UNIVERSITY of York

RESEARCH

CHE

Centre For Health Economics

FUNDED BY NIHR | National Institute for Health Research

UKaid from the British people

Who Benefits from Health insurance? Uncovering Heterogeneous Policy Impacts Using Causal Machine Learning

Noemi Kreif, Andrew Mirelman, Rodrigo Moreno-Serra, Taufik Hidayat, Karla DiazOrdaz, Marc Suhrcke

CHE Research Paper 173

Who benefits from health insurance? Uncovering heterogeneous policy impacts using causal machine learning

[a]Noemi Kreif
[b]Andrew Mirelman
[a]Rodrigo Moreno-Serra
[c]Taufik Hidayat
[d]Karla DiazOrdaz
[ae]Marc Suhrcke

[a]Centre for Health Economics, University of York, UK
[b]World Health Organization, Geneva, Switzerland
[c]Center for Health Economics and Policy Studies (CHEPS), Faculty of Public Health, Universitas Indonesia, Depok, Indonesia
[d]Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
[e]Luxembourg Institute of Socio-economic Research, Luxembourg

October 2020

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

We gratefully acknowledge the extremely valuable input from Ryota Nakamura, Mardiati Nadjib, Budi Hidayat, Darius Erlangga, and Peter Smith.

No ethical approval was needed.

Funding
This study was partly funded by the Bill and Melinda Gates Foundation, through the International Decision Support Initiative (iDSI), and by the National Institute for Health Research (NIHR) (NIHR Global Health Econometrics & Economics Group 16/137/90) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the funders. KDO was supported by a Royal Society-Wellcome Trust Sir Henry Dale Fellowship (218554/Z/19/Z).

Further copies
Only the latest electronic copy of our reports should be cited. Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK

www.york.ac.uk/che

Abstract

To be able to target health policies more efficiently, policymakers require knowledge about which individuals benefit most from a particular programme. While traditional approaches for subgroup analyses are constrained only to consider a small number of arbitrarily set, pre-defined subgroups, recently proposed causal machine learning (CML) approaches help explore treatment-effect heterogeneity in a more flexible yet principled way. This paper illustrates one such approach – 'causal forests' – in evaluating the effect of mothers' health insurance enrolment in Indonesia. Contrasting two health insurance schemes (subsidised and contributory) to no insurance, we find beneficial average impacts of enrolment in contributory health insurance on maternal health care utilisation and infant mortality. For subsidised health insurance, however, both effects were smaller and not statistically significant. The causal forest algorithm identified significant heterogeneity in the impacts of the contributory insurance scheme: disadvantaged mothers (i.e. with lower wealth quintiles, lower educated, or in rural areas) benefit the most in terms of increased health care utilisation. No significant heterogeneity was found for the subsidised scheme, even though this programme targeted vulnerable populations. Our study demonstrates the power of CML approaches to uncover the heterogeneity in programme impacts, hence providing policymakers with valuable information for programme design.

Keywords: policy evaluation, machine learning, heterogeneous treatment effects, health insurance

1. Introduction

Policymakers around the world are implementing health system policies to promote access to essential health care and to meet the health-related Sustainable Development Goals (Sachs 2012). To assess the impact of these policies on population health and health inequalities, they require evidence that is explicit about the complexity of policy impacts, in particular with respect to the potential heterogeneity in their effects. Here, we focus on heterogeneity in terms of observed *effect modifiers*, i.e. covariates that can modify the causal effect of a policy. Some such evidence can be obtained from subgroup analyses, comparing the effects of interventions across different population groups, characterised, for instance, by their socio-economic status (Mackenbach 2003). However, impact evaluations tend not to present such disaggregate analysis, due to the concern that subgroup analysis, unless pre-specified in an analysis plan using pre-existing theory, may be viewed as post-hoc 'data dredging', leading to spurious findings and publication bias towards significant estimates (Petticrew et al. 2012). Another common approach to assess treatment effect heterogeneity according to pre-specified covariates is by adding interaction terms in outcome regression models, but this implies making parametric assumptions (typically linearity), which may not be plausible (Hainmueller and Mummolo 2019).

Recently developed machine learning (ML) approaches have been proposed as a way to pre-empt the criticism of arbitrariness and to estimate treatment effect heterogeneity based on an unbiased exploration of the data (Athey and Imbens, 2017). The 'causal forests' approach is one specific, increasingly popular ML approach, developed by Athey, Wager and colleagues (Nie and Wager 2017, Wager and Athey 2018, Athey et al. 2019a). The causal forest approach combines the flexibility of ML methods with the rigour of semi-parametric statistical theory. The method flexibly captures how treatment effects vary according to observed covariates, by nonparametrically estimating a so-called 'conditional treatment effect function' (CATE). The algorithm reports individual treatment effects given observed covariates, and these individual treatment effects can be aggregated to provide average estimates for subgroups of interest. The first major benefit of this approach compared to traditional methods is in that researchers do not need to either split the data into many subgroups for stratified analysis or make parametric assumptions about interactions in an outcome model. The second advantage, which is unique to the ML component of the approach, is that it can describe CATEs for ex-ante unknown subgroups, whose importance is discovered by the algorithm, by providing a measure of the importance of a wide range of variables in predicting treatment effects.

