., 2013. Income, rainfall shocks and health: An instrumental variable approach. SSRN Scholarly Paper.
URL http://papers.ssrn.com/abstract=2280049

Scazufca, M., Menezes, P. R., Vallada, H., Araya, R., 2009. Validity of the self reporting questionnaire-20 in epidemiological studies with older adults. Social Psychiatry and Psychiatric Epidemiology 44 (3), 247–254.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2011. GPCC full data reanalysis version 6.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data.
URL http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V6_050

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theoretical and Applied Climatology 115 (1–2), 15–40.

Scholte, W. F., Verduin, F., van Lammeren, A., Rutayisire, T., Kamperman, A. M., 2011. Psychometric properties and longitudinal validation of the self-reporting questionnaire (SRQ-20) in a Rwandan community setting: a validation study. BMC Medical Research Methodology 11, 116.

Schrödle, B., Held, L., 2011. A primer on disease mapping and ecological regression using INLA. Computational Statistics 26 (2), 241–258.

Schultz, P. T., Tansel, A., 1997. Wage and labor supply effects of illness in Côte d'Ivoire and Ghana: instrumental variable estimates for days disabled. Journal of Development Economics 53 (2), 251–286.

Sen, A., 1976. Poverty: An ordinal approach to measurement. Econometrica 44 (2), 219–231.

Sen, A., 1985. Commodities and Capabilities. Elsevier Science, New York.

Sen, A., 2002. Health: perception versus observation. BMJ 324 (7342), 860–861.

Seth, S., Alkire, S., 2014. Did poverty reduction reach the poorest of the poor? Assessment methods in the counting approach. Oxford Poverty & Human Development Initiative (OPHI) Working Paper No. 77.
URL http://www.ophi.org.uk/did-poverty-reduction-reach-the-poorest-of-the-poor-assessment-methods-in-the-counting-approach/

Shafique, S., Akhter, N., Stallkamp, G., de Pee, S., Panagides, D., Bloem, M. W., 2007. Trends of under- and overweight among rural and urban poor women indicate the double burden of malnutrition in Bangladesh. International Journal of Epidemiology 36 (2), 449–457.

Shmueli, A., 2002. Reporting heterogeneity in the measurement of health and health-related quality of life. PharmacoEconomics 20 (6), 405–412.

Shrimpton, R., Rokx, C., 2013. The Double Burden of Malnutrition in Indonesia. World Bank Jakarta Report No. 76192-ID.
URL https://openknowledge.worldbank.org/handle/10986/17007

Simkiss, D. E., Blackburn, C. M., Mukoro, F. O., Read, J. M., Spencer, N. J., 2011. Childhood disability and socio-economic circumstances in low and middle income countries: systematic review. BMC Pediatrics 11, 119.

Skalická, V., van Lenthe, F., Bambra, C., Krokstad, S., Mackenbach, J., 2009. Material, psychosocial, behavioural and biomedical factors in the explanation of relative socio-economic inequalities in mortality: evidence from the HUNT study. International Journal of Epidemiology 38 (5), 1272–1284.

Smith, J. P., 1999. Healthy bodies and thick wallets: The dual relation between health and economic status. Journal of Economic Perspectives 13 (2), 145–166.

Snijders, T. A. B., Bosker, R. J., 2012. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, 2nd Edition. SAGE, London.

Soares Magalhães, R. J., Clements, A. C. A., Patil, A. P., Gething, P. W., Brooker, S., 2011. The applications of model-based geostatistics in helminth epidemiology and control. Advances in Parasitology 74, 267–296.

Somi, M. F., Butler, J. R., Vahid, F., Njau, J. D., Kachur, S. P., Abdulla, S., 2008. Use of proxy measures in estimating socioeconomic inequalities in malaria prevalence. Tropical Medicine & International Health 13 (3), 354–364.

Sonko, S. T., Jaiteh, M., Jafali, J., Jarju, L. B. S., D'Alessandro, U., Camara, A., Komma-Bah, M., Saho, A., 2014. Does socio-economic status explain the differentials in malaria parasite prevalence? Evidence from The Gambia. Malaria Journal 13, 449.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Linde, A. V. D., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583–639.

Stowasser, T., Heiss, F., McFadden, D., Winter, J., 2011. "Healthy, Wealthy and Wise?" Revisited: An Analysis of the Causal Pathways from Socioeconomic Status to Health. In: Investigations in the Economics of Aging. NBER Chapters. National Bureau of Economic Research, United States of America, pp. 267–317.

Stratton, K. J., Aggen, S. H., Richardson, L. K., Acierno, R., Kilpatrick, D. G., Gaboury, M. T., Tran, T. L., Trung, L. T., Tam, N. T., Tuan, T., Buoi, L. T., Ha, T. T., Thach, T. D., Amstadter, A. B., 2013. Evaluation of the psychometric properties of the Self-Reporting Questionnaire (SRQ-20) in a sample of Vietnamese adults. Comprehensive Psychiatry 54 (4), 398–405.

Strauss, J., Beegle, K., Dwiyanto, A., Herawati, Y., Pattinasarany, D., Satriawan, E., Sikoki, B., Sukamdi, Witoelar, F., 2004. Indonesian Living Standards Before and After the Financial Crisis: Evidence from the Indonesia Family Life Survey. ISEAS Publishing, Singapore.

Stresman, G. H., 2010. Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. Acta Tropica 116 (3), 167–172.

Subramanian, S., Huijts, T., Avendano, M., 2010. Self-reported health assessments in the 2002 World Health Survey: how do they correlate with education? Bulletin of the World Health Organization 88 (2), 131–138.

Subramanian, S. V., Kawachi, I., 2004. Income inequality and health: What have we learned so far? Epidemiologic Reviews 26 (1), 78–91.

Subramanian, S. V., Kawachi, I., Smith, G. D., 2007. Income inequality and the double burden of under- and overnutrition in India. Journal of Epidemiology and Community Health 61 (9), 802–809.

Subramanian, S. V., Perkins, J. M., Khan, K. T., 2009. Do burdens of underweight and overweight coexist among lower socioeconomic groups in India? American Journal of Clinical Nutrition 90 (2), 369–376.

Subramanian, S. V., Smith, G. D., 2006. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. American Journal of Clinical Nutrition 84 (3), 633–640.

Sumarto, S., Vothknecht, M., Wijaya, L., 2014. Explaining regional heterogeneity of poverty: evidence from a decentralized Indonesia. In: Hill, H. (Ed.), Regional Dynamics in a Decentralized Indonesia. ISEAS Publishing, Singapore, pp. 285–314.

Sumawinata, I. W., Bernadeta, Leksana, B., Sutamihardja, A., Subianto, P. B., Sekartuti, Fryauff, D. J., Baird, J. K., 2003. Very high risk of therapeutic failure with chloroquine for uncomplicated Plasmodium falciparum and P. vivax malaria in Indonesian Papua. American Journal of Tropical Medicine and Hygiene 68 (4), 416–420.

Surbakti, P., 1995. Indonesia's National Socio-economic Survey: A Continual Data Source for Analysis on Welfare Development. Badan Pusat Statistik, Jakarta, Indonesia.

Székely, M., 2003. Lo que dicen los pobres. Cuadernos de desarrollo humano No. 13. Secretaría de Desarrollo Social, México.

Tampubolon, G., Hanandita, W., 2014. Poverty and mental health in Indonesia. Social Science & Medicine 106, 20–27.

Tandon, A., Murray, C. J. L., Solomon, J. A., King, G., 2002. Statistical models for enhancing cross-population comparability. Global Programme on Evidence for Health Policy Discussion Paper No. 42, World Health Organization.

Teklehaimanot, A., Meija, P., 2008. Malaria and poverty. Annals of the New York Academy of Sciences 1136 (1), 32–37.

Terza, J. V., Basu, A., Rathouz, P. J., 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics 27 (3), 531–543.

Thabrany, H., 2006. Human resources in decentralized health systems in Indonesia: Challenges for equity. Regional Health Forum 10 (1), 75–88.

Thomas, D., 1990. Intra-household resource allocation: An inferential approach. Journal of Human Resources 25 (4), 635–664.

Thomas, D., Strauss, J., 1992. Prices, infrastructure, household characteristics and child height. Journal of Development Economics 39 (2), 301–331.

Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., Suriastini, W., 2012. Cutting the costs of attrition: Results from the Indonesia Family Life Survey. Journal of Development Economics 98 (1), 108–123.

Thomson, M. C., Connor, S. J., D'Alessandro, U., Rowlingson, B., Diggle, P., Cresswell, M., Greenwood, B., 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. American Journal of Tropical Medicine and Hygiene 61 (1), 2–8.

Tjitra, E., Anstey, N. M., Sugiarto, P., Warikar, N., Kenangalem, E., Karyana, M., Lampah, D. A., Price, R. N., 2008. Multidrug-resistant Plasmodium vivax associated with severe and fatal malaria: A prospective study in Papua, Indonesia. PLoS Medicine 5 (6), e128.

Tol, W. A., Reis, R., Susanty, D., de Jong, J. T. V. M., 2010. Communal violence and child psychosocial well-being: Qualitative findings from Poso, Indonesia. Transcultural Psychiatry 47 (1), 112–135.

Townsend, P., 1979. Poverty in the United Kingdom. Allen Lane and Penguin Books, London.
URL http://www.poverty.ac.uk/free-resources-books/poverty-united-kingdom

Trani, J.-F., Biggeri, M., Mauro, V., 2013. The multidimensionality of child poverty: Evidence from Afghanistan. Social Indicators Research 112 (2), 391–416.

Tusting, L. S., Willey, B., Lucas, H., Thompson, J., Kafy, H. T., Smith, R., Lindsay, S. W., 2013. Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. Lancet 382 (9896), 963–972.

UNDP, 2010. The real wealth of nations: Pathways to human development. Macmillan, New York.

UNDP, 2014. Why is the MPI better than the Human Poverty Index (HPI) which was previously used in the Human Development Reports? Frequently Asked Questions, United Nations Development Programme [online, accessed 27 February 2015].
URL http://hdr.undp.org/en/faq-page/multidimensional-poverty-index-mpi#t295n138

UNSCN, 2006. Tackling the Double Burden of Malnutrition: A Global Agenda. United Nations System Standing Committee on Nutrition, SCN News No. 32.

URL http://www.unscn.org/layout/modules/resources/files/scnnews32.pdf

UNSD, 2014. Purchasing power parities (PPP) conversion factor, local currency unit to international dollar. United Nations Statistics Division.
URL http://mdgs.un.org/unsd/mdg/SeriesDetail.aspx?srid=699

Ura, K., Alkire, S., Zangmo, T., Wangdi, K., 2012. An extensive analysis of GNH index. Centre for Bhutan Studies.
URL http://www.grossnationalhappiness.com/wp-content/uploads/2012/10/An%20Extensive%20Analysis%20of%20GNH%20Index.pdf

Vaezghasemi, M., Öhman, A., Eriksson, M., Hakimi, M., Weinehall, L., Kusnanto, H., Ng, N., 2014. The effect of gender and social capital on the dual burden of malnutrition: A multilevel study in Indonesia. PLoS ONE 9 (8), e103849.

van de Walle, D., 1988. On the use of the susenas for modelling consumer behaviour. Bulletin of Indonesian Economic Studies 24 (2), 107–121.

van Doorslaer, E., Koolman, X., Jones, A. M., 2004. Explaining income-related inequalities in doctor utilisation in Europe. Health Economics 13 (7), 629–647.

van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., Smith, J. P., 2011. Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 575–595.

Vathesatogkit, P., Batty, G. D., Woodward, M., 2014. Socioeconomic disadvantage and disease-specific mortality in Asia: systematic review with meta-analysis of population-based cohort studies. Journal of Epidemiology and Community Health 68 (4), 375–383.

Victor, B., Blevins, M., Ndatimana, E., González-Calvo, L., Fischer, E. F., Vergara, A. E., Vermund, S. H., Olupona, O., Moon, T. D., 2014. Multidimensional poverty in rural Mozambique: A new metric for evaluating public health interventions. PLoS ONE 9 (9), e108654.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., Sachdev, H. S., 2008. Maternal and child undernutrition: consequences for adult health and human capital. Lancet 371 (9609), 340–357.

Villermé, L.-R., 1830. De la mortalité dans divers quartiers de la ville de Paris. Annales d'hygiene publique 3, 294–341.

Virchow, R., 1848. Report on the typhus epidemic in Upper Silesia. In: Rather, L. J. (Ed.), Rudolf Virchow: Collected Essays on Public Health and Epidemiology. Science History Publications, Canton, MA, pp. 205–220.

Voňková, H., Hullegie, P., 2011. Is the anchoring vignette method sensitive to the domain and choice of the vignette? Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 597–620.

Wall, M. M., 2004. A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference 121 (2), 311–324.

Wallace, R. B., Herzog, A. R., 1995. Overview of the health measures in the health and retirement study. Journal of Human Resources 30, 84–107.

Wand, J., 2013. Credible comparisons using interpersonally incomparable data: Nonparametric scales with anchoring vignettes. American Journal of Political Science 57 (1), 249–262.

Wand, J., King, G., Lau, O., 2011. anchors: Software for anchoring vignette data. Journal of Statistical Software 42 (3).

Webb, P., Block, S., 2004. Nutrition information and formal schooling as inputs to child nutrition. Economic Development and Cultural Change 52 (4), 801–820.

Webster, D., 2001. Malaria kills one child every 30 seconds. Journal of Public Health Policy 22 (1), 23–33.

Weiss, D. J., Mappin, B., Dalrymple, U., Bhatt, S., Cameron, E., Hay, S. I., Gething, P. W., 2015. Re-examining environmental correlates of Plasmodium falciparum malaria endemicity: a data-intensive variable selection approach. Malaria Journal 14, 68.

Whelan, C. T., Layte, R., Maître, B., 2004. Understanding the mismatch between income poverty and deprivation: A dynamic comparative analysis. European Sociological Review 20 (4), 287–302.

Whitehead, M., 1992. The concepts and principles of equity and health. International Journal of Health Services 22 (3), 429–445.

WHO, 2010. Indonesia 2010 Landscape Analysis Country Assessment Report. World Health Organization.
URL `http://www.who.int/nutrition/landscape_analysis/Indonesia/en/`

WHO, 2012a. Global Report for Research on Infectious Diseases of Poverty. World Health Organization.
URL http://www.who.int/tdr/publications/global_report/en/

WHO, 2012b. Malaria and its control in the WHO South-East Asia Region. World Health Organization, Regional Office for South-East Asia.
URL http://www.searo.who.int/entity/malaria/topics/Malaria_factsheetWMD2012.pdf

WHO, 2013. Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. World Health Organization [online, accessed 9 May 2016].
URL http://www.who.int/gho/health_equity/handbook/en/

WHO, 2014. World Malaria Report 2014. World Health Organization.
URL http://www.who.int/malaria/publications/world_malaria_report/en/

WHO Expert Consultation, 2004. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 363 (9403), 157–163.

Wilkinson, R. G., Pickett, K., 2010. The Spirit Level: Why Equality is Better for Everyone, revised Edition. Penguin, London.

Windmeijer, F. A. G., Santos Silva, J. M. C., 1997. Endogeneity in count data models: An application to demand for health care. Journal of Applied Econometrics 12 (3), 281–294.

Winkvist, A., Nurdiati, D. S., Stenlund, H., Hakimi, M., 2000. Predicting under- and overnutrition among women of reproductive age: a population-based study in central Java, Indonesia. Public Health Nutrition 3 (2), 193–200.

Withrow, D., Alter, D. A., 2011. The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. Obesity Reviews 12 (2), 131–141.

Wolff, F.-C., Maliki, 2008. Evidence on the impact of child labor on child health in Indonesia, 1993–2000. Economics & Human Biology 6 (1), 143–169.

World Bank, 2014a. New Poverty Frontier in Indonesia: Reduction Slows, Inequality Rises. World Bank News.
URL http://www.worldbank.org/en/news/feature/2014/09/23/new-poverty-frontier-in-indonesia-reduction-slows-inequality-rises

World Bank, 2014b. World Development Indicators (WDI). The World Bank Online Database.
URL `http://data.worldbank.org/indicator`

Worrall, E., Basu, S., Hanson, K., 2005. Is malaria a disease of poverty? A review of the literature. Tropical Medicine & International Health 10 (10), 1047–1059.

Yadav, K., Dhiman, S., Rabha, B., Saikia, P. K., Veer, V., 2014. Socio-economic determinants for malaria transmission risk in an endemic primary health centre in Assam, India. Infectious Diseases of Poverty 3, 19.

Yamey, G., 2012. What are the barriers to scaling up health interventions in low and middle income countries? A qualitative study of academic leaders in implementation science. Globalization and Health 8, 11.

Yang, D., Choi, H., 2007. Are remittances insurance? Evidence from rainfall shocks in the Philippines. World Bank Economic Review 21 (2), 219–248.

Yu, J., 2013. Multidimensional poverty in China: Findings based on the CHNS. Social Indicators Research 112 (2), 315–336.

Zimmerman, F. J., Katon, W., 2005. Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship? Health Economics 14 (12), 1197–1215.