In this paper, we demonstrate how such an ML approach can provide beneficial information for health policymaking decisions aimed at improving overall health and reducing health inequalities. We apply the causal forests method to explore effect heterogeneity in two types of public health insurance programmes in Indonesia: subsidised health insurance targeting the poor and the near poor, and contributory health insurance for employees of the formal sector. We use the sequential implementation of health insurance that preceded the establishment of the unified National Health Insurance programme (Jamima Kesehatan National (JKN)) in 2014 as 'natural experiments' to investigate how changes in health insurance status have influenced health outcomes and health care utilisation. We first focus on infant mortality as the health outcome of interest, as this is arguably more sensitive to changes in access to health care services, compared to other health outcomes (Currie and Gruber 1996, Dow and Schmeer 2003). If health insurance is expected to reduce infant mortality, one of the channels through which this could occur is a differential increase in health care utilisation by insured mothers compared to uninsured ones. Birth assisted by a skilled health professional has been found to be a predictor of infant mortality in the neonatal stage (Lawn et al. 2005). We first estimate average treatment effects for these outcomes, using regression and propensity score approaches.

Given the notable geographical, ethnic, and economic disparities within Indonesia, it is expected that average policy effects mask important heterogeneity in the effect of health insurance programmes. For optimal targeting of interventions, health policymakers need to know how the impact of health insurance varies across different subgroups, in particular for those groups most vulnerable in terms of disease burden and access to health care (Lagomarsino et al. 2012): mothers with low education, those in the bottom socioeconomic quintiles, and those living in remote, rural communities. We therefore also estimate individual treatment effects using the causal forests approach and aggregate these to estimate subgroup-average treatment effects. We do so both for ex-ante specified subgroups following traditional practice, and via a data-driven ML approach that characterises those variables most associated with heterogeneity. We use data from the Indonesian Family Life Survey (IFLS) (Strauss et al. 2004, Strauss et al. 2009, Strauss et al. 2016), a rich and high-quality longitudinal survey of Indonesian individuals and households that allows controlling for observed confounders of the causal relationship between health insurance and health care utilisation (and health outcomes).

This paper makes three main contributions. First, through an evaluation of the impact of health insurance on health care utilisation, we illustrate the value of using novel causal ML methods for health policy evaluation. In particular, this is the first study that uses the causal forest approach in the context of a health policy evaluation to characterise the drivers of treatment heterogeneity for health insurance programmes. Second, we contribute to empirical evidence of health insurance expansions in Indonesia, by characterising its heterogeneous impacts on health care utilisation, by socioeconomic subgroups. Third, we add to the scarce rigorous evidence available about the effects of health insurance on health outcomes and health care utilisation in low- and middle-income countries (LMICs).

In the following sections, we present the institutional setting in Indonesia (Section 2.1), briefly review the literature on the impact evaluations of health insurance, both worldwide and in Indonesia (Section 2.2), and present the data used in the study (2.3). Then we describe the methods (Section 3), with a focus on the theory and practical implementation of the causal forest approach, present the results (Section 4) and discuss the findings and future avenues of research (Section 5).

2. The evaluation of the National Health Insurance expansion in Indonesia

2.1 Institutional setting

With an estimated population of over 270 million in 2019[1], Indonesia is the fourth most populous country in the world. Total health spending was 3.1% of gross domestic product (GDP) in 2016 (WHO 2019), with a relatively small share of total health expenditures being publicly funded (39%) (Mahendradhata et al. 2017). While on average, health indicators have improved significantly over the last decades – life expectancy rising from 63 to 71, and infant mortality falling from 41 to 26 deaths per 1000 live births, between 1990 and 2012 (Mahendradhata, Trisnantoro et al. 2017) – there remain considerable health inequalities (Agustina et al. 2019). To address unmet health care needs, and high out-of-pocket and catastrophic health spending, Indonesia launched an ambitious health system reform in 2014, the JKN, comprising a wide range of policies, including a unified benefit package, premium subsidies for the poor, a national formulary, and a provider contracting and payment system. The JKN reform was preceded by a series of subsidised health insurance expansions programmes – the focus of our study – starting from the 1990s. We briefly review the landscape of health insurance in Indonesia over our study period (2000-2014) (see Figure 1), describing the main contributory and subsidised health insurance schemes.