Appendix A:
Supplementary data for Chapter 2

Figure A.1: Title page of paper 1

Does poverty reduce mental health? An instrumental variable analysis

Wulung Hanandita*, Gindo Tampubolon

Institute for Social Change, University of Manchester, Humanities Bridgeford Street Building 2F, Oxford Road, Manchester M13 9PL, United Kingdom

ABSTRACT

That poverty and mental health are negatively associated in developing countries is well known among epidemiologists. Whether the relationship is causal or associational, however, remains an open question. This paper aims to estimate the causal effect of poverty on mental health by exploiting a natural experiment induced by weather variability across 440 districts in Indonesia ($N = 577,548$). Precipitation anomaly in two climatological seasons is used as an instrument for poverty status, which is measured using per capita household consumption expenditure. Results of an instrumental variable estimation suggest that poverty causes poor mental health: halving one's consumption expenditure raises the probability of suffering mental illness by 0.06 point; in terms of elasticity, a 1% decrease in consumption brings about 0.62% more symptoms of common mental disorders. This poverty effect is approximately five times stronger than that obtained prior to instrumenting and is robust to alternative distributional assumption, model specification, sample stratification and estimation technique. An individual's mental health is also negatively correlated with district income inequality, suggesting that income distribution may have a significant influence upon mental health over and above the effect of poverty. The findings imply that mental health can be improved not only by influencing individuals' health knowledge and behaviour but also by implementing a more equitable economic policy.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The negative association between poverty and mental health in developing countries has been increasingly documented. Research from various parts of the world generally shows that low levels of income, education, and assets as well as low social class are correlated with a higher probability of having common mental disorders (Lund et al., 2010). However, empirical evidence regarding the causal effect of the association remains scarce. Few studies have investigated the strength or the direction of causality between poverty and mental health in developing countries, although such study clearly benefits the formulation of public policy aimed at improving the health of the population. In encouraging study of this topic in the United States, Stowasser et al. (2011, p.2) note that ' … if causal links between wealth and health were confirmed, society would likely benefit from more universal access to health care and redistributive economic policy. Yet, if such causal links were rebutted, resources would be better spent on influencing health knowledge, preferences, and ultimately the

behavior of individuals.' Considering both the growing burden of disease attributed to mental illness (IHME, 2013) and tightly constrained health budgets (Patel, 2007), it is important to understand whether poverty reduces mental health in developing countries.

The fact that poverty is negatively associated with mental health in low- and middle-income countries is hardly surprising, but to reach a convincing estimate of its causal effect is certainly not an easy task. Two-way or simultaneous causation may come into play (Smith, 1999), inflating the estimated effect and making it impossible for researchers looking at observational data to separate the effect of wealth on mental health (social causation hypothesis) from that of the reverse (social selection hypothesis). Secondly, the observed wealth–health relationship may be confounded by unobserved common causes that accidentally induce a spurious correlation. Genetic frailty, early childhood environment, family background and preference or taste for lifestyle may impact both an individual's ability to work (and hence accumulate wealth) and his or her susceptibility to mental illness (Stowasser et al., 2011). The study on the mental health effect of poverty may also suffer from what is generally known as the attenuation bias. More often than not, wealth is measured with error, as a noisy, low signal-to-noise ratio variable which could trivially result in a downward-biased parameter estimate (Cameron and Trivedi, 2005). Because these endogeneity problems might be working at the same time, it is

* Corresponding author. Institute for Social Change, University of Manchester, HBS Room 2.12A, Oxford Road, Manchester M13 9PL, United Kingdom.
E-mail address: wulung.hanandita@postgrad.manchester.ac.uk (W. Hanandita).

Table A.1: The 20-item Self-Reporting Questionnaire (SRQ-20)

No.	Item
1.	Do you often have headaches?
2.	Is your appetite poor?
3.	Do you sleep badly?
4.	Are you easily frightened?
5.	Do you feel nervous, tense or worried?
6.	Do your hands shake?
7.	Is your digestion poor?
8.	Do you have trouble thinking clearly?
9.	Do you feel unhappy?
10.	Do you cry more than usual?
11.	Do you find it difficult to enjoy your daily activities?
12.	Do you find it difficult to make decisions?
13.	Is your daily work suffering?
14.	Are you unable to play a useful part in life?
15.	Have you lost interest in things?
16.	Do you feel that you are a worthless person?
17.	Has the thought of ending your life been on your mind?
18.	Do you feel tired all the time?
19.	Do you have uncomfortable feelings in your stomach?
20.	Are you easily tired?

Table A.2: Exploratory factor analysis of district deprivation index

Proportion of village without ...	Factor loading	Summary statistics	
Communication facilities	0.86	Explained variance	88%
Electricity	0.81	Cronbach's α	0.82
Street lighting	0.76	Eigenvalue	3.58
Healthcare facilities	0.75	KMO	0.80
TV signal coverage	0.73	N	454
Education facilities	0.65		
Entertainment facilities	0.30		

Figure A.2: Coverage of the Indonesia Family Life Survey (IFLS) 2007 (shaded area = sampled, unshaded area = not sampled)

District

IFLS

Non-IFLS

Instrumental variable estimation via Generalized Method of Moments (GMM)

Linear and linear probability models

Assume a population moment condition:

$$\mathrm{E}[u_i|z_i] = \mathbf{0} \Rightarrow \mathrm{E}[z_i u_i] = \mathbf{0} \Rightarrow \mathrm{E}[z_i(y_i - x_i'\boldsymbol{\beta})] = \mathbf{0}$$

where z is a set of exogenous variable, u is the error term, y is the dependent variable, x is the set of covariate excluding the instrument, and $\boldsymbol{\beta}$ is the parameter vector of interest.

Equate this population moment condition with its sample analogue and then construct a quadratic form:

$$Q(\boldsymbol{\beta}) = \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i\boldsymbol{\beta})z_i\right]'}_{1\times M} \cdot \underbrace{W}_{M\times M} \cdot \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i\boldsymbol{\beta})z_i\right]}_{M\times 1}$$

$$= \underbrace{g(\boldsymbol{\beta})'}_{1\times M}\underbrace{W}_{M\times M}\underbrace{g(\boldsymbol{\beta})}_{M\times 1}$$

where W is a symmetric positive definite weighting matrix. Since in an overidentified instrumental variables regression model there are more equations (the number of moment conditions) than there are unknowns (the number of coefficients in $\boldsymbol{\beta}$)—that is $\dim(z) > \dim(\boldsymbol{\beta})$, one cannot uniquely solve out for $\boldsymbol{\beta}$ using the original Method of Moments. Instead, the Generalized Method of Moments (GMM) defines its objective function as $Q(\boldsymbol{\beta})$ above, and finds $\hat{\boldsymbol{\beta}}$ that brings this quantity as close to zero as possible.

$$\hat{\boldsymbol{\beta}}_{\mathrm{GMM}} = \arg\min_{\beta}\left\{Q(\boldsymbol{\beta})\right\}$$

$$= \arg\min_{\beta}\left\{g(\boldsymbol{\beta})'Wg(\boldsymbol{\beta})\right\}$$

The GMM estimator is consistent for any weighting matrix W, but efficiency is not guaranteed. Conduct a two-step GMM estimation and apply cluster-robust variance estimator to obtain efficient estimates that are robust to arbitrary heteroscedasticity and autocorrelation within cluster. Detailed exposition is available in Baum et al. (2007).

Poisson model

Assume a population moment condition:

$$\mathrm{E}\big[u_i \big| z_i\big] = \mathbf{0} \Rightarrow \mathrm{E}\big[z_i u_i\big] = \mathbf{0} \Rightarrow \mathrm{E}\big[z_i(y_i - \exp\{x_i'\beta\})\big] = \mathbf{0}$$

where z is a set of exogenous variable, u is the error term, y is the dependent variable, x is the set of covariate excluding the instrument, and β is the parameter vector of interest.

Equate this population moment condition with its sample analogue and then construct a quadratic form:

$$\underbrace{Q(\beta)}_{1\times 1} = \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \exp\{x_i\beta\}\right)z_i\right]'}_{1\times M} \cdot \underbrace{W}_{M\times M} \cdot \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \exp\{x_i\beta\}\right)z_i\right]}_{M\times 1}$$

$$= \underbrace{g(\beta)'}_{1\times M} \underbrace{W}_{M\times M} \underbrace{g(\beta)}_{M\times 1}$$

where W is a symmetric positive definite weighting matrix. Since in an overidentified instrumental variables regression model there are more equations (the number of moment conditions) than there are unknowns (the number of coefficients in β)—that is $\dim(z) > \dim(\beta)$, one cannot uniquely solve out for β using the original Method of Moments. Instead, the Generalized Method of Moments (GMM) defines its objective function as $Q(\beta)$ above, and finds $\hat{\beta}$ that brings this quantity as close to zero as possible.

$$\hat{\beta}_{\mathrm{GMM}} = \arg\min_{\beta}\{Q(\beta)\}$$
$$= \arg\min_{\beta}\left\{g(\beta)'\,W g(\beta)\right\}$$

The GMM estimator is consistent for any weighting matrix W, but efficiency is not guaranteed. Conduct a two-step GMM estimation and apply cluster-robust variance estimator to obtain efficient estimates that are robust to arbitrary heteroscedasticity and autocorrelation within cluster. Detailed exposition is available in Windmeijer and Santos Silva (1997).

Instrumental variable estimation for binary dependent variable via Maximum Likelihood (ML)

Specify a two-equation (Probit and linear) model:

$$
\begin{aligned}
y_{1i}^* &= x_i'\beta + \delta y_{2i} + u_i \\
y_{2i} &= z_i'\Pi + e_i \\
y_{1i} &= \begin{cases} 1 & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \leq 0 \end{cases}
\end{aligned}
$$

where y_1^* is the dependent variable, x is the set of exogenous covariate, y_2 is the endogenous variable, z is the instrument set, β, δ and Π are the parameter vectors, and u and e are the error terms that are assumed to come from a bivariate normal distribution:

$$
\begin{pmatrix} u \\ e \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_e \\ \rho\sigma_e & \sigma_e^2 \end{pmatrix} \right]
$$

Estimate the parameters using Maximum Likelihood and apply cluster-robust variance estimator to get efficient estimates that are robust to arbitrary heteroscedasticity and autocorrelation within cluster.

Probit instead of logistic regression was considered because the probit has a normally distributed error that is easier to exploit than the logistic. More specifically, the bivariate normal error offers the ability to conduct a test of the null hypothesis of exogeneity ($H_0 : \rho = 0$), which is very convenient for the purpose of this chapter. Detailed exposition is available in Cameron and Trivedi (2010).

Instrumental variable estimation via Control Function

Firstly, fit a reduced-form regression using OLS:

$$
y_{2i} = z_i'\Pi + e_i
$$

Then obtain the residual and plug it to the main equation (intuitively, making the unobservables observed):

$$
y_{1i} = x_i'\beta + \delta y_{2i} + \lambda \hat{e}_i + u_i
$$

Fit the main model and adjust the standard error for the two-step estimation using bootstrap. In these equations: y_2 is the endogenous variable, z is the instrument set, Π, β, δ, and λ are the parameter vectors, y_1 is the dependent variable, x is the set of exogenous covariate, u and e are the error terms, and $\hat{e}$ is the residual. Note that the main dependent variable y_1 can be continuous (OLS) or discrete (Probit, Poisson), but this approach only works if the endogenous variable y_2 is continuous. Detailed exposition is available in Imbens and Wooldridge (2007) and Terza et al. (2008).

On the use of linear model

Linear model is not an optimal choice for modelling the two outcomes of interest: the raw mental health score and the binary indicator of mental illness. Applied to the raw mental health score that is strictly non-negative and heavily right-skewed, linear model could give unrealistic prediction (negative predicted values). Similarly, applied to the binary outcome, the model may yield predicted probabilities that are not bounded between zero and one. On top of that, formal tests and inference obtained from the model are hardly trustworthy because they are based on the assumption of homoscedastic normal error.

Linear model, however, was given a place in the analysis because of two reasons. Firstly, it permits the use of several diagnostics tests that are essential for instrumental variable estimation such as test of validity of overidentifying instruments as well as test for weak instruments. These tests are still under extensive research and they are not yet available for generalised linear models. Secondly, linear model was presented in the analysis for the purpose of comparison to existing studies—mainly with those employing the classical two-stage least squares (2SLS) estimator.

Table A.3: Estimation results, Riskesdas 2007 rural sample aged 15 or older (N = 363,356)

Predictors	Mental Health Score				Probable Caseness			
	Linear	Linear-IV	Poisson	Poisson-IV	LPM	LPM-IV	Probit	Probit-IV
Log(PCE)	-0.21 ± 0.07†	-1.45 ± 0.70†	-0.10 ± 0.03†	-0.66 ± 0.33†	-0.02 ± 0.01†	-0.10 ± 0.05*	-0.02 ± 0.01‡	-0.10 ± 0.06*
Age 25–34	0.05 ± 0.02†	0.06 ± 0.03†	0.02 ± 0.01*	0.02 ± 0.01*	0.00 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	0.00 ± 0.00
Age 35–44	0.16 ± 0.03‡	0.22 ± 0.05‡	0.08 ± 0.02‡	0.11 ± 0.02‡	0.01 ± 0.00†	0.01 ± 0.00†	0.01 ± 0.00†	0.01 ± 0.00†
Age 45–54	0.34 ± 0.04‡	0.46 ± 0.08‡	0.16 ± 0.02‡	0.22 ± 0.04‡	0.02 ± 0.00‡	0.02 ± 0.01‡	0.02 ± 0.00‡	0.02 ± 0.01‡
Age 55–64	0.60 ± 0.05‡	0.68 ± 0.07‡	0.26 ± 0.02‡	0.29 ± 0.03‡	0.04 ± 0.00‡	0.04 ± 0.00‡	0.04 ± 0.00‡	0.04 ± 0.01‡
Age 65+	1.00 ± 0.07‡	1.00 ± 0.08‡	0.38 ± 0.03‡	0.37 ± 0.03‡	0.08 ± 0.01‡	0.08 ± 0.01‡	0.06 ± 0.01‡	0.06 ± 0.01‡
Female	0.60 ± 0.03‡	0.59 ± 0.03‡	0.28 ± 0.01‡	0.28 ± 0.01‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.05 ± 0.00‡
Never married	-0.11 ± 0.03‡	-0.05 ± 0.04	-0.07 ± 0.02‡	-0.05 ± 0.02†	-0.00 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.00
Divorced	0.30 ± 0.06‡	0.27 ± 0.06‡	0.10 ± 0.02‡	0.09 ± 0.02‡	0.03 ± 0.01‡	0.03 ± 0.01‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Widowed	0.30 ± 0.04‡	0.28 ± 0.04‡	0.06 ± 0.01‡	0.05 ± 0.02‡	0.03 ± 0.00‡	0.03 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Middle school	-0.25 ± 0.03‡	-0.08 ± 0.10	-0.13 ± 0.02‡	-0.04 ± 0.04	-0.02 ± 0.00‡	-0.00 ± 0.01	-0.02 ± 0.00‡	-0.01 ± 0.01
High school	-0.37 ± 0.04‡	-0.02 ± 0.19	-0.19 ± 0.02‡	-0.03 ± 0.08	-0.03 ± 0.00‡	-0.00 ± 0.01	-0.03 ± 0.00‡	-0.01 ± 0.02
College	-0.39 ± 0.06‡	0.21 ± 0.34	-0.19 ± 0.03‡	0.07 ± 0.15	-0.03 ± 0.01‡	0.01 ± 0.02	-0.03 ± 0.01‡	0.01 ± 0.03
Unemployed	0.29 ± 0.03‡	0.25 ± 0.04‡	0.12 ± 0.01‡	0.11 ± 0.01‡	0.03 ± 0.00‡	0.03 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Less physical activity	0.14 ± 0.04‡	0.21 ± 0.05‡	0.06 ± 0.02‡	0.09 ± 0.02‡	0.01 ± 0.00‡	0.02 ± 0.00‡	0.01 ± 0.00‡	0.02 ± 0.00‡
Frequent smoker	0.22 ± 0.03‡	0.21 ± 0.03‡	0.10 ± 0.01‡	0.10 ± 0.01‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Heavy drinker	0.51 ± 0.12‡	0.47 ± 0.12‡	0.23 ± 0.05‡	0.22 ± 0.05‡	0.04 ± 0.01‡	0.04 ± 0.01‡	0.04 ± 0.01‡	0.04 ± 0.01‡
Chronic illness	1.36 ± 0.05‡	1.44 ± 0.06‡	0.45 ± 0.01‡	0.49 ± 0.02‡	0.11 ± 0.01‡	0.12 ± 0.01‡	0.08 ± 0.00‡	0.09 ± 0.01‡
Household size	-0.03 ± 0.01†	-0.13 ± 0.06†	-0.01 ± 0.00†	-0.06 ± 0.03†	-0.00 ± 0.00‡	-0.01 ± 0.00†	-0.00 ± 0.00†	-0.01 ± 0.01*
District deprivation	0.00 ± 0.06	-0.02 ± 0.07	0.00 ± 0.03	-0.01 ± 0.03	-0.00 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.01
District inequality	3.39 ± 1.24†	2.79 ± 1.33†	1.54 ± 0.56†	1.11 ± 0.66*	0.29 ± 0.11†	0.25 ± 0.11†	0.29 ± 0.11‡	0.25 ± 0.11*
Intercept	1.59 ± 0.06‡	1.42 ± 0.11‡	0.49 ± 0.03‡	0.39 ± 0.08‡	0.06 ± 0.00‡	0.05 ± 0.01‡		0.25 ± 0.12†
Estimator	OLS	GMM	GMM	GMM	OLS	GMM	ML	ML
Instruments' validity		0.79		0.76		0.62		
Log(PCE)'s exogeneity		3.12*				2.73*		2.24
Instruments' strength		21.40				21.40		

Note: * p < 0.10, † p < 0.05, ‡ p < 0.01.