Historically, health insurance in Indonesia was available as contributory schemes, for those employed in the formal sector and their family dependants (Achadi et al. 2014). *Askes* was a mandatory health insurance programme for active and retired civil servants, and military personal, with a contribution of 2% from payroll salary or pension (Thabrany 2001), while *Jamsostek* was an optional social security scheme for private employees**,** with a 3-6% salary contribution (Hidayat 2004)[2]. For poor households that were not eligible for these health insurance programmes, from 1994 a *Health Card* programme provided free basic health care at public health facilities (Johar 2009).

The *Askeskin* scheme, established in 2005, was the first national, subsidised health insurance programme, basing eligibility on a combination of geographic and individual-level criteria (Sparrow et al. 2013). The insurance scheme covered a comprehensive package of health services (outpatient care and inpatient care, mobile health services, immunisation and medications), with the premium fully subsidised by central government (Sparrow et al. 2013). The scheme left a large group of households without health coverage, i.e. those not poor enough to be eligible but also not having access to contributory health insurance in the formal sector. In 2008, the *Askeskin* programme was re-organised, and the resulting *Jamkesmas* expanded the eligible population, targeting the poor and 'near poor', based on a combination of means testing (using 14 assets recorded in a National Poverty Census Survey indicators) and local government eligibility criteria (Harimurti, et al. 2013). However, not all households eligible for the programme possessed a membership card due to perceived stigmatisation from health care providers and concerns about long waiting times (Harimurti, et al. 2013). Despite the means testing, a significant 'leakage' occurred, resulting in households in higher income quantiles also receiving free health insurance (Harimurti, et al. 2013). In principle, *Jamkesmas* continued to provide a comprehensive package, but in reality, the availability of services in the benefit package was limited, especially in rural areas, thereby contributing to large geographic inequalities in access (Harimurti, et al. 2013). To compensate for the large gaps in insurance status, district governments provided decentralised health care financing schemes offering subsidised

[1] See http://worldpopulationreview.com/countries/indonesiahttp://worldpopulationreview.com/countries/indonesia, Retrieved 2019-08-07.

[2] Unlike *Askes*, it was possible to opt out of this scheme, and purchase private health insurance instead, leading to a relatively low overall coverage: only 1.5% of the population was covered by *Jamsostek* in 2001 (Thabrany 2001).

health insurance, known as *Jamkesda*) (Sparrow et al. 2017).[3] In 2014, all health insurance schemes were absorbed into a single national health insurance scheme, JKN, aimed at continuing to expand health insurance coverage to the total population, with the original stated objective to achieve universal health coverage by 2019.

2.2 Related literature on the impact of health insurance on health outcomes and utilisation

There is a growing body of evidence that health care utilisation increases as a result of providing health insurance (e.g. Trujillo et al. 2010, Yilma et al. 2015), and there is some support for a financial protection enhancing impact (e.g. Cheng, et al. 2015, Bai and Wu 2014, Aryeeteyet al. 2016). While country-level analyses have found that increasing health coverage through national-level health spending is beneficial for health, particularly within a system of risk-pooling (Moreno-Serra and Smith 2015), evaluations using micro data provide more mixed findings. For example, while in the USA and Taiwan, publicly funded health insurance programmes have been found to reduce infant and young children mortality (Currie & Gruber 1996, Chou et al., 2014), evidence from randomised insurance experiments suggest positive impacts to access and utilisation but little health benefit (Baicker et al. 2011). For LMICs, systematic reviews (Acharya et al. 2013, Erlangga et al. 2019a) found only a minority of evaluations that looked at health outcomes, with conclusive evidence of a health-improving impact in less than half of the primary studies reviewed (e.g. on health-related quality of life (Wang et al. 2009), infant mortality (Mensah et al., 2010) and glucose control (Sosa-Rubi et al. 2009)). The rest found either no evidence of a positive impact (e.g. Dow and Schmeer 2003, Chen and Jin 2012), or adverse impacts (Fink et al. 2013).

For Indonesia, quantitative impact evaluations of the different stages of health insurance expansions also reveal a mixed picture. Johar (2009) finds that the Health Card programme did not significantly increase health care utilisation among the poor, and attributes this finding to inelastic demand amongst the recipients. Evaluations of the *Askeskin* programme found some increase in financial protection (Aji et al. 2013), but only a modest impact on health care utilisation among the beneficiaries (Sparrow et al. 2013). An evaluation of the early implementation of the JKN programme (between 2007 and 2014) found that while contributory health insurance increased both inpatient and outpatient utilisation, subsidised health insurance only increased inpatient utilisation, and to a smaller extent (Erlangga et. al. 2019b).

There are various reasons why impact evaluations of health insurance expansions may not always demonstrate measurable improvements in health outcomes. First, establishing the causal effect of health insurance programmes is challenging due to selection into the health insurance programmes on observed and unobserved characteristics. Correcting for such selection bias requires exploiting 'natural experiments' through quasi-experimental econometric techniques (Wagstaff 2010). Second, demand-side policies (e.g. via health insurance expansion) on their own may not improve health, due to contextual factors of the health system (e.g. supply side constraints in terms of health care providers) and ineffective implementation of insurance programmes (Moreno-Serra and Smith 2012), and this concern has been thought to explain the modest impacts of Jamkesmas scheme in Indonesia (Harimurti et al. 2013). Third, the availability of health insurance may affect specific sub populations differently.

[3] As of 2013, around 12% of the population was estimated to have been covered by the *Jamkesda* schemes (32 million covered in 2013 out of a population of 252 million in that year). http://gnhe.org/blog/wp-content/uploads/2015/05/GNHE-UHC-assessment_Indonesia-1.pdf.http://gnhe.org/blog/wp-content/uploads/2015/05/GNHE-UHC-assessment_Indonesia-1.pdf. A further subsidised scheme (*Jampersal*) aimed to cover uninsured pregnant women and newborns was launched in 2011 with the specific aim of filling the gap in delivery services for maternal and neonatal health (Achadi, Achadi et al. 2014), and this insurance status was universal and not means tested.

Understanding the impact of health insurance for different populations is crucial to inform current and future health policymaking in Indonesia, where a large segment of the population is still uninsured, and where the government intends to provide subsidised health insurance for the poor households. Previous evaluations tended not to conduct subgroup analysis, with the exception of a few studies. Erlangga et al. (2019b) looked at impacts by subgroups, and found that the lowest income quintiles did not benefit from improved in-patient utilisation, with no effects in areas with low density of healthcare facilities. Anindya et al (2020) identified significant impacts of the JKN programme on maternal health care utilisation (skilled birth attendance, institutional deliver, antenatal care visits), and – in a subgroup analysis by socioeconomic quintiles and geographical regions – found that mothers from lower socioeconomic quintiles and more deprived regions benefitted more from health insurance.

2.3 Data

The IFLS household dataset includes respondents living in 13 out of the 27 Indonesian provinces, initially using the sampling frame of the 1993 national household socioeconomic survey (Survei Sosial Ekonomi Nasional – Susenas) from the Central Bureau of Statistics[4]. The first round of the survey was in 1993 (IFLS1), covering 7,224 households. Subsequent rounds were conducted with the same respondents and their new household members in 1997 (IFLS2), late 1998 (IFLS2+ with a 25% subsample), 2000 (IFLS3), 2007/2008 (IFLS4) and 2014/2015 (IFLS5). In order to exploit temporal variation in the availability of the health insurance schemes, we use the IFLS waves which were collected in the pre-*Askeskin* period (IFLS 3), in the pre-*Jamkesmas* period (IFLS4), and in the post-*Jamkesmas* period, covering the start of the *JKN* programme up to 2015 (IFLS5) (See Figure 1 for the links between the various policy reforms and survey waves).

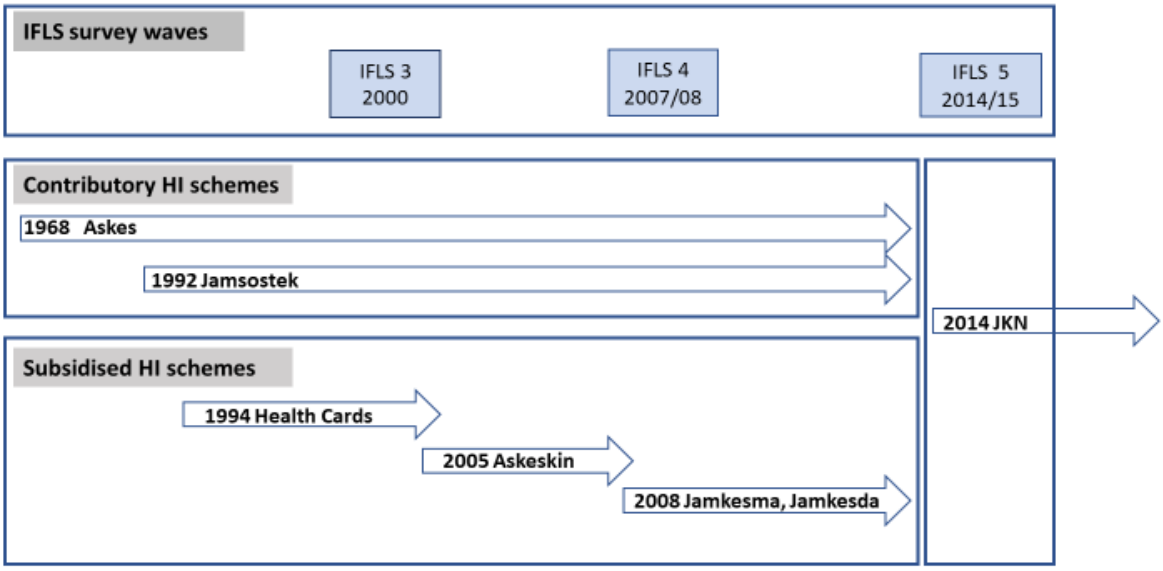


Figure 1 Health insurance expansions in Indonesia and the IFLS household survey waves

Notes: IFLS: Indonesia Family Life Survey, HI: health insurance, JKN: Jamima Kesehatan National (National Health Insurance)

[4] The sample is stratified in provinces and rural-urban areas within provinces. There are some randomly selected enumeration areas (EA) within the strata and households within enumeration areas. The aim of the selection of the provinces was to be cost-effective given the size of the country without neglecting the representation of the population, the ability to illustrate the cultural and socioeconomic diversity of Indonesia. In addition, the survey was designed to have a panel structured at the household level.