Table A.4: Estimation results, Riskesdas 2007 urban sample aged 15 or older (N = 214,192)

Predictors	Mental Health Score				Probable Caseness			
	Linear	Linear-IV	Poisson	Poisson-IV	LPM	LPM-IV	Probit	Probit-IV
Log(PCE)	-0.26 ± 0.07‡	-1.28 ± 0.67*	-0.13 ± 0.03‡	-0.64 ± 0.34*	-0.02 ± 0.00‡	-0.09 ± 0.05*	-0.02 ± 0.00‡	-0.09 ± 0.06
Age 25–34	-0.04 ± 0.03	-0.04 ± 0.03	-0.03 ± 0.02	-0.03 ± 0.02	-0.01 ± 0.00†	-0.01 ± 0.00†	-0.01 ± 0.00‡	-0.01 ± 0.00‡
Age 35–44	-0.04 ± 0.04	-0.00 ± 0.05	-0.02 ± 0.02	-0.00 ± 0.03	-0.01 ± 0.00‡	-0.01 ± 0.00†	-0.02 ± 0.00‡	-0.01 ± 0.00‡
Age 45–54	-0.02 ± 0.05	0.13 ± 0.11	-0.01 ± 0.03	0.06 ± 0.06	-0.02 ± 0.00‡	-0.01 ± 0.01	-0.02 ± 0.00‡	-0.01 ± 0.01
Age 55–64	0.07 ± 0.06	0.21 ± 0.11*	0.03 ± 0.03	0.10 ± 0.05*	-0.01 ± 0.01†	-0.00 ± 0.01	-0.01 ± 0.01†	-0.00 ± 0.01
Age 65+	0.39 ± 0.08‡	0.47 ± 0.09‡	0.13 ± 0.03‡	0.17 ± 0.04‡	0.01 ± 0.01*	0.02 ± 0.01†	0.00 ± 0.01	0.01 ± 0.01
Female	0.63 ± 0.04‡	0.66 ± 0.05‡	0.34 ± 0.01‡	0.35 ± 0.02‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.05 ± 0.00‡	0.06 ± 0.01‡
Never married	-0.11 ± 0.04†	-0.01 ± 0.08	-0.07 ± 0.02†	-0.03 ± 0.04	-0.00 ± 0.00	0.00 ± 0.01	-0.00 ± 0.00	0.00 ± 0.01
Divorced	0.45 ± 0.06‡	0.43 ± 0.07‡	0.17 ± 0.02‡	0.15 ± 0.03‡	0.04 ± 0.01‡	0.04 ± 0.01‡	0.03 ± 0.00‡	0.03 ± 0.01‡
Widowed	0.18 ± 0.05‡	0.17 ± 0.05†	0.03 ± 0.02	0.02 ± 0.02	0.02 ± 0.01‡	0.02 ± 0.01‡	0.01 ± 0.00†	0.01 ± 0.00‡
Middle school	-0.40 ± 0.04‡	-0.21 ± 0.13	-0.18 ± 0.02‡	-0.08 ± 0.07	-0.03 ± 0.00‡	-0.02 ± 0.01‡	-0.03 ± 0.00‡	-0.02 ± 0.01
High school	-0.57 ± 0.04‡	-0.18 ± 0.25	-0.27 ± 0.02‡	-0.08 ± 0.13	-0.05 ± 0.00‡	-0.02 ± 0.02	-0.04 ± 0.00‡	-0.02 ± 0.02
College	-0.69 ± 0.05‡	-0.08 ± 0.39	-0.36 ± 0.03‡	-0.05 ± 0.20	-0.05 ± 0.00‡	-0.01 ± 0.03	-0.06 ± 0.01‡	-0.02 ± 0.03
Unemployed	0.23 ± 0.04‡	0.14 ± 0.07*	0.10 ± 0.02‡	0.06 ± 0.03†	0.02 ± 0.00‡	0.02 ± 0.01†	0.02 ± 0.00‡	0.01 ± 0.01†
Less physical activity	0.18 ± 0.05‡	0.27 ± 0.08‡	0.09 ± 0.03‡	0.13 ± 0.04‡	0.02 ± 0.00‡	0.02 ± 0.01‡	0.02 ± 0.00‡	0.02 ± 0.01‡
Frequent smoker	0.29 ± 0.03‡	0.27 ± 0.03‡	0.16 ± 0.02‡	0.15 ± 0.02‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Heavy drinker	0.59 ± 0.20†	0.61 ± 0.21†	0.28 ± 0.08‡	0.29 ± 0.09‡	0.05 ± 0.02†	0.05 ± 0.02†	0.05 ± 0.02‡	0.05 ± 0.02‡
Chronic illness	1.13 ± 0.04‡	1.19 ± 0.05‡	0.44 ± 0.02‡	0.47 ± 0.03‡	0.10 ± 0.00‡	0.10 ± 0.01‡	0.07 ± 0.00‡	0.08 ± 0.01‡
Household size	-0.04 ± 0.01†	-0.11 ± 0.05†	-0.02 ± 0.01†	-0.06 ± 0.03†	-0.00 ± 0.00‡	-0.01 ± 0.00†	-0.00 ± 0.00‡	-0.01 ± 0.00*
District deprivation	-0.01 ± 0.08	-0.04 ± 0.10	-0.00 ± 0.04	-0.01 ± 0.05	-0.01 ± 0.01	-0.01 ± 0.01	-0.01 ± 0.01	-0.01 ± 0.01
District inequality	3.93 ± 1.49‡	5.60 ± 1.88†	1.98 ± 0.73†	2.74 ± 0.87†	0.27 ± 0.11†	0.38 ± 0.15†	0.26 ± 0.11†	0.39 ± 0.17†
Intercept	1.82 ± 0.09‡	1.86 ± 0.10‡	0.56 ± 0.05‡	0.55 ± 0.05‡	0.09 ± 0.01‡	0.10 ± 0.01‡		
Estimator	OLS	GMM	GMM	GMM	OLS	GMM	ML	ML
Instruments' validity		0.31		0.20		0.25		
Log(PCE)'s exogeneity		2.87*				2.18		1.26
Instruments' strength		10.10				10.10		

Note: * p < 0.10, † p < 0.05, ‡ p < 0.01.

199

Table A.5: Control function estimates

Predictors	Linear Model			Poisson Model		
	Rural	Urban	National	Rural	Urban	National
Log(PCE)	-1.41 ± 0.68†	-1.14 ± 0.79	-1.19 ± 0.59†	-0.65 ± 0.33†	-0.59 ± 0.43	-0.58 ± 0.30*
ê	1.22 ± 0.67*	0.90 ± 0.80	0.98 ± 0.59*	0.57 ± 0.33*	0.47 ± 0.43	0.48 ± 0.30
Age 25–34	0.06 ± 0.02†	-0.04 ± 0.03	0.03 ± 0.02	0.03 ± 0.01†	-0.03 ± 0.02	0.01 ± 0.01
Age 35–44	0.22 ± 0.04‡	-0.01 ± 0.05	0.13 ± 0.04‡	0.11 ± 0.02‡	-0.01 ± 0.03	0.06 ± 0.02‡
Age 45–54	0.46 ± 0.08‡	0.10 ± 0.13	0.32 ± 0.08‡	0.22 ± 0.04‡	0.06 ± 0.07	0.16 ± 0.04‡
Age 55–64	0.68 ± 0.07‡	0.18 ± 0.13	0.50 ± 0.07‡	0.30 ± 0.03‡	0.09 ± 0.07	0.23 ± 0.04‡
Age 65+	1.00 ± 0.07‡	0.46 ± 0.10‡	0.83 ± 0.07‡	0.38 ± 0.03‡	0.17 ± 0.05‡	0.31 ± 0.03‡
Female	0.60 ± 0.03‡	0.66 ± 0.05‡	0.63 ± 0.03‡	0.28 ± 0.01‡	0.35 ± 0.02‡	0.31 ± 0.01‡
Never married	-0.06 ± 0.04	-0.02 ± 0.09	-0.02 ± 0.05	-0.05 ± 0.02†	-0.02 ± 0.05	-0.03 ± 0.03
Divorced	0.26 ± 0.06‡	0.43 ± 0.07‡	0.33 ± 0.05‡	0.09 ± 0.02‡	0.16 ± 0.03‡	0.11 ± 0.02‡
Widowed	0.28 ± 0.04‡	0.17 ± 0.05†	0.25 ± 0.04‡	0.05 ± 0.01‡	0.03 ± 0.02	0.05 ± 0.01‡
Middle school	-0.09 ± 0.09	-0.24 ± 0.15	-0.15 ± 0.09*	-0.05 ± 0.05	-0.09 ± 0.08	-0.07 ± 0.05
High school	-0.04 ± 0.18	-0.13 ± 0.19	-0.13 ± 0.19	-0.04 ± 0.09	-0.10 ± 0.16	-0.07 ± 0.09
College	0.18 ± 0.33	-0.04 ± 0.47	-0.04 ± 0.32	0.07 ± 0.16	-0.08 ± 0.25	-0.04 ± 0.16
Unemployed	0.25 ± 0.04‡	0.14 ± 0.08*	0.19 ± 0.04‡	0.10 ± 0.02‡	0.06 ± 0.04	0.08 ± 0.02‡
Less physical activity	0.21 ± 0.05‡	0.26 ± 0.09†	0.22 ± 0.05‡	0.09 ± 0.02‡	0.13 ± 0.05†	0.10 ± 0.03‡
Frequent smoker	0.22 ± 0.03‡	0.28 ± 0.03‡	0.25 ± 0.03‡	0.10 ± 0.01‡	0.16 ± 0.02‡	0.13 ± 0.01‡
Heavy drinker	0.47 ± 0.12‡	0.61 ± 0.20†	0.52 ± 0.11‡	0.22 ± 0.05‡	0.29 ± 0.08‡	0.24 ± 0.05‡
Chronic illness	1.44 ± 0.06‡	1.18 ± 0.06‡	1.32 ± 0.05‡	0.49 ± 0.02‡	0.46 ± 0.03‡	0.47 ± 0.02‡
Household size	-0.13 ± 0.06†	-0.10 ± 0.06*	-0.11 ± 0.05†	-0.06 ± 0.03†	-0.05 ± 0.03*	-0.05 ± 0.02†
District deprivation	-0.01 ± 0.06	-0.03 ± 0.09	-0.02 ± 0.06	-0.00 ± 0.03	-0.02 ± 0.04	-0.01 ± 0.03
District inequality	2.76 ± 1.27†	5.58 ± 1.92†	4.18 ± 1.13‡	1.24 ± 0.58†	2.84 ± 0.97†	1.99 ± 0.53‡
Urban			0.21 ± 0.19			0.10 ± 0.10
Intercept	1.43 ± 0.11‡	1.86 ± 0.09‡	1.52 ± 0.11‡	0.42 ± 0.06‡	0.58 ± 0.05‡	0.45 ± 0.06‡
Estimator	OLS	OLS	OLS	GMM	GMM	GMM
N	363,356	214,192	577,548	363,356	214,192	577,548

Note: Standard errors are not adjusted to the fact that is a generated regressor. * p < 0.10, † p < 0.05, ‡ p < 0.01.

Table A.6: Marginal vs. Conditional model

Predictors	Linear-IV		LPM-IV	
	Marginal	Conditional	Marginal	Conditional
Log(PCE)	-1.31 ± 0.55†	-0.97 ± 0.67	-0.09 ± 0.04†	-0.09 ± 0.06
Age 25–34	0.03 ± 0.02	0.04 ± 0.02†	-0.00 ± 0.00	-0.00 ± 0.00
Age 35–44	0.14 ± 0.04‡	0.17 ± 0.03‡	0.00 ± 0.00	0.01 ± 0.00*
Age 45–54	0.34 ± 0.07‡	0.36 ± 0.08‡	0.01 ± 0.01†	0.02 ± 0.01†
Age 55–64	0.51 ± 0.07‡	0.55 ± 0.07‡	0.03 ± 0.01‡	0.03 ± 0.01‡
Age 65+	0.83 ± 0.07‡	0.90 ± 0.04‡	0.06 ± 0.01‡	0.06 ± 0.00‡
Female	0.63 ± 0.03‡	0.58 ± 0.01‡	0.05 ± 0.00‡	0.05 ± 0.00‡
Never married	-0.01 ± 0.05	-0.01 ± 0.04	0.00 ± 0.00	0.01 ± 0.00*
Divorced	0.33 ± 0.05‡	0.27 ± 0.04‡	0.03 ± 0.00‡	0.03 ± 0.00‡
Widowed	0.25 ± 0.04‡	0.28 ± 0.02‡	0.03 ± 0.00‡	0.03 ± 0.00‡
Middle school	-0.13 ± 0.09	-0.14 ± 0.08*	-0.01 ± 0.01	-0.01 ± 0.01
High school	-0.09 ± 0.17	-0.16 ± 0.17	-0.01 ± 0.01	-0.01 ± 0.01
College	0.03 ± 0.29	-0.19 ± 0.32	-0.00 ± 0.02	-0.01 ± 0.03
Unemployed	0.18 ± 0.04‡	0.22 ± 0.03‡	0.02 ± 0.00‡	0.02 ± 0.00‡
Less physical activity	0.23 ± 0.05‡	0.11 ± 0.03†	0.02 ± 0.00‡	0.01 ± 0.00‡
Frequent smoker	0.24 ± 0.02‡	0.16 ± 0.01‡	0.02 ± 0.00‡	0.01 ± 0.00‡
Heavy drinker	0.52 ± 0.11‡	0.42 ± 0.06‡	0.04 ± 0.01‡	0.03 ± 0.01‡
Chronic illness	1.33 ± 0.05‡	1.15 ± 0.03‡	0.11 ± 0.00‡	0.10 ± 0.00‡
Household size	-0.12 ± 0.04†	-0.10 ± 0.06*	-0.01 ± 0.00†	-0.01 ± 0.00*
District deprivation	-0.03 ± 0.06	-0.03 ± 0.04	-0.01 ± 0.00	-0.00 ± 0.00
District inequality	4.18 ± 1.17‡	3.54 ± 0.91‡	0.32 ± 0.10‡	0.30 ± 0.08‡
Urban	0.24 ± 0.19	0.07 ± 0.12	0.02 ± 0.01	0.01 ± 0.01
Intercept	1.50 ± 0.11†	1.66 ± 0.06‡	0.06 ± 0.01‡	0.07 ± 0.01‡
Estimator	GMM	GLS	GMM	GLS
N	577, 548	577, 548	577, 548	577, 548

Note: Sampling weight is not applied to models with random district effects. * p < 0.10, † p < 0.05, ‡ p < 0.01.

```
    ********************************************************************
    *            STATA CODE TO REPLICATE THE RESULTS OF
    *              Does poverty reduce mental health?
    *                An instrumental variable analysis
    ********************************************************************

    * DATA SOURCE

    * The National Basic Health Research (RISKESDAS) 2007:
    *        http://www.litbang.depkes.go.id/

    * The National Socio-economic Survey (SUSENAS) 2007:
    *        http://microdata.bps.go.id/

    * The Village Census (PODES) 2008:
    *        http://microdata.bps.go.id/

    * GPCC Climatology Version 2011 at 0.5 degree
    *        http://dx.doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_025

    * Spatial polygon data of administrative boundaries:
    *        http://www.gadm.org/

    ********************************************************************
    * VARIABLE DESCRIPTION
    ********************************************************************
    * stress        SRQ-20 score
    * depressed     SRQ-20 score > 6
    * JJA07AN       June-July-August 2007 precipitation anomaly
    * MAM07AN       March-April-May 2007 precipitation anomaly
    * age2534       dummy variable for age 25-34
    * age3544       dummy variable for age 35-44
    * age4554       dummy variable for age 45-54
    * age5564       dummy variable for age 55-64
    * age65up       dummy variable for age 65+
    * female        dummy variable for female
    * single        dummy variable for never married
    * divorced      dummy variable for divorced
    * widowed       dummy variable for widowed
    * compulEdu     dummy variable for junior high school
    * highSchool    dummy variable for senior high school
    * college       dummy variable for college and higher
    * unemp         dummy variable for unemployed
    * lpa           dummy variable for lack of physical activity
    * smokeDaily    dummy variable for daily smoker
    * drinkDaily    dummy variable for daily drinker
    * chronic       dummy variable for chronic illness
    * schizo        dummy variable for schizophrenia
    * hhSize        household size
    * urban         dummy for urban
    * depriv        index of deprivation
    * kabGini       Gini coefficient
    * logpce        log per capita consumption expenditure
    * BPS07         2007 district ID (440)
    * weight        sampling weight
    ********************************************************************

    * define macro for the instruments
    global rain JJA07AN MAM07AN

    * define macro for exogenous regressors
    global exog age2534 age3544 age4554 age5564 age65up ///
              female single divorced widowed ///
              compulEdu highSchool college unemp ///
              lpa smokeDaily drinkDaily chronic ///
              hhSize depriv kabGini

    global exogA age2534 age3544 age4554 age5564 age65up ///
               female single divorced widowed ///
               compulEdu highSchool college ///
```

202

```
                lpa smokeDaily drinkDaily chronic ///
                hhSize depriv kabGini

global exogB age2534 age3544 age4554 age5564 age65up ///
             female single divorced widowed ///
             compulEdu highSchool college unemp ///
             hhSize depriv kabGini

global exogC age2534 age3544 age4554 age5564 age65up ///
             female single divorced widowed ///
             compulEdu highSchool college unemp ///
             lpa smokeDaily drinkDaily chronic ///
             hhSize

global exogD age2534 age3544 age4554 age5564 age65up ///
             female single divorced widowed ///
             compulEdu highSchool college  ///
             hhSize depriv kabGini

global exogE age2534 age3544 age4554 age5564 age65up ///
             female single divorced widowed ///
             compulEdu highSchool college  ///
             hhSize

*****************************************************************
* RURAL SAMPLE
*                        >= 15 years old , no schizophrenia record
*****************************************************************

* Linear
reg stress logpce $exog                                  ///
    if urban==0 & age >=15 & schizo ==0 [pw=weight],      ///
    cluster(BPS07)

* Linear-IV
ivreg2 stress (logpce = $rain) $exog                     ///
       if urban==0 & age >=15 & schizo ==0 [pw=weight],   ///
       ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Poisson
gmm (stress - exp({xb: logpce $exog} + {b0}))            ///
    if urban==0 & age >=15 & schizo ==0 [pw=weight],      ///
    instruments(logpce $exog)                            ///
    vce(cluster BPS07)                                   ///
    derivative(/xb = -1*exp({xb:} + {b0}))               ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

* Poisson-IV
gmm (stress - exp({xb: logpce $exog} + {b0}))            ///
    if urban==0 & age >=15 & schizo ==0 [pw=weight],      ///
    instruments($exog $rain)                             ///
    vce(cluster BPS07)                                   ///
    derivative(/xb = -1*exp({xb:} + {b0}))               ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid

* LPM
reg depressed logpce $exog                               ///
    if urban==0 & age >=15 & schizo ==0 [pw=weight],      ///
    cluster(BPS07)

* LPM-IV
ivreg2 depressed (logpce = $rain) $exog                  ///
       if urban==0 & age >=15 & schizo ==0 [pw=weight],   ///
       ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Probit
probit depressed logpce $exog                            ///
       if urban==0 & age >=15 & schizo ==0 [pw=weight],   ///
       cluster(BPS07)
```

```
mfx
margins, dydx(_all) predict(p) post

* Probit-IV
ivprobit depressed (logpce = $rain) $exog                   ///
        if urban==0 & age>=15 & schizo==0 [pw=weight],      ///
        cluster(BPS07)
mfx, predict(p) eq(depressed)
margins, dydx(_all) predict(p) post

********************************************************************
* URBAN SAMPLE
*                       >= 15 years old, no schizophrenia record
********************************************************************

* Linear
reg stress logpce $exog                                     ///
    if urban==1 & age>=15 & schizo==0 [pw=weight],          ///
    cluster(BPS07)

* Linear-IV
ivreg2 stress (logpce = $rain) $exog                        ///
        if urban==1 & age>=15 & schizo==0 [pw=weight],      ///
        ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Poisson
gmm (stress - exp({xb: logpce $exog} + {b0}))               ///
    if urban==1 & age>=15 & schizo==0 [pw=weight],          ///
    instruments(logpce $exog)                               ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

* Poisson-IV
gmm (stress - exp({xb: logpce $exog} + {b0}))               ///
    if urban==1 & age>=15 & schizo==0 [pw=weight],          ///
    instruments($exog $rain)                                ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid

* LPM
reg depressed logpce $exog                                  ///
    if urban==1 & age>=15 & schizo==0 [pw=weight],          ///
    cluster(BPS07)

* LPM-IV
ivreg2 depressed (logpce = $rain) $exog                     ///
        if urban==1 & age>=15 & schizo==0 [pw=weight],      ///
        ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Probit
probit depressed logpce $exog                               ///
        if urban==1 & age>=15 & schizo==0 [pw=weight],      ///
        cluster(BPS07)
mfx
margins, dydx(_all) predict(p) post

* Probit-IV
ivprobit depressed (logpce = $rain) $exog                   ///
        if urban==1 & age>=15 & schizo==0 [pw=weight],      ///
        cluster(BPS07)
mfx, predict(p) eq(depressed)
margins, dydx(_all) predict(p) post

********************************************************************
* NATIONAL SAMPLE
*                       >= 15 years old, no schizophrenia record
********************************************************************
```

```
* Linear
reg stress logpce $exog urban                               ///
    if age >=15 & schizo==0 [pw=weight],                    ///
    cluster(BPS07)

* Linear-IV
ivreg2 stress (logpce = $rain) $exog urban                  ///
        if age >=15 & schizo==0 [pw=weight],                ///
        ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Poisson
gmm (stress - exp({xb: logpce $exog urban} + {b0}))         ///
    if age >=15 & schizo==0 [pw=weight],                    ///
    instruments(logpce $exog urban)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

* Poisson-IV
gmm (stress - exp({xb: logpce $exog urban} + {b0}))         ///
    if age >=15 & schizo==0 [pw=weight],                    ///
    instruments($exog urban $rain)                          ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid

* LPM
reg depressed logpce $exog urban                            ///
    if age >=15 & schizo==0 [pw=weight],                    ///
    cluster(BPS07)

* LPM-IV
ivreg2 depressed (logpce = $rain) $exog urban               ///
        if age >=15 & schizo==0 [pw=weight],                ///
        ffirst gmm2s cluster(BPS07) endogtest(logpce)

* Probit
probit depressed logpce $exog urban                         ///
        if age >=15 & schizo==0 [pw=weight],                ///
        cluster(BPS07)
mfx
margins, dydx(_all) predict(p) post

* Probit-IV
ivprobit depressed (logpce = $rain) $exog urban             ///
          if age >=15 & schizo==0 [pw=weight],              ///
          cluster(BPS07) first
mfx, predict(p) eq(depressed)
margins, dydx(_all) predict(p) post

****************************************************************
* CONTROL FUNCTION
*               Standard Error Not Adjusted to Two-step Estimation
****************************************************************

*** SETUP

* reduced form (rural)
reg logpce $rain $exog ///
    if urban==0 & age >=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)
predict double vhatRR, residual

* reduced form (urban)
reg logpce $rain $exog ///
    if urban==1 & age >=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)
predict double vhatUR, residual
```

```
* reduced form (national)
reg logpce urban $rain $exog ///
    if age>=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)
predict double vhatNN, residual

*** LINEAR MODEL (OLS)

* rural
reg stress logpce vhatRR $exog ///
    if urban==0 & age>=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)

* urban
reg stress logpce vhatUR $exog ///
    if urban==1 & age>=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)

* national
reg stress logpce vhatNN urban $exog ///
    if age>=15 & schizo==0 [pw=weight], ///
    cluster(BPS07)

*** POISSON MODEL (ML)

* rural
poisson stress logpce vhatRR $exog ///
        if urban==0 & age>=15 & schizo==0 [pw=weight], ///
        cluster(BPS07)

* urban
poisson stress logpce vhatUR $exog ///
        if urban==1 & age>=15 & schizo==0 [pw=weight], ///
        cluster(BPS07)

* national
poisson stress logpce vhatNN urban $exog ///
        if age>=15 & schizo==0 [pw=weight], ///
        cluster(BPS07)

*** POISSON MODEL (GMM)

* rural
gmm (stress - exp({xb: logpce vhatRR $exog} + {b0}))        ///
    if urban==0 & age>=15 & schizo==0 [pw=weight],          ///
    instruments(logpce vhatRR $exog)                        ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

* urban
gmm (stress - exp({xb: logpce vhatUR $exog} + {b0}))        ///
    if urban==1 & age>=15 & schizo==0 [pw=weight],          ///
    instruments(logpce vhatUR $exog)                        ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

* national
gmm (stress - exp({xb: logpce vhatNN urban $exog} + {b0}))  ///
    if age>=15 & schizo==0 [pw=weight],                     ///
    instruments(logpce vhatNN urban $exog)                  ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))

****************************************************************
* ROBUSTNESS CHECK NATIONAL
****************************************************************
```

```
* LINEAR-IV A
ivreg2 stress (logpce = $rain) $exogA urban                 ///
      if age >=15 & schizo ==0 [pw=weight],                 ///
      ffirst gmm2s cluster(BPS07) endogtest(logpce)


* LINEAR-IV B
ivreg2 stress (logpce = $rain) $exogB urban                 ///
      if age >=15 & schizo ==0 [pw=weight],                 ///
      ffirst gmm2s cluster(BPS07) endogtest(logpce)


* LINEAR-IV C
ivreg2 stress (logpce = $rain) $exogC urban                 ///
      if age >=15 & schizo ==0 [pw=weight],                 ///
      ffirst gmm2s cluster(BPS07) endogtest(logpce)


* LINEAR-IV D
ivreg2 stress (logpce = $rain) $exogD urban                 ///
      if age >=15 & schizo ==0 [pw=weight],                 ///
      ffirst gmm2s cluster(BPS07) endogtest(logpce)


* LINEAR-IV E
ivreg2 stress (logpce = $rain) $exogE urban                 ///
      if age >=15 & schizo ==0 [pw=weight],                 ///
      ffirst gmm2s cluster(BPS07) endogtest(logpce)


* POISSON-IV A
gmm (stress - exp({xb: logpce $exogA urban} + {b0}))        ///
    if age >=15 & schizo ==0 [pw=weight],                   ///
    instruments($exogA urban $rain)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid


* POISSON-IV B
gmm (stress - exp({xb: logpce $exogB urban} + {b0}))        ///
    if age >=15 & schizo ==0 [pw=weight],                   ///
    instruments($exogB urban $rain)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid


* POISSON-IV C
gmm (stress - exp({xb: logpce $exogC urban} + {b0}))        ///
    if age >=15 & schizo ==0 [pw=weight],                   ///
    instruments($exogC urban $rain)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid


* POISSON-IV D
gmm (stress - exp({xb: logpce $exogD urban} + {b0}))        ///
    if age >=15 & schizo ==0 [pw=weight],                   ///
    instruments($exogD urban $rain)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid


* POISSON-IV E
gmm (stress - exp({xb: logpce $exogE urban} + {b0}))        ///
    if age >=15 & schizo ==0 [pw=weight],                   ///
    instruments($exogE urban $rain)                         ///
    vce(cluster BPS07)                                      ///
    derivative(/xb = -1*exp({xb:} + {b0}))                  ///
    derivative(/b0 = -1*exp({xb:} + {b0}))
estat overid
```

```
* LPM -IV A
ivreg2 depressed (logpce = $rain) $exogA urban                ///
      if age >=15 & schizo ==0 [pw=weight],                   ///
      ffirst gmm2s cluster (BPS07) endogtest (logpce)

* LPM -IV B
ivreg2 depressed (logpce = $rain) $exogB urban                ///
      if age >=15 & schizo ==0 [pw=weight],                   ///
      ffirst gmm2s cluster (BPS07) endogtest (logpce)

* LPM -IV C
ivreg2 depressed (logpce = $rain) $exogC urban                ///
      if age >=15 & schizo ==0 [pw=weight],                   ///
      ffirst gmm2s cluster (BPS07) endogtest (logpce)

* LPM -IV D
ivreg2 depressed (logpce = $rain) $exogD urban                ///
      if age >=15 & schizo ==0 [pw=weight],                   ///
      ffirst gmm2s cluster (BPS07) endogtest (logpce)

* LPM -IV E
ivreg2 depressed (logpce = $rain) $exogE urban                ///
      if age >=15 & schizo ==0 [pw=weight],                   ///
      ffirst gmm2s cluster (BPS07) endogtest (logpce)

* PROBIT -IV A
ivprobit depressed (logpce = $rain) $exogA urban              ///
        if age >=15 & schizo ==0 [pw=weight],                 ///
        cluster (BPS07) first
mfx , predict(p) eq(depressed)
margins , dydx(_all) predict(p) post

* PROBIT -IV B
ivprobit depressed (logpce = $rain) $exogB urban              ///
        if age >=15 & schizo ==0 [pw=weight],                 ///
        cluster (BPS07) first
mfx , predict(p) eq(depressed)
margins , dydx(_all) predict(p) post

* PROBIT -IV C
ivprobit depressed (logpce = $rain) $exogC urban              ///
        if age >=15 & schizo ==0 [pw=weight],                 ///
        cluster (BPS07) first
mfx , predict(p) eq(depressed)
margins , dydx(_all) predict(p) post

* PROBIT -IV D
ivprobit depressed (logpce = $rain) $exogD urban              ///
        if age >=15 & schizo ==0 [pw=weight],                 ///
        cluster (BPS07) first
mfx , predict(p) eq(depressed)
margins , dydx(_all) predict(p) post

* PROBIT -IV E
ivprobit depressed (logpce = $rain) $exogE urban              ///
        if age >=15 & schizo ==0 [pw=weight],                 ///
        cluster (BPS07) first
mfx , predict(p) eq(depressed)
margins , dydx(_all) predict(p) post

* DISTRICT RE
xtset BPS07

* DISTRICT RE: LINEAR -IV
xtivreg stress (logpce = $rain) $exog urban                   ///
      if age >=15 & schizo ==0,                               ///
      re first theta

* DISTRICT RE: LPM -IV
```

```
xtivreg depressed (logpce = $rain) $exog urban            ///
        if age>=15 & schizo==0,                           ///
        re first theta

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

Appendix B:
Supplementary data for Chapter 3

Figure B.1: Title page of paper 2

Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis

Wulung Hanandita[1] · Gindo Tampubolon[1]

Abstract

Purpose Studies on self-rated health outcomes are fraught with problems when individuals' reporting behaviour is systematically biased by demographic, socio-economic, or cultural factors. Analysing the data drawn from the Indonesia Family Life Survey 2007, this paper aims to investigate the extent of differential health reporting behaviour by demographic and socio-economic status among Indonesians aged 40 and older ($N = 3735$).

Methods Interpersonal heterogeneity in reporting style is identified by asking respondents to rate a number of vignettes that describe varying levels of health status in targeted health domains (mobility, pain, cognition, sleep, depression, and breathing) using the same ordinal response scale that is applied to the self-report health question. A compound hierarchical ordered probit model is fitted to obtain health differences by demographic and socio-economic status. The obtained regression coefficients are then compared to the standard ordered probit model.

Results We find that Indonesians with more education tend to rate a given health status in each domain more negatively than their less-educated counterparts. Allowing for such differential reporting behaviour results in relatively stronger positive education effects.

Conclusion There is a need to correct for differential reporting behaviour using vignettes when analysing self-rated health measures in older adults in Indonesia. Unless such an adjustment is made, the salutary effect of education will be underestimated.

Keywords Self-rated health · Socio-economic status · Reporting heterogeneity · Anchoring vignette · Indonesia

Introduction

Both resource constraints and the multidimensionality of health concepts being studied often necessitate the collection of self-rated health (SRH) data. SRH measures, which ask individuals to report their health status either in general or on a specific health domain using an ordinal response scale, require no specialist intervention during data collection, are relatively cheap and quick to obtain, and are feasible to implement in large-scale surveys. In addition to the belief that SRH can capture aspects of health that cannot be tapped by objective measure [35], research has shown that SRH is highly correlated with assessments provided by health professionals [9] and that is also a strong predictor of mortality [15] as well as health care utilisation [30].

Notwithstanding these benefits, the use of SRH in the study of socio-economic inequalities in health becomes fraught with serious problems when individuals have different expectations, knowledge, or standards of what constitutes a good health. For example, when experiencing an identically severe health problem, poor individuals may paradoxically report better health than their richer counterparts (Fig. 1) simply because the poor have a much higher tolerance to health problems than the rich [28]. This is known in the literature as 'reporting heterogeneity' [29], 'differential item functioning' [19], 'response

✉ Wulung Hanandita
wulung.hanandita@postgrad.manchester.ac.uk

Gindo Tampubolon
gindo.tampubolon@manchester.ac.uk

[1] Cathie Marsh Institute for Social Research, University of Manchester, Humanities Bridgeford Street Building 2F, Oxford Road, Manchester M13 9PL, UK

⌂ Springer

Self-report health question

1. Overall in the last 30 days, how much of a problem did [name of person/you] have with moving around?

2. Overall in the last 30 days, how much of bodily aches or pains did you have?

3. Overall in the last 30 days overall how much difficulty did you have with remembering things?

4. In the last 30 days, how much difficulty did you have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?

5. Overall in the last 30 days, how much of a problem did you have with feeling sad, low, or depressed?

6. In the last 30 days, how much of a problem did you have because of shortness of breath?

Vignette description

Mobility

1. Pak Taryono/Bu Taryini is able to walk distances of up to 200 metres without any problems but feels tired after walking one kilometer. He has no problems with day-to-day activities, such as carrying food from the market.

2. Pak Tumino/Bu Tumini does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry the groceries and do some light household work.

3. Pak Sidik/Bu Endah has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy.

Pain

1. Pak Budiarto/Bu Budiarti has a headache once a month that is relieved after taking a pill. During the headache she can carry on with her day-to-day affairs.

2. Pak Sumarno/Bu Sumarni has pain that radiates down her right arm and wrist during her day at work. This is slightly relieved in the evenings when she is no longer working on her computer.

3. Pak Mulyono/Bu Mulyanti has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things.

Cognition

1. Pak Taryono/Bu Taryini can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week he forgets where his keys or glasses are, but finds them within five minutes.

2. Pak Suwarso/Bu Suwarsih is keen to learn new recipes but finds that she often makes mistakes and has to reread several times before she is able to do them properly.

3. Pak Mugiono/Bu Mugianti cannot concentrate for more than 15 minutes and has difficulty paying attention to what is being said to him. Whenever he starts a task, he never manages to finish it and often forgets what he was doing. He is able to learn the names of people he meets.

Sleep

1. Pak Partono/Bu Partini falls asleep easily at night, but two nights a week she wakes up in the middle of the night and cannot go back to sleep for the rest of the night.

2. Pak Darma/Bu Darmi wakes up almost once every hour during the night. When he wakes up in the night, it takes around 15 minutes for him to go back to sleep. In the morning he does not feel well-rested.

3. Pak Parto/Bu Parti takes about two hours every night to fall asleep. He wakes up once or twice a night feeling panicked and takes more than one hour to fall asleep again.

Depression

1. Pak Arman/Bu Lina enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities.

2. Pak Sukarso/Bu Sukarsih feels nervous and anxious. He worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests him. When he is alone he tends to feel useless and empty.

3. Pak Rano/Bu Rina feels depressed most of the time. She weeps frequently and feels hopeless about the future. She feels that she has become a burden on others and that she would be better dead.

Breathing

1. Pak Sugiarto/Bu Suwarsih has no problems while walking slowly. He gets out of breath easily when climbing uphill for 20 meters or a flight of stairs.

2. Pak Ramlan/Bu Badriah suffers from respiratory infections about once every year. He is short of breath 3 or 4 times a week and had to be admitted in hospital twice in the past month with a bad cough that required treatment with antibiotics.

3. Pak Hamid/Bu Karsini has been a heavy smoker for 30 years and wakes up with a cough every morning. He gets short of breath even while resting and does not leave the house anymore. He often needs to be put on oxygen.