We constructed a birth-level dataset, and linked outcome information on births to insurance information of the mother, as well as her demographic information, her household and community. Complete pregnancy histories are available for women aged 15 to 49, including the date of birth of each child, whether the child is still alive, and if not, the age at death. In line with conventions, the death of a child is classified as 'infant death', if the death occurred before the first birthday. Restricting the recall period to 6 years to minimise recall bias, we collated births between 2002 and 2007 from IFLS4, and between 2008 and 2014 from IFLS5. We define an 'assisted birth' variable indicating whether the birth has been attended by either a midwife or a doctor, or both, regardless of place of delivery (both in and out of hospital). A mother can be 'uninsured', 'have subsidised insurance', and 'have contributory insurance' in the year of the birth. Subsidised insurance is defined as reporting enrolment in the *Health card*, *Askeskin, Jamkesmas, Jamkesda* or *JKN* schemes; contributory insurance is defined as reporting access to the *Askes* or *Jamsostek* or other employer provided insurance[5] (see Figure 1). Finally, 'uninsured' is defined as not having reported any subsidised or contributory insurance. Those births for which a mother reports having both subsidised and contributory insurance were excluded from the analysis[6], as such double insurance, while it did occur in practice, was not formally allowed[7].

We extracted a rich set of variables to control for confounding, focussing on the characteristics of the mothers, households and communities, which contribute to the eligibility and enrolment in the health insurance schemes, and which are also independently associated with health care utilisation or infant mortality[8]. Following previous studies (Dow and Schmeer 2003, Shrestha 2010), we included the following predictors of infant mortality: mother's education (categorised as primary, senior, secondary, and university), mother's literacy (ability to write a letter in Indonesian), age at birth, sex of the child, birth order of the child, and whether a household was urban or rural. To capture the means-testing eligibility criteria of the subsidised health insurance programmes (Johar 2009), we construct an asset index (O'Donnell et al. 2007), using principal component analysis (PCA) to classify households into wealth quintiles based on asset ownership and household characteristics (see Appendix A for specific variables used in the PCA). We also created a binary variable from the self-reported health of the mother (1 if good or excellent, 0 otherwise). To capture further indicators of socioeconomic deprivation, which could play a role in the eligibility for subsidised health insurance, we capture participation in three major social assistance programs: a subsidised rice ('Raskin') programme, an unconditional cash transfer programme, and a 'poor card' programme. We also added a variable capturing whether the household had been seriously affected by a natural disaster in the preceding five years.

[5] We have recoded mothers who have reported no health insurance if they were eligible based on the insurance status of their spouse or household head, and they reported being insured.
[6] Such double insurance constituted 5% of our overall sample.
[7] Because of the universal availability of the Jampersal programme, both insured and uninsured mothers may have reported "having" Jampersal. Hence, we did not include it in the definition of the health insurance variable.
[8] Variables that are only expected to affect enrolment in the subsidised or contributory health insurance schemes, but are unlikely to have a direct effect on infant mortality (or be affected by infant mortality themselves), were assessed as candidates for instrumental variables. However, none of them were found strong and valid at the same time.

Following previous work (Johar 2009), we also capture whether community members have access to a village midwife, a birth clinic, a hospital, a public health centre or private health care providers.[9] We have also extracted indicators for province of residence for the mother, at the time of the survey, to control for unobserved heterogeneity across provinces (e.g. in terms of geography), which may influence access to health care. A year of birth variable seeks to control for time trends affecting changes in infant mortality (e.g. technological innovations in neonatal intensive care), that may have coincided with the gradual expansion of health insurance. For the pre-specified subgroup analysis, we selected three widely used socioeconomic proxies to be able to assess the impact of insurance for those most vulnerable in terms of disease burden and access to health care (Lagomarsino et al. 2012): mothers with low education, those in the bottom socioeconomic quintiles, and those living in remote, rural communities.

[9] In order to strengthen our causal assumptions, we require that the observable variables included in our regressions are measured before a child is born, but also before a decision about enrolling in health insurance has been made. Hence, for births recorded in IFLS4 (2002-2007), we take measurements of individual and household level variables from IFLS3 (2000). Similarly, for births recorded in IFLS5 (2008-2014), we measure individual and household level variables from IFLS4 (2007/2008). For individuals who did not have a measurement in the previous wave, because they were not part of the IFLS sample yet (approximately 30% of the total sample), we take the current measurements as proxies. We follow a similar logic for missing household level covariates in the case of new people entering the IFLS sample (5% of the total sample missing). We construct indicator variables for these cases of missingness and include them in our analyses.

3. Methods

3.1 Potential outcomes and causal parameters

We are interested in estimating causal effects of a mother being enrolled in one of two health insurance types (subsidised or contributory) versus no health insurance, on one health outcome (infant mortality) and one health care utilisation outcome (assisted delivery) for a given birth, henceforth referred to as a unit. We conduct these analyses separately, and use a common notation Y for both outcomes, and W for both health insurance schemes. Following the Neyman-Rubin potential outcomes framework (see e.g. Abadie and Cattaneo, 2018), the potential outcome for a given birth i under no health insurance is denoted by $Y_i(0)$ and under a given health insurance programme by $Y_i(1)$. The causal individual treatment effect of health insurance for each unit is defined as the difference between the two potential outcomes, $\tau_i = Y_i(1) - Y_i(0)$. Since for a given unit only one of the two potential treatment states is observed, this individual treatment effect cannot be estimated from the observed data. Therefore, we aim for average estimands which can be identified under no interference, consistency[10], and no unobserved confounding after adjusting for the sufficient variable set X; is, that the potential outcomes are independent of the observed insurance status W, conditional on X:

$$Y(1), Y(0) \perp W \mid X \ .$$

Further, we require that the overlap assumption holds, that is, there must be a positive probability to be enrolled in a given health insurance programme, but this probability must be strictly smaller than 1: no covariate combination should fully determine a mother's insurance status:

$$0 < \Pr(W = 1 \mid X) < 1 \, .$$

Under these assumptions, the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the controls (ATC) can be identified. These three estimands answer different policy evaluation questions. The ATE estimand, defined as $E[Y(1) - Y(0)]$ contrasts the potential outcomes in a world where everyone has a given insurance, and where no one has insurance, and takes the average of these contrasts over the pooled population of the uninsured and the insured. The ATT, defined as $E[Y(1) - Y(0)|W = 1]$ answers the question: how much did those who had a certain insurance type benefit from having that health insurance, compared to not having insurance? Finally, the ATC, defined as $E[Y(1) - Y(0)|W = 0]$ aims to answer the question: how much the uninsured would have benefitted from having a given insurance type. The ATC estimand also allows us to contrast the impacts of the two insurance types, as the population for whom the benefits are calculated is held constant at the uninsured, representing a large portion of the population in Indonesia in the study period, including subgroups from all socioeconomic quintiles.

Beyond population average treatment effects, we also aim to evaluate how treatment effects vary across different subpopulations, by estimating the conditional average treatment effect (CATE) function, defined as:

$$\tau(x) = E[Y(1) - Y(0)|X = x].$$

[10] The no interference assumption requires that a unit's outcome is not affected by the treatment received by other units (Tchetgen and VanderWeele 2012). The consistency assumption requires that the observed outcome corresponds to the potential outcome under the observed treatment (VanderWeele 2009).

The CATE can be conceptualised as a function that takes a combination of observed covariates that are assumed to modify the effect of the treatment, at a selected covariate profile x, and outputs a treatment effect that corresponds to this covariate profile. In the context of health insurance, we expect that a range of the observed covariates can modify the treatment effect, beyond the socioeconomic factors listed above. The geographical availability of health services may be one such example.

3.2 Estimation of average treatment effects

3.2.1 Linear outcome regression

In the first set of analysis, we assume a linear probability model for each outcome of interest:

$$Y_i = \beta X_i + W_i\tau + \varepsilon_i\,, \quad [1]$$

where Y_i indicates (a) the survival status of infant i born in year t at 12 months after birth (b) whether the birth was attended by health professional, and the vector $X_i = (Z_{mt}, Z_{ht}, Z_{ct}, \delta_p, \alpha_t)$ includes several components: Z_{mt} denotes the characteristics of the mother (e.g. education) , Z_{ht} captures household characteristics (e.g. household asset quintile, social assistance), Z_{ct} community level variables (e.g. availability of hospital or birth clinic in the neighbourhood, or availability of a village midwife in the year of birth), δ_p are the effects of unobserved time-constant factors at the province level, α_t is the birth cohort indicator capturing shocks over time. W_i is the treatment of interest, i.e. whether in birth year of child i, the mother had a given health insurance ($W \in (0,1)$), τ is the treatment effect of interest. The residual term ε_i is assumed normally distributed, mean zero, and captures a composite of any unobserved province, community, household, mother and child level shocks**.** It follows from the previously stated assumptions that W_i is uncorrelated with the ε_i implying that any unobserved health shock to the mother, or income shock to the household, beyond those captured by the year fixed effects is unrelated to whether a mother is enrolled in health insurance in a given year. We employ separate regressions to estimate the impacts of subsidised and contributory health insurance, compared to no insurance, while in Appendix B we present an analysis where we pool the treatment groups, and estimate the impact on the outcomes of having any insurance vs. no insurance.