```
#############################################################################
#                      R CODE TO REPLICATE THE RESULTS OF
#             Does reporting behaviour bias the measurement of
#          social inequalities in self-rated health in Indonesia?
#############################################################################

# DATA SOURCE

# The Indonesia Family Life Survey (IFLS) IV, 2007:
#        http://www.rand.org/labor/FLS/IFLS.html


#---------------------------------------------------------------------------#
# VARIABLE DESCRIPTION
#---------------------------------------------------------------------------#
# srMOB        self-rated health measure of mobility
# srPAI        self-rated health measure of pain
# srCOG        self-rated health measure of cognition
# srSLE        self-rated health measure of sleep
# srDEP        self-rated health measure of depression
# srBRE        self-rated health measure of breathing
# age50s       dummy variable for age 50-59
# age60s       dummy variable for age 60-69
# age70up      dummy variable for age 70+
# female       dummy variable for female
# unmarried    dummy variable for not in marriage
# bigFamily    household size > 4
# educated     dummy variable for the completion of junior high school
# logassetPC   log per capita household assets value
# urban        dummy variable for urban
# vgMOB        vignette rating of mobility
# vgPAI        vignette rating of pain
# vgCOG        vignette rating of cognition
# vgSLE        vignette rating of sleep
# vgDEP        vignette rating of depression
# vgBRE        vignette rating of breathing
#---------------------------------------------------------------------------#


# library
require(foreign)
require(ZeligChoice)
require(anchors)

# set formula
MOBmodel <- srMOB ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

PAImodel <- srPAI ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

COGmodel <- srCOG ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

SLEmodel <- srSLE ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

DEPmodel <- srDEP ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

BREmodel <- srBRE ~ age50s + age60s + age70up + female + unmarried +
                    bigFamily + educated + logassetPC + urban

TAUmodel <- ~ age50s + age60s + age70up + female + unmarried +
              bigFamily + educated + logassetPC + urban


#---------------------------------------------------------------------------#
# OPROBIT - IGNORING REPORTING HETEROGENEITY - NAIVE ANALYSIS
#---------------------------------------------------------------------------#

# fit oprobit self report
oprobitMOB <- zelig(formula = MOBmodel, model = "oprobit", data = VGDATA)
```

```
oprobitPAI <- zelig(formula = PAImodel , model = "oprobit", data = VGDATA)
oprobitCOG <- zelig(formula = COGmodel , model = "oprobit", data = VGDATA)
oprobitSLE <- zelig(formula = SLEmodel , model = "oprobit", data = VGDATA)
oprobitDEP <- zelig(formula = DEPmodel , model = "oprobit", data = VGDATA)
oprobitBRE <- zelig(formula = BREmodel , model = "oprobit", data = VGDATA)

#-----------------------------------------------------------------------------#
# CHOPIT - ACCOUNTING FOR REPORTING HETEROGENEITY - ALL VIGNETTES
#-----------------------------------------------------------------------------#

# set CHOPIT formula
fMOB <- list(self = MOBmodel ,
             vign = cbind(vgMOB1 , vgMOB2 , vgMOB3) ~ 1,
             tau  = TAUmodel)

fPAI <- list(self = PAImodel ,
             vign = cbind(vgPAI1 , vgPAI2 , vgPAI3) ~ 1,
             tau  = TAUmodel)

fCOG <- list(self = COGmodel ,
             vign = cbind(vgCOG1 , vgCOG2 , vgCOG3) ~ 1,
             tau  = TAUmodel)

fSLE <- list(self = SLEmodel ,
             vign = cbind(vgSLE1 , vgSLE2 , vgSLE3) ~ 1,
             tau  = TAUmodel)

fDEP <- list(self = DEPmodel ,
             vign = cbind(vgDEP1 , vgDEP2 , vgDEP3) ~ 1,
             tau  = TAUmodel)

fBRE <- list(self = BREmodel ,
             vign = cbind(vgBRE1 , vgBRE2 , vgBRE3) ~ 1,
             tau  = TAUmodel)

# fit CHOPIT
chopitML.MOB <- chopit(formula = fMOB, data = VGDATA,
                       options = anchors.options(use.linear = TRUE))
chopitGN.MOB <- chopit(formula = fMOB, data = VGDATA,
                       options = anchors.options(use.linear = TRUE,
                                                 optimizer = "genoud",
                                                 start = chopitML.MOB$parm,
                                                 print.level = 1))

chopitML.PAI <- chopit(formula = fPAI, data = VGDATA,
                       options = anchors.options(use.linear = TRUE))
chopitGN.PAI <- chopit(formula = fPAI, data = VGDATA,
                       options = anchors.options(use.linear = TRUE,
                                                 optimizer = "genoud",
                                                 start = chopitML.PAI$parm,
                                                 print.level = 1))

chopitML.COG <- chopit(formula = fCOG, data = VGDATA,
                       options = anchors.options(use.linear = TRUE))
chopitGN.COG <- chopit(formula = fCOG, data = VGDATA,
                       options = anchors.options(use.linear = TRUE,
                                                 optimizer = "genoud",
                                                 start = chopitML.COG$parm,
                                                 print.level = 1))

chopitML.SLE <- chopit(formula = fSLE, data = VGDATA,
                       options = anchors.options(use.linear = TRUE))
chopitGN.SLE <- chopit(formula = fSLE, data = VGDATA,
                       options = anchors.options(use.linear = TRUE,
                                                 optimizer = "genoud",
                                                 start = chopitML.SLE$parm,
                                                 print.level  = 1))

chopitML.DEP <- chopit(formula = fDEP, data = VGDATA,
                       options = anchors.options(use.linear = T))
```

```
chopitGN.DEP <- chopit(formula = fDEP, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.DEP$parm,
                                                  print.level = 1))

chopitML.BRE <- chopit(formula = fBRE, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.BRE <- chopit(formula = fBRE, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.BRE$parm,
                                                  print.level = 1))

#-----------------------------------------------------------------------------#
# CHOPIT - SINGLE VIGNETTE
#-----------------------------------------------------------------------------#

# set CHOPIT formula
fMOB1 <- list(self = MOBmodel,
              vign = cbind(vgMOB1) ~ 1,
              tau  = TAUmodel)

fMOB2 <- list(self = MOBmodel,
              vign = cbind(vgMOB2) ~ 1,
              tau  = TAUmodel)

fMOB3 <- list(self = MOBmodel,
              vign = cbind(vgMOB3) ~ 1,
              tau  = TAUmodel)

fPAI1 <- list(self = PAImodel,
              vign = cbind(vgPAI1) ~ 1,
              tau  = TAUmodel)

fPAI2 <- list(self = PAImodel,
              vign = cbind(vgPAI2) ~ 1,
              tau  = TAUmodel)

fPAI3 <- list(self = PAImodel,
              vign = cbind(vgPAI3) ~ 1,
              tau  = TAUmodel)

fCOG1 <- list(self = COGmodel,
              vign = cbind(vgCOG1) ~ 1,
              tau  = TAUmodel)

fCOG2 <- list(self = COGmodel,
              vign = cbind(vgCOG2) ~ 1,
              tau  = TAUmodel)

fCOG3 <- list(self = COGmodel,
              vign = cbind(vgCOG3) ~ 1,
              tau  = TAUmodel)

fSLE1 <- list(self = SLEmodel,
              vign = cbind(vgSLE1) ~ 1,
              tau  = TAUmodel)

fSLE2 <- list(self = SLEmodel,
              vign = cbind(vgSLE2) ~ 1,
              tau  = TAUmodel)

fSLE3 <- list(self = SLEmodel,
              vign = cbind(vgSLE3) ~ 1,
              tau  = TAUmodel)

fDEP1 <- list(self = DEPmodel,
              vign = cbind(vgDEP1) ~ 1,
              tau  = TAUmodel)
```

```
fDEP2 <- list(self = DEPmodel,
              vign = cbind(vgDEP2) ~ 1,
              tau  = TAUmodel)


fDEP3 <- list(self = DEPmodel,
              vign = cbind(vgDEP3) ~ 1,
              tau  = TAUmodel)


fBRE1 <- list(self = BREmodel,
              vign = cbind(vgBRE1) ~ 1,
              tau  = TAUmodel)


fBRE2 <- list(self = BREmodel,
              vign = cbind(vgBRE2) ~ 1,
              tau  = TAUmodel)


fBRE3 <- list(self = BREmodel,
              vign = cbind(vgBRE3) ~ 1,
              tau  = TAUmodel)


# fit chopit
chopitML.MOB1 <- chopit(formula = fMOB1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.MOB1 <- chopit(formula = fMOB1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.MOB1$parm,
                                                  print.level = 1))


chopitML.MOB2 <- chopit(formula = fMOB2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.MOB2 <- chopit(formula = fMOB2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.MOB2$parm,
                                                  print.level = 1))


chopitML.MOB3 <- chopit(formula = fMOB3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.MOB3 <- chopit(formula = fMOB3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.MOB3$parm,
                                                  print.level = 1))


chopitML.PAI1 <- chopit(formula = fPAI1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.PAI1 <- chopit(formula = fPAI1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.PAI1$parm,
                                                  print.level = 1))


chopitML.PAI2 <- chopit(formula = fPAI2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.PAI2 <- chopit(formula = fPAI2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.PAI2$parm,
                                                  print.level = 1))


chopitML.PAI3 <- chopit(formula = fPAI3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.PAI3 <- chopit(formula = fPAI3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.PAI3$parm,
                                                  print.level = 1))
```

```
chopitML.COG1 <- chopit(formula = fCOG1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.COG1 <- chopit(formula = fCOG1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.COG1$parm,
                                                  print.level = 1))

chopitML.COG2 <- chopit(formula = fCOG2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.COG2 <- chopit(formula = fCOG2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.COG2$parm,
                                                  print.level = 1))

chopitML.COG3 <- chopit(formula = fCOG3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.COG3 <- chopit(formula = fCOG3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.COG3$parm,
                                                  print.level = 1))

chopitML.SLE1 <- chopit(formula = fSLE1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.SLE1 <- chopit(formula = fSLE1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.SLE1$parm,
                                                  print.level = 1))

chopitML.SLE2 <- chopit(formula = fSLE2, data = VGDATA,
                        options = anchors.options(use.linear = T))
chopitGN.SLE2 <- chopit(formula = fSLE2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.SLE2$parm,
                                                  print.level = 1))

chopitML.SLE3 <- chopit(formula = fSLE3, data = VGDATA,
                        options = anchors.options(use.linear = T))
chopitGN.SLE3 <- chopit(formula = fSLE3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.SLE3$parm,
                                                  print.level = 1))

chopitML.DEP1 <- chopit(formula = fDEP1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.DEP1 <- chopit(formula = fDEP1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.DEP1$parm,
                                                  print.level = 1))

chopitML.DEP2 <- chopit(formula = fDEP2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.DEP2 <- chopit(formula = fDEP2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.DEP2$parm,
                                                  print.level = 1))

chopitML.DEP3 <- chopit(formula = fDEP3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.DEP3 <- chopit(formula = fDEP3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.DEP3$parm,
```

```
                                                        print.level = 1))

chopitML.BRE1 <- chopit(formula = fBRE1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.BRE1 <- chopit(formula = fBRE1, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  max.generations = 512,
                                                  start = chopitML.BRE1$parm,
                                                  print.level = 1))

chopitML.BRE2 <- chopit(formula = fBRE2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.BRE2 <- chopit(formula = fBRE2, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.BRE2$parm,
                                                  print.level = 1))

chopitML.BRE3 <- chopit(formula = fBRE3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE))
chopitGN.BRE3 <- chopit(formula = fBRE3, data = VGDATA,
                        options = anchors.options(use.linear = TRUE,
                                                  optimizer = "genoud",
                                                  start = chopitML.BRE3$parm,
                                                  print.level = 1))


*********************************************************************
*               STATA CODE TO REPLICATE THE RESULTS OF
*          Does reporting behaviour bias the measurement of
*       social inequalities in self-rated health in Indonesia?
*********************************************************************

* DATA SOURCE

* The Indonesia Family Life Survey (IFLS) IV, 2007:
*          http://www.rand.org/labor/FLS/IFLS.html


*********************************************************************
* VARIABLE DESCRIPTION
*********************************************************************
* Same with previous R code, just one addition:
* pidlink          individual ID
*********************************************************************

* define macro
global xvars age50s age60s age70up female unmarried bigFamily ///
              educated logassetPC urban


*********************************************************************
* CHOPIT MODEL using GLLAMM
* Note: Run this code for each health domain
*********************************************************************

* reshape long
ren srMOB vgMOB4
ds
reshape long vgMOB, i(pidlink) j(item)

* drop missing values
drop if vgMOB==.

* list
list pidlink item vgMOB in 1/10

* generate dummies
tab item, gen(i)

* sort
sort pidlink item
```

220

```
* list
list pidlink item i1-i4 in 1/10

* prepare GLLAMM
rename i4 self
for var $xvars: gen s_X = self*X
global thexvars s_age50s s_age60s s_age70up s_female ///
                s_unmarried s_bigFamily s_educated ///
                s_logassetPC s_urban

gen vign = 1-self
eq het: vign self
constraint def 1 [lns1]self=0
eq thresh: $xvars

destring pidlink, gen(id)

* fit CHOPIT
gllamm vgMOB $thexvars i1 i2 i3, ///
        i(id) link(soprobit) s(het) constr(1) ethresh(thresh) ///
        init
est store chopitMOB

* test of reporting homogeneity
* H0: all gammas equal to zero or simply there is no DIF
test ([_cut11])([_cut12])([_cut13])([_cut14])

* test of reporting homogeneity for each variable
* H0: gamma X equal to zero or simply there is no DIF by X
test [_cut11]age50s      [_cut12]age50s ///
     [_cut13]age50s      [_cut14]age50s
test [_cut11]age60s      [_cut12]age60s ///
     [_cut13]age60s      [_cut14]age60s
test [_cut11]age70up     [_cut12]age70up ///
     [_cut13]age70up     [_cut14]age70up
test [_cut11]female      [_cut12]female ///
     [_cut13]female      [_cut14]female
test [_cut11]unmarried   [_cut12]unmarried ///
     [_cut13]unmarried   [_cut14]unmarried
test [_cut11]bigFamily   [_cut12]bigFamily ///
     [_cut13]bigFamily   [_cut14]bigFamily
test [_cut11]educated    [_cut12]educated ///
     [_cut13]educated    [_cut14]educated
test [_cut11]logassetPC [_cut12]logassetPC ///
     [_cut13]logassetPC [_cut14]logassetPC
test [_cut11]urban       [_cut12]urban ///
     [_cut13]urban       [_cut14]urban

* test of parallel shift
* H0: each gamma is equal across all cut-points
test [_cut11 = _cut12 = _cut13 = _cut14]

* test of parallel shift for each variable
* H0: gamma X is the same in all cut-points
test [_cut11]age50s      = [_cut12]age50s = ///
     [_cut13]age50s      = [_cut14]age50s
test [_cut11]age60s      = [_cut12]age60s = ///
     [_cut13]age60s      = [_cut14]age60s
test [_cut11]age70up     = [_cut12]age70up = ///
     [_cut13]age70up     = [_cut14]age70up
test [_cut11]female      = [_cut12]female = ///
     [_cut13]female      = [_cut14]female
test [_cut11]unmarried   = [_cut12]unmarried = ///
     [_cut13]unmarried   = [_cut14]unmarried
test [_cut11]bigFamily   = [_cut12]bigFamily = ///
     [_cut13]bigFamily   = [_cut14]bigFamily
test [_cut11]educated    = [_cut12]educated = ///
     [_cut13]educated    = [_cut14]educated
test [_cut11]logassetPC = [_cut12]logassetPC = ///
```

```
      [_cut13]logassetPC = [_cut14]logassetPC
test [_cut11]urban      = [_cut12]urban = ///
      [_cut13]urban      = [_cut14]urban


**********************************************************************
*        CHOPIT PROGRAM using MAXIMUM LIKELIHOOD EVALUATOR
**********************************************************************