3.2.2 Propensity score weighted outcome regression

The outcome regression, eq. [1], assumes a homogenous additive treatment, hence τ cannot be directly interpreted as estimating either one of the ATE, ATT or ATC defined before. Moreover, the model relies on the linear relationship between the outcome and the covariates being correct (Ho et al. 2007). To address these restrictions, we also implement an inverse probability weighted (IPW) estimation of the outcome regression, using propensity scores (Rosenbaum and Rubin, 1983), defined as $p(X) = (W = 1|X)$, estimated via logistic regression including all the sufficient covariates as in [1][11]. We then use these inverse propensity scores to weight the linear outcome regression models, and obtain the estimated ATE, ATE and ATC by predicting both potential outcomes for each unit, and taking the relevant averages. This estimator is considered double-robust: consistent if either the regression or the propensity score model is correctly specified (Kang and Schafer, 2008). We implement this regression using the teffects command in Stata.

[11] Instead of province dummies, we use region dummies to adjust for confounding due to geographic region, due to convergence issues experienced in the weighted parametric regression models used by the `teffects` package. The use of region vs. province dummies made no difference to the results.

3.3 Estimation of heterogeneous treatment effects using causal forests

The ATE, ATT and ATC estimands allow for the causal effects to be different for those insured and uninsured, but do not capture their variation over the observed X covariates. Therefore, we now focus on the CATE estimand $\tau(x)$ that allows the treatment effects to vary as a function of covariates. We begin by considering a partially linear model for the outcome of interest, as before, that is:

$$Y_i = f(X_i) + W\tau + \varepsilon_i, \qquad [2]$$

with $f(X)$ an unspecified function, and initially, that τ, the treatment effect, is constant in X. Following Robinson (1988), we can re write this model in a 'centred' or residualised form as follows

$$Y_i - m(X_i) = \big(W_i - p(X_i)\big)\tau + \epsilon_i, \qquad [3]$$

where $p(X_i)$ is the propensity score as before, and $m(X_i) = E[Y_i|X_i]$ the conditional expectation of the outcome, marginalised over the treatment. The expressions $m(.)$ and $p(.)$ are often referred to as 'nuisance functions', and they can be estimated with any prediction algorithm, including ML methods. The causal effect τ can be estimated by solving eq. [3], and plugging in the predictions for $m(X_i)$ and $p(X_i)$ in the following formula:

$$\hat{\tau} = \frac{\sum_{i=1}^{n}\{(W_i - \hat{p}(X_i))(Y_i - \hat{m}(X_i))\}}{\sum\{W_i - \hat{p}(X_i)\}^2}. \qquad [4]$$

This corresponds to running a regression of the Y-residual on the W-residual. Such 'residualising' decreases the sensitivity of the resulting estimator to the errors in the estimates of the nuisance functions (Chernozhukov et al. 2018a). This can be extended to allow for heterogeneous treatment effects, assuming a sufficiently small neighbourhood $N(x)$ such that $\tau(x)$ is constant, which allows us to rewrite [4] as

$$\hat{\tau}(x) = \frac{\sum_{\{i:X_i \epsilon\, N(x)\}}\big\{\big(W_i - \hat{p}(X_i)\big)\big(Y_i - \hat{m}(X_i)\big)\big\}}{\sum_{\{i:X_i \epsilon N(x)\}}\{W_i - \hat{p}(X_i)\}^2} \qquad [5]$$

The main challenge for CATE estimation is how to choose $N(x)$. To solve this, Athey et al. (2019a) propose a generalised random forest approach, which conceptualises these neighbourhoods as a locally weighted set of neighbouring observations for a given value of x. The weights are estimated by performing a modification of the Random Forest regression algorithm (Breiman, 2001). In short, random forests calculate a predicted outcome for a unit by averaging the outcome of other units that are similar enough in covariates. The group of similar units are referred to as a leaf of a tree, and leaves are decided on by splitting the data based on cut-off values of the predictors, where the predictors to split on and cut-offs are decided so that the resulting splits minimise the prediction error in the sample. To reduce the noise stemming from using individual trees as predictors, this is done many times over bootstrapped samples of the data, and final predictions for each observation are obtained as the average of predictions over the bootstrap samples.