* CHOPIT in Stata
* Adapted from Applied Health Economics, 2nd Ed. (2013)

* PROGRAM FOR CHOPIT, joint estimation of vignette component and
* own health model with cutpts determined by vignette component
* 1. CUTPTS FUNCTIONS OF INDIVIDUAL CHARACTERISTICS,
*    HEALTH INDEX FUNCTION OF VIGNETTE DUMMIES
* 2. CUTPTS DEFINED BY VIGNETTE COMPONENT,
*    HEALTH INDEX FUNCTION OF INDIVIDUAL CHARACTERISTICS

* IMPORTANT:
* this program parameterises tau_k+1 = tau_k + xb
* which differs from King et al. (2004) parameterisation
* original CHOPIT uses tau_k+1 = tau_k + exp(xb)
* in order to guarantee that the threshold is positive
* this program normalises the scale to the vignette components
* therefore the beta are not comparable with R or GLLAMM
*           the beta are not comparable with oprobit if sigma!=1
* R has more options and less pain in data manipulation :)

* change delimiter
#delimit ;

* define program
cap program drop chopit;
program define chopit;
args lnf b s
        b_2 b_3
        m1 m2 m3 m4 ;

        tempvar  b_1   p1_1 p2_1 p3_1 p4_1 p5_1
                 p1_2 p2_2 p3_2 p4_2 p5_2
                 p1_3 p2_3 p3_3 p4_3 p5_3
                 p1   p2   p3   p4   p5;
quietly {;

        gen double `p1_1' = 0; gen double `p2_1' = 0;
        gen double `p3_1' = 0; gen double `p4_1' = 0;
        gen double `p5_1' = 0;
        gen double `p1_2' = 0; gen double `p2_2' = 0;
        gen double `p3_2' = 0; gen double `p4_2' = 0;
        gen double `p5_2' = 0;
        gen double `p1_3' = 0; gen double `p2_3' = 0;
        gen double `p3_3' = 0; gen double `p4_3' = 0;
        gen double `p5_3' = 0;

        gen double `p1' = 0; gen double `p2' = 0;
        gen double `p3' = 0; gen double `p4' = 0;
        gen double `p5' = 0;

        gen double `b_1' = 0;

        replace `p1_1' = normal(`m1'-`b_1');
        replace `p2_1' = normal(`m2'-`b_1') - normal(`m1'-`b_1');
        replace `p3_1' = normal(`m3'-`b_1') - normal(`m2'-`b_1');
        replace `p4_1' = normal(`m4'-`b_1') - normal(`m3'-`b_1');
        replace `p5_1' = 1 - normal(`m4'-`b_1');

        replace `p1_2' = normal(`m1'-`b_2');
        replace `p2_2' = normal(`m2'-`b_2') - normal(`m1'-`b_2');
        replace `p3_2' = normal(`m3'-`b_2') - normal(`m2'-`b_2');
        replace `p4_2' = normal(`m4'-`b_2') - normal(`m3'-`b_2');
```

```
            replace 'p5_2' = 1 - normal('m4'-'b_2');

            replace 'p1_3' = normal('m1'-'b_3');
            replace 'p2_3' = normal('m2'-'b_3') - normal('m1'-'b_3');
            replace 'p3_3' = normal('m3'-'b_3') - normal('m2'-'b_3');
            replace 'p4_3' = normal('m4'-'b_3') - normal('m3'-'b_3');
            replace 'p5_3' = 1 - normal('m4'-'b_3');


            replace 'p1' = normal(('m1'-'b')/'s') ;
            replace 'p2' = normal(('m2'-'b')/'s') - normal(('m1'-'b')/'s');
            replace 'p3' = normal(('m3'-'b')/'s') - normal(('m2'-'b')/'s');
            replace 'p4' = normal(('m4'-'b')/'s') - normal(('m3'-'b')/'s');
            replace 'p5' = 1 - normal(('m4'-'b')/'s');

            replace 'lnf' = ((vig1==1)*ln('p1_1') + (vig1==2)*ln('p2_1') +
                        (vig1==3)*ln('p3_1') + (vig1==4)*ln('p4_1') +
                        (vig1==5)*ln('p5_1')) +
                        ((vig2==1)*ln('p1_2') + (vig2==2)*ln('p2_2') +
                        (vig2==3)*ln('p3_2') + (vig2==4)*ln('p4_2') +
                        (vig2==5)*ln('p5_2')) +
                        ((vig3==1)*ln('p1_3') + (vig3==2)*ln('p2_3') +
                        (vig3==3)*ln('p3_3') + (vig3==4)*ln('p4_3') +
                        (vig3==5)*ln('p5_3')) +
                        ((y==1)*ln('p1') + (y==2)*ln('p2') +
                        (y==3)*ln('p3') + (y==4)*ln('p4') +
                        (y==5)*ln('p5'));
        };
end;


* return delimiter to default
#delimit cr


********************************************************************
* NAIVE OPROBIT
* Note: Run this code for each health domain
********************************************************************


* fit naive oprobit
oprobit srMOB $xvars, nolog
est store opro_MOB

* calculate marginal effects, P(very good health)
quietly ///
nlcom (PE5_age50s: normal(_b[/cut4]-_b[logassetPC]*15) -     ///
                normal(_b[/cut4]-_b[logassetPC]*15 -        ///
                    _b[age50s]))                            ///
    (PE5_age60s: normal(_b[/cut4]-_b[logassetPC]*15) -     ///
                normal(_b[/cut4]-_b[logassetPC]*15 -        ///
                    _b[age60s]))                            ///
    (PE5_age70up: normal(_b[/cut4]-_b[logassetPC]*15) -    ///
                 normal(_b[/cut4]-_b[logassetPC]*15 -        ///
                     _b[age70up]))                          ///
    (PE5_female: normal(_b[/cut4]-_b[logassetPC]*15) -     ///
                normal(_b[/cut4]-_b[logassetPC]*15 -        ///
                    _b[female]))                            ///
    (PE5_unmarried: normal(_b[/cut4]-_b[logassetPC]*15) -  ///
                    normal(_b[/cut4]-_b[logassetPC]*15 -    ///
                        _b[unmarried]))                     ///
    (PE5_bigFamily: normal(_b[/cut4]-_b[logassetPC]*15) -  ///
                    normal(_b[/cut4]-_b[logassetPC]*15 -    ///
                        _b[bigFamily]))                     ///
    (PE5_educated: normal(_b[/cut4]-_b[logassetPC]*15) -   ///
                    normal(_b[/cut4]-_b[logassetPC]*15 -    ///
                        _b[educated]))                      ///
    (PE5_logassetPC: normal(_b[/cut4]-_b[logassetPC]*15) - ///
                    normal(_b[/cut4]-_b[logassetPC]*15 -    ///
                        _b[logassetPC]))                    ///
    (PE5_urban: normal(_b[/cut4]-_b[logassetPC]*15) -      ///
                normal(_b[/cut4]-_b[logassetPC]*15 -        ///
```

```
                              _b[urban])), post
nlcom


*******************************************************************
* CHOPIT using MAXIMUM LIKELIHOOD EVALUATOR
* Note: Run this code for each health domain
*******************************************************************

* prepare dataset
clonevar     y = srMOB
clonevar vig1 = vgMOB1
clonevar vig2 = vgMOB2
clonevar vig3 = vgMOB3

gen vigdum1 = (vgMOB1!=.)
gen vigdum2 = (vgMOB2!=.)
gen vigdum3 = (vgMOB3!=.)

* set up the model
ml model lf chopit (xb: $xvars)   ///
                   (sig:)         ///
                   (vigdum2:)     ///
                   (vigdum1:)     ///
                   (mu1: $xvars)  ///
                   (mu2: $xvars)  ///
                   (mu3: $xvars)  ///
                   (mu4: $xvars)

* fit the model
ml search
ml maximize
est store cho1_MOB

* calculate marginal effects, P(very good health)
quietly ///
nlcom (PE5_age50s: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15-_b[xb:age50s])/_b[sig:_cons])) ///
      (PE5_age60s: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15-_b[xb:age60s])/_b[sig:_cons])) ///
      (PE5_age70up: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15-_b[xb:age70up])/_b[sig:_cons])) ///
      (PE5_female: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15-_b[xb:female])/_b[sig:_cons])) ///
      (PE5_unmarried: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15-_b[xb:unmarried])/_b[sig:_cons])) ///
      (PE5_bigFamily: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                          15)/_b[sig:_cons]) -              ///
                   normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                          15-_b[xb:_cons]-_b[xb:logassetPC]* ///
```

```
                                  15-_b[xb:bigFamily])/_b[sig:_cons])) ///
              (PE5_educated: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15)/_b[sig:_cons]) -                  ///
                    normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15-_b[xb:educated])/_b[sig:_cons])) ///
              (PE5_logassetPC: normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15)/_b[sig:_cons]) -                  ///
                    normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15-_b[xb:logassetPC])/_b[sig:_cons])) ///
              (PE5_urban: normal((_b[mu4:_cons]+_b[mu4:logassetPC]*  ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15)/_b[sig:_cons]) -                  ///
                    normal((_b[mu4:_cons]+_b[mu4:logassetPC]* ///
                             15-_b[xb:_cons]-_b[xb:logassetPC]* ///
                             15-_b[xb:urban])/_b[sig:_cons])), ///
         post
nlcom

* test of reporting homogeneity
est restore cho1_MOB
test([mu1])         ([mu2])         ([mu3])         ([mu4])
test [mu1]age50s    [mu2]age50s    [mu3]age50s    [mu4]age50s
test [mu1]age60s    [mu2]age60s    [mu3]age60s    [mu4]age60s
test [mu1]age70up   [mu2]age70up   [mu3]age70up   [mu4]age70up
test [mu1]female    [mu2]female    [mu3]female    [mu4]female
test [mu1]unmarried [mu2]unmarried [mu3]unmarried [mu4]unmarried
test [mu1]bigFamily [mu2]bigFamily [mu3]bigFamily [mu4]bigFamily
test [mu1]educated  [mu2]educated  [mu3]educated  [mu4]educated
test [mu1]logassetPC [mu2]logassetPC [mu3]logassetPC [mu4]logassetPC
test [mu1]urban     [mu2]urban     [mu3]urban     [mu4]urban

* test of parallel cut-point shift
est restore cho1_MOB
test [mu1           = mu2          = mu3          = mu4]
test [mu1]age50s    =[mu2]age50s   =[mu3]age50s   =[mu4]age50s
test [mu1]age60s    =[mu2]age60s   =[mu3]age60s   =[mu4]age60s
test [mu1]age70up   =[mu2]age70up  =[mu3]age70up  =[mu4]age70up
test [mu1]female    =[mu2]female   =[mu3]female   =[mu4]female
test [mu1]unmarried =[mu2]unmarried =[mu3]unmarried =[mu4]unmarried
test [mu1]bigFamily =[mu2]bigFamily =[mu3]bigFamily =[mu4]bigFamily
test [mu1]educated  =[mu2]educated  =[mu3]educated  =[mu4]educated
test [mu1]logassetPC=[mu2]logassetPC=[mu3]logassetPC=[mu4]logassetPC
test [mu1]urban     =[mu2]urban    =[mu3]urban    =[mu4]urban


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

Appendix C:
Supplementary data for Chapter 4

Figure C.1: Title page of paper 3

Contents lists available at ScienceDirect

SSM -Population Health

journal homepage: www.elsevier.com/locate/ssmph

Article

The double burden of malnutrition in Indonesia: Social determinants and geographical variations

Wulung Hanandita *, Gindo Tampubolon

Cathie Marsh Institute for Social Research, University of Manchester, Humanities Bridgeford Street Building 2F, Oxford Road, Manchester M13 9PL, United Kingdom

ABSTRACT

The presence of simultaneous under- and over-nutrition has been widely documented in low- and middle-income countries, but global nutritional research has seen only a few large-scale population studies from Indonesia. We investigate the social determinants as well as the geographical variations of under- and over-nutrition in Indonesia using the largest public health study ever conducted in the country, the National Basic Health Research 2007 (N=645,032). Multilevel multinomial logistic regression and quantile regression models are fitted to estimate the association between nutritional status and a number of socio-economic indicators at both the individual and district levels. We find that: (1) education and income reduce the odds of being underweight by 10–30% but at the same time increase those of overweight by 10–40%; (2) independent from the compositional effect of poverty, income inequality is detrimental to population health: a 0.1 increase in the Gini coefficient is associated with an 8–12% increase in the odds of an individual's being both under- and overweight; and (3) the effects that these determinants have upon nutritional status are not necessarily homogeneous along the continuum of body mass index. Equally important, our analysis reveals that there is substantial spatial clustering of areas with elevated risk of under- or over-nutrition across the 17,000-island archipelago. As of 2007, under-nutrition in Indonesia remains a 'disease of poverty', while over-nutrition is one of affluence. The income inequality accompanying Indonesia's economic growth may aggravate the dual burden of under- and over-nutrition. A more equitable economic policy and a policy that improves living standards may be effective for addressing the double burden.

1. Introduction

The simultaneous presence of under- and over-nutrition within populations of developing countries undergoing rapid economic transition has been widely documented (Gillespie & Haddad, 2003; Jehn & Brewis, 2009). The changes in dietary intake patterns and leisure-time activities associated with industrialisation and urbanisation are known to have contributed to an increased prevalence of obesity in numerous countries (Popkin, 1998, 1999); at the same time, the problem of under-nutrition remains undefeated. This dual burden, which may also exist within a single household (Doak, Adair, Bentley, Monteiro, & Popkin, 2005; Lee, Houser, Must, de Fulladolsa, & Bermudez, 2012), is costly for the health as well as the economy of a nation. Under-nutrition impairs cognition (Sandjaja et al., 2013) and physical development (Mani, 2012), reduces economic productivity (Victora et al., 2008), raises

the mortality rate, and even induces an intergenerational cycle of malnutrition (Barker, 1997); on the other extreme of the nutritional spectrum, over-nutrition is known to increase the risk of non-communicable diseases, inflate health care costs (Cawley & Meyerhoefer, 2012; Withrow & Alter, 2011), and reduce overall quality of life.

The body of nutritional epidemiology and development economics research suggests that, over and above the biological aspects of age and sex, socio-economic status, along with a number of ecological factors such as urban environment, area-level economic development and income inequality, seems to consistently determine the social distribution of malnutrition (Doak et al., 2005; Ha et al., 2011; Lee et al., 2012; Rahmanian et al., 2014; Roemling & Qaim, 2013; Shafique et al., 2007; Subramanian, Kawachi, & Smith, 2007; Vaezghasemi et al., 2014). Notwithstanding the increasing number of studies in this stream of research, the literature, however, does not yet include sufficient evidence from Indonesia, which is the most populous developing country after China and India. To date, empirical evidence tends to come from South Asia, Africa and Latin America (see for example

* Corresponding author.
E-mail addresses: wulung.hanandita@postgrad.manchester.ac.uk (W. Hanandita), gindo.tampubolon@manchester.ac.uk (G. Tampubolon).

```
*********************************************************************
*              STATA CODE TO REPLICATE THE RESULTS OF
*         The double burden of malnutrition in Indonesia:
*         Social determinants and geographical variations
*********************************************************************

* DATA SOURCE

* The National Basic Health Research (RISKESDAS) 2007:
*         http://www.litbang.depkes.go.id/

* The National Socio-economic Survey (SUSENAS) 2007:
*         http://microdata.bps.go.id/

* The Village Census (PODES) 2008:
*         http://microdata.bps.go.id/

* Spatial polygon data of administrative boundaries:
*         http://www.gadm.org/

*********************************************************************
* VARIABLE DESCRIPTION
*********************************************************************
* BMI           body mass index
* BMIcat        indicator of nutritional status
* BMIsamp       dummy variable for sample inclusion
* age2534       dummy variable for age 25-34
* age3544       dummy variable for age 35-44
* age4554       dummy variable for age 45-54
* age5564       dummy variable for age 55-64
* age65up       dummy variable for age 65+
* female        dummy variable for female
* single        dummy variable for never married
* divorced      dummy variable for divorced
* widowed       dummy variable for widowed
* compulEdu     dummy variable for junior high school
* highSchool    dummy variable for senior high school
* college       dummy variable for college and higher
* unemp         dummy variable for unemployed
* lpa           dummy variable for lack of physical activity
* hhSize        household size
* urban         dummy for urban
* depriv        index of deprivation
* inequality    Gini coefficient
* logpce        log per capita consumption expenditure
* medpce6       median per capita consumption expenditure
* neko_kpi      quintile per capita consumption expenditure
* BPS07         2007 district ID (440)
*********************************************************************


*********************************************************************
* Multilevel Multinomial Logistic Regression - NULL MODEL
*********************************************************************

* gsem setup
clonevar bmi = BMIcat
clonevar district = BPS07

* define macro
global nullX age2534 age3544 age4554 age5564 age65up female

* NULL MODEL
* calculate RE correlation manually
gsem (1.bmi <- $nullX RE1[district])  ///
     (3.bmi <- $nullX RE3[district]) if BMIsamp==1, ///
     cov(RE1[district]*RE3[district]) mlogit ///
     intm(mcaghermite) tech(dfp 10 nr 5 dfp 5 nr 5) difficult ///
     latent(RE1 RE3) nocapslatent
estat eform 1.bmi 3.bmi, cformat(%9.2f)
```

```
*****************************************************************
* Multilevel Multinomial Logistic Regression - FULL MODEL 1
*****************************************************************

* define macro
global theX logpce inequality ///
            age2534 age3544 age4554 age5564 age65up ///
            female single divorced widowed ///
            compulEdu highSchool college unemp lpa ///
            hhSize urban depriv medpce6

* FULL MODEL with log(PCE)
gsem (1.bmi <- $theX RE1[district])  ///
     (3.bmi <- $theX RE3[district]) if BMIsamp==1, ///
     cov(RE1[district]*RE3[district]) mlogit ///
     intm(mcaghermite) tech(dfp 10 nr 5 dfp 5 nr 5) difficult ///
     latent(RE1 RE3) nocapslatent
estat eform 1.bmi 3.bmi, cformat(%9.2f)

* extract random effects
predict ranef1 ranef3, latent mode

* calculate correlation between district RE
mat li b
mat REcorrFull = b[1,48] / ( sqrt(b[1,46]) * sqrt(b[1,47]) )
mat li REcorrFull

*****************************************************************
* Multilevel Multinomial Logistic Regression - FULL MODEL 2
*****************************************************************

* define macro
global someX inequality ///
            age2534 age3544 age4554 age5564 age65up ///
            female single divorced widowed ///
            compulEdu highSchool college unemp lpa ///
            hhSize urban depriv medpce6

* FULL MODEL with INCOME QUINTILE
gsem (1.bmi <- i.neko_kpi $someX RE1[district])  ///
     (3.bmi <- i.neko_kpi $someX RE3[district]) if BMIsamp==1, ///
     cov(RE1[district]*RE3[district]) mlogit ///
     intm(mcaghermite) tech(dfp 12 nr 5 dfp 5 nr 5) difficult ///
     latent(RE1 RE3) nocapslatent
estat eform 1.bmi 3.bmi, cformat(%9.2f)

* extract random effects
predict ranef1qui ranef3qui, latent mode

*****************************************************************
* Multilevel Multinomial Logistic Regression - INTERACTION MDL
*****************************************************************

* create interaction terms
gen femXinc  = female*logpce
gen femXineq = female*inequality

* define macro
global intX femXinc femXineq logpce inequality ///
            age2534 age3544 age4554 age5564 age65up ///
            female single divorced widowed ///
            compulEdu highSchool college unemp lpa ///
            hhSize urban depriv medpce6

* INTERACTION MODEL
gsem (1.bmi <- $intX RE1[district])  ///
     (3.bmi <- $intX RE3[district]) if BMIsamp==1, ///
     cov(RE1[district]*RE3[district]) mlogit ///
     intm(mcaghermite) tech(dfp 12 nr 5 dfp 5 nr 5) difficult ///
     latent(RE1 RE3) nocapslatent
```

```
            estat eform 1.bmi 3.bmi, cformat(%9.2f)

            * extract random effects
            predict ranef1int ranef3int, latent mode

            * calculate correlation between district RE
            mat li b
            mat REcorrInt = b[1,52] / ( sqrt(b[1,51]) * sqrt(b[1,50]) )
            mat li REcorrInt


            * joint test
            test age2534 age3544 age4554 age5564 age65up
            test single divorced widowed
            test logpce
            test inequality
            test female
            test logpce femXinc
            test inequality femXinc
            test compulEdu highSchool college
            test depriv

            test [1.bmi]inequality = [3.bmi]inequality
            test [1.bmi]logpce = [3.bmi]logpce
            test [1.bmi]female = [3.bmi]female

            ****************************************************************
            * Multilevel Multinomial Logistic Regression - BY URBAN/RURAL
            ****************************************************************

            * define macro
            global STRAX inequality ///
                        age2534 age3544 age4554 age5564 age65up ///
                        female single divorced widowed ///
                        compulEdu highSchool college unemp lpa ///
                        hhSize depriv medpce6

            * mark sample
            gen BMIsampR = (BMIsamp==1 & urban==0)
            tab BMIsampR if BMIsamp==1
            gen BMIsampU = (BMIsamp==1 & urban==1)
            tab BMIsampU if BMIsamp==1

            * STRATIFIED MODEL: URBAN SAMPLE
            gsem (1.bmi <- i.neko_kpi $STRAX RE1[district])  ///
                 (3.bmi <- i.neko_kpi $STRAX RE3[district]) if BMIsampU==1, ///
                 cov(RE1[district]*RE3[district]) mlogit ///
                 intm(mcaghermite) tech(dfp 15 nr 5 dfp 10 nr 5) difficult ///
                 latent(RE1 RE3) nocapslatent
            estat eform 1.bmi 3.bmi, cformat(%9.2f)

            * STRATIFIED MODEL: RURAL SAMPLE
            gsem (1.bmi <- i.neko_kpi $STRAX RE1[district])  ///
                 (3.bmi <- i.neko_kpi $STRAX RE3[district]) if BMIsampR==1, ///
                 cov(RE1[district]*RE3[district]) mlogit ///
                 intm(mcaghermite) tech(dfp 15 nr 5 dfp 10 nr 5) difficult ///
                 latent(RE1 RE3) nocapslatent
            estat eform 1.bmi 3.bmi, cformat(%9.2f)

            ****************************************************************
            * Multilevel Multinomial Logistic Regression - BY GENDER
            ****************************************************************

            * keep complete observation
            drop if logpce == . | inequality == . | age2534 == . | ///
                    age3544 == . | age4554 == . | age5564 == . | ///
                    age65up == . | single == . | divorced == . | ///
                    widowed == . | compulEdu == . | highSchool == . | ///
                    college == . | unemp == . | lpa == . | ///
                    hhSize == . | urban == . | depriv == . | ///
                    medpce6 == . | female == . | neko_kpi == . | bmi == .
```

230

```
* create income quintile dummy
tab neko_kpi, gen(q)

* female model
runmlwin bmi q2 q3 q4 q5 inequality ///
          age2534 age3544 age4554 age5564 age65up ///
          single divorced widowed ///
          compulEdu highSchool college unemp lpa ///
          hhSize urban depriv medpce6 cons ///
          if female == 1, ///
        level2(BPS07: cons) ///
        level1(idart:) ///
        discrete(d(multinomial) l(mlogit) denom(cons) ///
              base(2) pql2) ///
        forcesort nopause corr rrr

* male model
runmlwin bmi q2 q3 q4 q5 inequality ///
          age2534 age3544 age4554 age5564 age65up ///
          single divorced widowed ///
          compulEdu highSchool college unemp lpa ///
          hhSize urban depriv medpce6 cons ///
          if female == 0, ///
        level2(BPS07: cons) ///
        level1(idart:) ///
        discrete(d(multinomial) l(mlogit) denom(cons) ///
              base(2) pql2) ///
        forcesort nopause corr rrr

*****************************************************************
* Quantile Regression
*****************************************************************

* define macro
global theX logpce inequality ///
          age2534 age3544 age4554 age5564 age65up ///
          female single divorced widowed ///
          compulEdu highSchool college unemp lpa ///
          hhSize urban depriv medpce6

* check percentile
_pctile BMI if BMIsamp==1, nq(100)
ret li

* ols
reg BMI $theX if BMIsamp==1, cluster(BPS07)

* quantile regression
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(2)
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(14)
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(67)
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(82)
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(92)
qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q(97)

forval i = 5(5)95 {
      qreg2 BMI $theX if BMIsamp==1, cluster(BPS07) q('i')
}

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

Appendix D:
Supplementary data for Chapter 5

Figure D.1: Title page of paper 4

International Journal of Health Geographics

RESEARCH

Open Access

CrossMark

Geography and social distribution of malaria in Indonesian Papua: a cross-sectional study

Wulung Hanandita[*] and Gindo Tampubolon

Abstract

Background: Despite being one of the world's most affected regions, only little is known about the social and spatial distributions of malaria in Indonesian Papua. Existing studies tend to be descriptive in nature; their inferences are prone to confounding and selection biases. At the same time, there remains limited malaria-cartographic activity in the region. Analysing a subset (N = 22,643) of the National Basic Health Research 2007 dataset (N = 987,205), this paper aims to quantify the district-specific risk of malaria in Papua and to understand how socio-demographic/economic factors measured at individual and district levels are associated with individual's probability of contracting the disease.

Methods: We adopt a Bayesian hierarchical logistic regression model that accommodates not only the nesting of individuals within the island's 27 administrative units but also the spatial autocorrelation among these locations. Both individual and contextual characteristics are included as predictors in the model; a normal conditional autoregressive prior and an exchangeable one are assigned to the random effects. Robustness is then assessed through sensitivity analyses using alternative hyperpriors.

Results: We find that rural Papuans as well as those who live in poor, densely forested, lowland districts are at a higher risk of infection than their counterparts. We also find age and gender differentials in malaria prevalence, if only to a small degree. Nine districts are estimated to have higher-than-expected malaria risks; the extent of spatial variation on the island remains notable even after accounting for socio-demographic/economic risk factors.

Conclusions: Although we show that malaria is geography-dependent in Indonesian Papua, it is also a disease of poverty. This means that malaria eradication requires not only biological (proximal) interventions but also social (distal) ones.

Keywords: Malaria, Map, Papua, Indonesia, Bayesian, Spatial, Multilevel

Background

Malaria, a mosquito-borne infectious disease that inflicts devastating health [1, 2] and economic [3–5] costs on society, remains a major problem in Indonesian Papua [6] (Fig. 1). This region of mixed-parasite endemicity is located in the easternmost part of the Indonesian archipelago (Fig. 2) and is classified by the World Health Organization (WHO) as hyper-endemic area with annual parasite incidence (API) greater than 10 % (nationwide API < 1 %; [7]) and parasite prevalence (PP) as high as 50–75 % (nationwide PP < 1 %; [8]). Malaria accounts for a considerable proportion (15–34 %) of total hospital workload in the region [9]; mortality due to severe anaemia [10] as well as multi-drug resistance with high rate of therapeutic failure (65–95 %) have been documented [11, 12]. In 2007, the Ministry of Health of the Republic of Indonesia [13] estimated that the infectious disease was prevalent among one-fifth (22.25 %) of the Papuan population—a figure that is seven times higher than the national average (Fig. 1). Perhaps nothing can highlight

*Correspondence: wulung.hanandita@postgrad.manchester.ac.uk
Cathie Marsh Institute for Social Research (CMIST), University Manchester, Oxford Road, Manchester M13 9PL, UK

BioMed Central

```
##########################################################################
#                    R CODE TO REPLICATE THE RESULTS OF
#   Geography and social distribution of malaria in Indonesian Papua
##########################################################################

# DATA SOURCE

# The National Basic Health Research (RISKESDAS) 2007:
#        http://www.litbang.depkes.go.id/

# The National Socio-economic Survey (SUSENAS) 2008:
#        http://microdata.bps.go.id/

# The Village Census (PODES) 2008:
#        http://microdata.bps.go.id/

# Spatial polygon data of administrative boundaries:
#        http://www.gadm.org/

#-----------------------------------------------------------------------#
# VARIABLE DESCRIPTION
#-----------------------------------------------------------------------#
# female                dummy variable for female
# agegr                 age-group indicator
# itn                   dummy variable for ITN use
# rural                 dummy variable for rural dweller
# lowland               dummy variable for lowland district
# pdpopjungle           proportion living in or near forest
# medpce5               quintile of district income
# GID                   district identifier
# papuaedit.graph       a 27 x 27 binary adjacency matrix
#-----------------------------------------------------------------------#

# library
require(INLA)

# listwise deletion
ori.sample <- dim(rkd.papua)[1]
rkd.papua <- na.omit(rkd.papua)
complete.sample <- dim(rkd.papua)[1]
ori.sample
complete.sample
complete.sample / ori.sample

#-----------------------------------------------------------------------#
# MODEL FITTING
#-----------------------------------------------------------------------#

# null model
f0 <- malaria ~ 1 +
                f(GID, model = "bym", graph = "papuaedit.graph",
                  scale.model = TRUE,
                  param = c(0.001,0.001,0.001,0.001))

system.time({
  (fit.0 <- inla(f0, family = "binomial", data = rkd.papua,
                 control.compute = list(dic = TRUE, cpo = TRUE)))
})

# full model
f1 <- malaria ~ 1 + factor(female) + factor(agegr) +
                factor(itn) + factor(rural) +
                factor(lowland) + I(pdpopjungle*10) +
                factor(medpce5) +
                f(GID, model = "bym",graph = "papuaedit.graph",
                  param = c(0.001, 0.001,0.001, 0.001),
                  scale.model = TRUE)

system.time({
  (fit.1 <- inla(f1, family = "binomial", data = rkd.papua,
```

234

```
                            control.compute = list(dic = TRUE, cpo = TRUE)))
        })

        #-------------------------------------------------------------------------#
        # SENSITIVITY TO HYPERPRIORS
        #-------------------------------------------------------------------------#

        f1.a <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym",graph = "papuaedit.graph",
                              param = c(1, 0.00005,1, 0.00005),
                              scale.model = TRUE)

        f1.b <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(1, 0.001, 1, 0.001),
                              scale.model = TRUE)

        f1.c <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(1, 0.01, 1, 0.01),
                              scale.model = TRUE)

        f1.d <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(1, 0.1, 1, 0.1),
                              scale.model = TRUE)


        f1.e <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(0.5, 0.0005, 0.5, 0.0005),
                              scale.model = TRUE)

        f1.f <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(0.01, 0.01, 0.01, 0.01),
                              scale.model = TRUE)

        f1.g <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(10, 0.35, 10, 0.35),
                              scale.model = TRUE)

        f1.h <- malaria ~ 1 + factor(female) + factor(agegr) +
                            factor(itn) + factor(rural) +
                            factor(lowland) + I(pdpopjungle*10) +
                            factor(medpce5) +
                            f(GID, model = "bym", graph = "papuaedit.graph",
                              param = c(0.001, 0.001, 0.001, 0.001),
```

```
                      scale.model = TRUE)

f1.i <- malaria ~ 1 + factor(female) + factor(agegr) +
                     factor(itn) + factor(rural) +
                     factor(lowland) + I(pdpopjungle*10) +
                     factor(medpce5) +
                     f(GID, model = "bym", graph = "papuaedit.graph",
                       param = c(0.00001, 0.00001, 0.00001, 0.00001),
                       scale.model = TRUE)

f1.j <- malaria ~ 1 + factor(female) + factor(agegr) +
                     factor(itn) + factor(rural) +
                     factor(lowland) + I(pdpopjungle*10) +
                     factor(medpce5) +
                     f(GID, model = "bym", graph = "papuaedit.graph",
                       param = c(0.0005, 0.0005, 0.0005, 0.0005),
                       scale.model = TRUE)

f1.k <- malaria ~ 1 + factor(female) + factor(agegr) +
                     factor(itn) + factor(rural) +
                     factor(lowland) + I(pdpopjungle*10) +
                     factor(medpce5) +
                     f(GID, model = "bym", graph = "papuaedit.graph",
                       param = c(1, 0.005, 1, 0.005),
                       scale.model = TRUE)

system.time({
  (fit.1.a <- inla(f1.a, family = "binomial", data = rkd.papua,
                   control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.a)

system.time({
  (fit.1.b <- inla(f1.b, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.b)

system.time({
  (fit.1.c <- inla(f1.c, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.c)

system.time({
  (fit.1.d <- inla(f1.d, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.d)

system.time({
  (fit.1.e <- inla(f1.e, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.e)

system.time({
  (fit.1.f <- inla(f1.f, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.f)

system.time({
  (fit.1.g <- inla(f1.g, family = "binomial", data = rkd.papua,
```

```
                          control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.g)

system.time({
  (fit.1.h <- inla(f1.h, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.h)

system.time({
  (fit.1.i <- inla(f1.i, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.i)

system.time({
  (fit.1.j <- inla(f1.j, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.j)

system.time({
  (fit.1.k <- inla(f1.k, family = "binomial", data = rkd.papua,
                    control.compute = list(dic = TRUE, cpo = TRUE)))
})

summary(fit.1.k)

####################################################################
#            WinBUGS CODE TO REPLICATE THE RESULTS OF
#   Geography and social distribution of malaria in Indonesian Papua
####################################################################

# Data source and variable description are the same with above R code
# However, instead of INLA, here the model is fitted using MCMC
# Please note that convergence may take considerably long time

model {

  for (i in 1:21772) {

        # data model
          malaria[i]  ~  dbern(p[i])

        # process model
          logit(p[i]) <- alpha                       # global intercept
                          + beta[1]  * female[i]
                          + beta[2]  * age2[i]
                          + beta[3]  * age3[i]
                          + beta[4]  * age4[i]
                          + beta[5]  * age5[i]
                          + beta[6]  * age6[i]
                          + beta[7]  * age7[i]
                          + beta[8]  * itn[i]
                          + beta[9]  * rural[i]
                          + beta[10] * lowland[i]
                          + beta[11] * jungle[i]
                          + beta[12] * income2[i]
                          + beta[13] * income3[i]
                          + beta[14] * income4[i]
                          + beta[15] * income5[i]
                          + S[GID[i]]               # spatial effects
                          + U[GID[i]]               # iid effects

  }
```

```
# Priors

  # vague prior for the overall intercept
    alpha  ~ dflat()

  # diffuse priors for regression coefficients
    for (k in 1:15) {beta[k] ~ dnorm(0,0.0001)}

  # CAR prior for spatial random effects
  # note that adj, weights, and num must be specified in data block
    S[1:27] ~ car.normal(adj[],weights[],num[],tau.S)

  # Gaussian prior for spatially unstructured random effects
    for (j in 1:27) {U[j] ~ dnorm(0,tau.U)}

# Hyperpriors

  # prior for the precision of spatially structured random effects
    tau.S ~ dgamma(0.001,0.001)

  # prior for the precision of spatially unstructured random effects
    tau.U ~ dgamma(0.001,0.001)

# Quantities of interest

  for (j in 1:27) {

  # sum of random effects
    xi[j] <- S[j] + U[j]

  # spatial residual odds ratio
    district.OR[j] <- exp(xi[j])

  # posterior probability that the residual odds ratio is above 1
    excess.risk[j] <- step(district.OR[j]-1)

  # baseline probability of malaria infection
    district.PR[j] <- exp(alpha + S[j] + U[j]) /
                      (1 + (exp(alpha + S[j] + U[j])))

  }

  # odds ratio of regression coefficients
    alpha.OR <- exp(alpha)
    for (k in 1:15) {beta.OR[k] <- exp(beta[k])}

  # conditional variance of spatially structured random effects
    sigma2.S <- 1/tau.S

  # unconditional variance of spatially structured random effects
    sigma2.S.un <- sd(S[])*sd(S[])

  # variance of spatially unstructured random effects
    sigma2.U <- 1/tau.U

  # pseudo variance partition coefficient
    vpc <- sigma2.S.un/(sigma2.S.un + sigma2.U)

}

###################################################################
```

Appendix E:
Supplementary data for Chapter 6

Figure E.1: Title page of paper 5

CrossMark

Multidimensional Poverty in Indonesia: Trend Over the Last Decade (2003–2013)

Wulung Hanandita[1] · Gindo Tampubolon[1]

Abstract The notion of poverty as an experience of multiple deprivation has been widely acknowledged. In Indonesia, however, poverty assessment has almost exclusively been conducted within the monetary space; even when multidimensionality is admitted, it has always been computed using variants of marginal method that are indifferent to joint deprivation. Applying a novel measurement method that is sensitive to both the incidence and the intensity of multiple deprivation to data from the National Socio-economic Survey (Susenas), this paper investigates the extent and the patterns of multidimensional poverty in Indonesia from 2003 to 2013 ($N = 7,148,964$). An Indonesian version of the multidimensional poverty index is constructed by augmenting the existing consumption poverty measure with information on health and education. Results suggest that there was an unambiguous poverty reduction over the last decade at both national and sub-national levels. The data also reveal that progress has been inclusive across population subgroups, although spatial variation remains notable. The new poverty measurement method proves to be easily adaptable to the Indonesian context and could complement the methods currently employed by the Indonesian Statistical Bureau.

✉ Wulung Hanandita
wulung.hanandita@postgrad.manchester.ac.uk

Gindo Tampubolon
gindo.tampubolon@manchester.ac.uk

[1] Cathie Marsh Institute for Social Research, University of Manchester, Humanities Bridgeford Street Building 2F, Oxford Road, Manchester M13 9PL, UK

🙼 Springer

```
*******************************************************************
*               STATA CODE TO REPLICATE THE RESULTS OF
*               Multidimensional poverty in Indonesia:
*               Trend over the last decade (2003-2013)
*******************************************************************

* DATA SOURCE

* The National Socio-economic Survey (SUSENAS) 2003-2013:
*         http://microdata.bps.go.id/

* Poverty lines at national and provincial level, 2008 onwards:
*         http://www.bps.go.id/

* Spatial polygon data of administrative boundaries:
*         http://www.gadm.org/

* Code for computing the inequality index can be obtained from
* Oxford Poverty & Human Development Initiative (OPHI):
*         http://www.ophi.org.uk/

*******************************************************************
* VARIABLE DESCRIPTION
*******************************************************************
* poorADB      per capita consumption expenditure <$1.51 PPP
* no_illdays   illness episode < 4 days
* no_morbidity morbidity < 3 diseases
* educated     completed primary school
* literate     can read and write latin characters
* YEAR         year indicator
* rural        dummy for rural area
* island       island indicator
* female       dummy for female
* CBID         census block ID
* HHID         household ID
* BPS03x       2003 district ID (346)
* weight       sampling weight
*******************************************************************

* deprivation matrix (g0)
gen inc = poorADB
gen he1 = 1 - no_illdays
gen he2 = 1 - no_morbidity
gen ed1 = 1 - educated
gen ed2 = 1 - literate

* national dashboard
tabstat inc he1 he2 ed1 ed2 [aw = weight], by(YEAR) nototal save

tabstatmat DAS_national,  nototal
mat2txt, mat(DAS_national) saving(DAS_national) replace

* urban dashboard
forval i = 0/1 {
        tabstat inc he1 he2 ed1 ed2 if rural == 'i' ///
        [aw = weight], by(YEAR) nototal save

        tabstatmat DAS_rural_'i',  nototal
        mat2txt, mat(DAS_rural_'i') saving(DAS_rural_'i') replace
}

* island dasboard
forval i = 1/6 {
        tabstat inc he1 he2 ed1 ed2 if island == 'i' ///
        [aw = weight], by(YEAR) nototal save

        tabstatmat DAS_island_'i',  nototal
        mat2txt, mat(DAS_island_'i') saving(DAS_island_'i') replace
}
```

```
* gender dashboard
forval i = 0/1 {
        tabstat inc he1 he2 ed1 ed2 if female == `i' ///
        [aw = weight], by(YEAR) nototal save

        tabstatmat DAS_female_`i',  nototal
        mat2txt, mat(DAS_female_`i') saving(DAS_female_`i') replace
}

* weighting scheme 1 (w1)
gen w1_inc = 1/3
gen w1_he1 = 1/6
gen w1_he2 = 1/6
gen w1_ed1 = 1/6
gen w1_ed2 = 1/6

* weighting scheme 2 (w2)
gen w2_inc = 1/3
gen w2_he1 = 1/6
gen w2_he2 = 1/6
gen w2_ed1 = 1/3
gen w2_ed2 = 0

* weighting scheme 1 (w3)
gen w3_inc = 1/3
gen w3_he1 = 1/3
gen w3_he2 = 0
gen w3_ed1 = 1/3
gen w3_ed2 = 0

* weighted deprivation matrix (g0)
global indicators inc he1 he2 ed1 ed2

foreach var of global indicators {
        gen d_w1_`var' = `var' * w1_`var'
        gen d_w2_`var' = `var' * w2_`var'
        gen d_w3_`var' = `var' * w3_`var'
}

drop w2_* w3_* // free RAM

* counting vector (ci)
foreach var in w1 w2 w3 {
        egen ci_`var' = rsum(d_`var'_inc ///
                             d_`var'_he1 d_`var'_he2 ///
                             d_`var'_ed1 d_`var'_ed2)
}

bys YEAR: tab ci_w1 [aw = weight], mis

* identification & aggregation using the second cutoff
foreach w in w1 w2 w3 {
        foreach k in 16 33 50 66 83 100 {

        gen H_`w'_`k'p = (ci_`w' >= `k' / 100)

        gen A_`w'_`k'p = ci_`w' if H_`w'_`k'p == 1

        gen M0_`w'_`k'p = 0
        replace M0_`w'_`k'p = ci_`w' if H_`w'_`k'p == 1
        }
}

drop ci_w2 ci_w3 // free RAM

* confidence interval

  // declare survey design
  svyset CBID [pw = weight], strata(rural) || HHID
```

```
// CI national
foreach k in 16 33 50 66 83 100 {
 gen N_L_M0_w1_`k'p = .
 gen N_U_M0_w1_`k'p = .
}

forval i = 2003/2013 {
 foreach k in 16 33 50 66 83 100 {
  svy: mean M0_w1_`k'p if YEAR == `i'
  replace N_L_M0_w1_`k'p = _b[M0_w1_`k'p] - ///
                     1.96*((_se[M0_w1_`k'p])) if YEAR == `i'
  replace N_U_M0_w1_`k'p = _b[M0_w1_`k'p] + ///
                     1.96*((_se[M0_w1_`k'p])) if YEAR == `i'
 }
}

tabstat M0_w1_33p N_L_M0_w1_33p N_U_M0_w1_33p [aw=weight], ///
        by(YEAR) nototal

// CI urban/rural
foreach k in 16 33 50 66 83 100 {
 gen UR_L_M0_w1_`k'p = .
 gen UR_U_M0_w1_`k'p = .
}

forval i = 2003/2013 {
 foreach k in 16 33 50 66 83 100 {
  svy: mean M0_w1_`k'p if YEAR == `i', over(rural)
  replace UR_L_M0_w1_`k'p = _b[0] - 1.96*((_se[0])) ///
                            if YEAR == `i' & rural == 0
  replace UR_U_M0_w1_`k'p = _b[0] + 1.96*((_se[0])) ///
                            if YEAR == `i' & rural == 0
  replace UR_L_M0_w1_`k'p = _b[1] - 1.96*((_se[1])) ///
                            if YEAR == `i' & rural == 1
  replace UR_U_M0_w1_`k'p = _b[1] + 1.96*((_se[1])) ///
                            if YEAR == `i' & rural == 1
 }
}

tabstat M0_w1_33p UR_L_M0_w1_33p UR_U_M0_w1_33p [aw=weight], ///
        by(YEAR) nototal

// CI island
foreach k in 16 33 50 66 83 100 {
 gen IS_L_M0_w1_`k'p = .
 gen IS_U_M0_w1_`k'p = .
}

forval i = 2003/2013 {
 foreach k in 16 33 50 66 83 100 {
  svy: mean M0_w1_`k'p if YEAR == `i', over(island)
  replace IS_L_M0_w1_`k'p = _b[1] - 1.96*((_se[1])) ///
                            if YEAR == `i' & island == 1
  replace IS_U_M0_w1_`k'p = _b[1] + 1.96*((_se[1])) ///
                            if YEAR == `i' & island == 1
  replace IS_L_M0_w1_`k'p = _b[2] - 1.96*((_se[2])) ///
                            if YEAR == `i' & island == 2
  replace IS_U_M0_w1_`k'p = _b[2] + 1.96*((_se[2])) ///
                            if YEAR == `i' & island == 2
  replace IS_L_M0_w1_`k'p = _b[3] - 1.96*((_se[3])) ///
                            if YEAR == `i' & island == 3
  replace IS_U_M0_w1_`k'p = _b[3] + 1.96*((_se[3])) ///
                            if YEAR == `i' & island == 3
  replace IS_L_M0_w1_`k'p = _b[4] - 1.96*((_se[4])) ///
                            if YEAR == `i' & island == 4
  replace IS_U_M0_w1_`k'p = _b[4] + 1.96*((_se[4])) ///
                            if YEAR == `i' & island == 4
  replace IS_L_M0_w1_`k'p = _b[5] - 1.96*((_se[5])) ///
                            if YEAR == `i' & island == 5
  replace IS_U_M0_w1_`k'p = _b[5] + 1.96*((_se[5])) ///
```

```
                                     if YEAR == 'i' & island == 5
     replace IS_L_M0_w1_'k'p = _b[6] - 1.96*((_se[6])) ///
                                     if YEAR == 'i' & island == 6
     replace IS_U_M0_w1_'k'p = _b[6] + 1.96*((_se[6])) ///
                                     if YEAR == 'i' & island == 6
 }
}

 forval i = 1/6 {
     tabstat M0_w1_33p IS_L_M0_w1_33p IS_U_M0_w1_33p ///
             [aw=weight] if island == 'i', by(YEAR) nototal
 }

 // CI male/female
 foreach k in 16 33 50 66 83 100 {
  gen MF_L_M0_w1_'k'p = .
  gen MF_U_M0_w1_'k'p = .
 }

 forval i = 2003/2013 {
  foreach k in 16 33 50 66 83 100 {
   svy: mean M0_w1_'k'p if YEAR == 'i', over(female)
   replace MF_L_M0_w1_'k'p = _b[0] - 1.96*((_se[0])) ///
                                   if YEAR == 'i' & female == 0
   replace MF_U_M0_w1_'k'p = _b[0] + 1.96*((_se[0])) ///
                                   if YEAR == 'i' & female == 0
   replace MF_L_M0_w1_'k'p = _b[1] - 1.96*((_se[1])) ///
                                   if YEAR == 'i' & female == 1
   replace MF_U_M0_w1_'k'p = _b[1] + 1.96*((_se[1])) ///
                                   if YEAR == 'i' & female == 1
  }
 }

 tabstat M0_w1_33p MF_L_M0_w1_33p MF_U_M0_w1_33p [aw=weight], ///
         by(YEAR) nototal

* dimensional breakdown

 // pick k
 local k = 33

 // censored headcount:
 // the proportion of the population that are poor with respect
 // to a certain cutoff and are deprived in that dimension at
 // the same time.
 foreach var of global indicators {
         gen ch_w1_'var'_'k'p = (ci_w1 >= 'k'/100 & 'var' == 1)
 }

 tabstat ch_* [aw=weight], by(YEAR) nototal save

 tabstatmat CH_national, nototal
 mat2txt, mat(CH_national) saving(CH_national) replace

 // calculate absolute and relative contribution
 // note: do it for each year
 local k = 33
 gen M0_w1_'k'p_y = .
 forval i = 2003/2013 {
         summ M0_w1_'k'p [aw = weight] if YEAR == 'i'
         replace M0_w1_'k'p_y = r(mean) if YEAR == 'i'
 }

 local k = 33
 foreach var of global indicators {
         gen con_a_w1_'k'p_'var' = .
         gen con_r_w1_'k'p_'var' = .
 }

 local k = 33
```

```
        foreach var of global indicators {
                replace con_a_w1_`k'p_`var' = ///
                        ch_w1_`var'_`k'p * w1_`var'
                replace con_r_w1_`k'p_`var' = ///
                        ch_w1_`var'_`k'p * w1_`var' / M0_w1_`k'p_y
    }

    tabstat con_* [aw=weight], by(YEAR) nototal save
    tabstatmat CON_national,  nototal
    mat2txt, mat(CON_national) saving(CON_national) replace

    drop ch_w1_*  // free RAM

* national MPI time-series
tabstat H_w1_33p A_w1_33p ///
        M0_w1_16p M0_w1_33p M0_w1_50p ///
        M0_w1_66p M0_w1_83p M0_w1_100p ///
        M0_w2_33p M0_w3_33p ///
        con_a_* con_r_* N_L_* N_U_* [aw = weight], ///
        by(YEAR) nototal save

tabstatmat MPI_national,  nototal
mat2txt, mat(MPI_national) saving(MPI_national) replace

* urban/rural time-series
forval i = 0/1 {
        tabstat H_w1_33p A_w1_33p ///
        M0_w1_16p M0_w1_33p M0_w1_50p ///
        M0_w1_66p M0_w1_83p M0_w1_100p ///
        M0_w2_33p M0_w3_33p UR_L_* UR_U_* ///
        [aw = weight] if rural == `i', ///
        by(YEAR) nototal save

        tabstatmat MPI_rural_`i',  nototal
        mat2txt, mat(MPI_rural_`i') saving(MPI_rural_`i') replace
}

* island MPI time-series
forval i = 1/6 {
        tabstat H_w1_33p A_w1_33p ///
        M0_w1_16p M0_w1_33p M0_w1_50p ///
        M0_w1_66p M0_w1_83p M0_w1_100p ///
        M0_w2_33p M0_w3_33p IS_L_* IS_U_* ///
        [aw = weight] if island == `i', ///
        by(YEAR) nototal save

        tabstatmat MPI_island_`i',  nototal
        mat2txt, mat(MPI_island_`i') saving(MPI_island_`i') replace
}

* gender MPI time-series
forval i = 0/1 {
        tabstat H_w1_33p A_w1_33p ///
        M0_w1_16p M0_w1_33p M0_w1_50p ///
        M0_w1_66p M0_w1_83p M0_w1_100p ///
        M0_w2_33p M0_w3_33p MF_L_* MF_U_* ///
        [aw = weight] if female == `i', ///
        by(YEAR) nototal save

        tabstatmat MPI_female_`i',  nototal
        mat2txt, mat(MPI_female_`i') saving(MPI_female_`i') replace
}

* H, A, M0 by Kabupaten: 2013
tabstat H_w1_33p A_w1_33p M0_w1_33p ///
        if YEAR == 2013 [aw = weight], ///
        by(BPS03x) nototal save

tabstatmat MPI_district_2013,  nototal
mat2txt, mat(MPI_district_2013) saving(MPI_district_2013) replace
```

```
* subgroup decomposition
* here we take the weighted proportion

  // urban/rural
  forval i = 2003/2013 {
        tab rural [aw = weight] if YEAR == 'i', matcell(f'i')
        mat2txt, mat(f'i') saving(f_rural_'i'.txt) replace
  }

  // island
  forval i = 2003/2013 {
        tab island [aw = weight] if YEAR == 'i', matcell(f'i')
        mat2txt, mat(f'i') saving(f_island_'i'.txt) replace
  }

  // gender
  forval i = 2003/2013 {
        tab female [aw = weight] if YEAR == 'i', matcell(f'i')
        mat2txt, mat(f'i') saving(f_female_'i'.txt) replace
  }

* multidimensionally poor but income non-poor by weights
foreach w in w1 w2 w3 {
   clonevar H_33p_'w' = H_'w'_33p
   replace H_33p_'w' = . if H_'w'_33p == .
}

foreach w in w1 w2 w3 {
   forval i = 2003/2013 {
      tab inc H_33p_'w' [aw = weight] if YEAR == 'i', ///
         matcell(fw'i')
      mat2txt, mat(fw'i') saving(f_'w'_'i'.txt) replace
   }
}


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