Generalised random forests build on this algorithm, but modify it in important aspects, to ultimately minimise the bias in the estimated CATE. First, the outcomes and treatment are residualised as described before. Second, the splits of the data ('the causal trees') are formed by running the local linear regressions (eq. [3]) in each candidate split. Instead of choosing splits to minimise prediction error, they are chosen so that within a leaf, estimated treatment effects are similar (corresponding to homogenous treatment effects within a leaf), while between leaves, they differ (capturing treatment effect heterogeneity across units with differing X values). This procedure is performed on many bootstrap samples, thus forming causal forests. The causal forests are then used to calculate

$\alpha_i(x)$ weights for each observation, based on how frequently an observation was used to estimate the treatment effect at x. The resulting weights are employed in an estimator of the CATE that modifies [4] as follows:

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)\{(W_i - p(X_i))(Y_i - m(X_i))\}}{\sum \alpha_i(x)\{W_i - p(X_i)\}^2}. \qquad [6]$$

Individual treatment effects $\hat{\tau}(X_i)$ can be estimated by evaluating $\hat{\tau}(\mathrm{x})$ at the covariate combination of each unit. Average treatment effects can also be obtained, by plugging in the estimated $\hat{\tau}(X_i)$ in a variant of the augmented inverse probability weighting estimator (Robins, Rotnitzky, and Zhao, 1994):

$$\hat{\tau} = \sum_{i=1}^{n^D} \hat{\tau}(X_i) + \frac{W_i - p(X_i)}{p(X_i)(1-p(X_i))} ((Y_i - m(X_i) - (W_i - p(X_i))\hat{\tau}(X_i)), \qquad [7]$$

where the summation is taken over n^D, that stands for the sample of the treated, the control or the treated plus control samples, depending on whether the causal estimand is the ATT the ATC or the ATE, respectively. This formula also provides the subgroup average treatment effects, constraining the summation for units in the subgroups of interest (e.g. women with primary education only).

We apply the causal forest approach as implemented in the grf R package (Tibshirani et al. 2018) and follow the steps, suggested by Athey and Wager (2019b), below:

1. Fit 'traditional' regression forests to estimate $m(X_i)$ and the $p(X_i)$, then calculate residualised outcomes using these quantities. We use 500 trees to select the tuning parameters, and 1000 trees to obtain the predictions.
2. Grow a causal forest on 1000 bootstrap samples (with 500 trees to select tuning parameters). We rank variables used in terms of variable importance in the resulting causal forest (based on count of the proportion of splits on the given variable) and select those with higher than mean variable importance measure.
3. Re-grow the causal forest, using only the selected effect modifiers, using 500 trees for tuning, and 3000 trees for predicting ITEs. The residualising step [1] ensures that using only a subset of the covariates will not result in incomplete adjustment for confounding.
4. Estimate individual level treatment effects by evaluating the resulting $\hat{\tau}(x)$ function for each unit's own covariate values. We estimate ATE, ATT, ATC, and subgroup ATCs for each pre-specified subgroup.
5. Assess the heterogeneity captured by the resulting causal forests the following ways:
    a. Plot the estimated individual level CATEs, with their estimated confidence intervals.
    b. Perform a test for the presence of overall heterogeneity captured by the $\hat{\tau}(x)$ estimate (Chernozhukov et al. 2018b). This test assesses whether $\hat{\tau}(x)$ captures any further information than simply using the ATE, $\hat{\tau}$ to "predict" the individual level treatment effects.
    c. Assess the final ranking of the variable importance measure, form further subgroups based on the top ranked variables, and contrast the differences in the average treatment effects across these subgroups.
    d. Split individuals into two groups based on their estimated CATEs (below and above median), and describe these groups in a number of key characteristics.

We implement this approach for the skilled birth attendance outcome variable, and fit separate causal forests for the subsidised and contributory health insurance. The covariates used in Steps 1-2 include all the individual, household, community level variables used in the previous analyses, as well as the year of birth dummies and province dummies.

4. Results

4.1 Descriptive statistics

Table 1 and Figure 2 highlight that while the majority of births recorded in our dataset were not covered by any insurance scheme, subsidised health insurance saw a steep increase from 2005, while infant mortality decreased and the proportion of births assisted by a midwife or physician demonstrated a clear upwards trend. In Table 3, we contrast the observed characteristics of the three groups: births insured by subsidised insurance, births insured by contributory insurance, and births not covered by health insurance in the year of birth, comparing the means and standardised differences for each treatment group to the control group. Most variables display large differences (standardised differences >10%), with births under subsidised insurance being more likely to be from a rural household and from mothers who are older at birth, less likely to have studied at university and more likely to have only elementary school education, belong to lower wealth quintiles, and receive social assistance programmes, compared to those without subsidised insurance. By contrast, while those mothers with contributory insurance are also somewhat older at the time of birth than the uninsured, they are also more likely to have a university education, and are overrepresented among households within the highest asset index quintiles. A quarter of these mothers received subsidised rice, while only a small fraction received cash transfer (7%) or held a 'Poor card' (4%). We interpret these large differences as indicative of a strong confounding of the relationship between health insurance and the outcomes of interest.

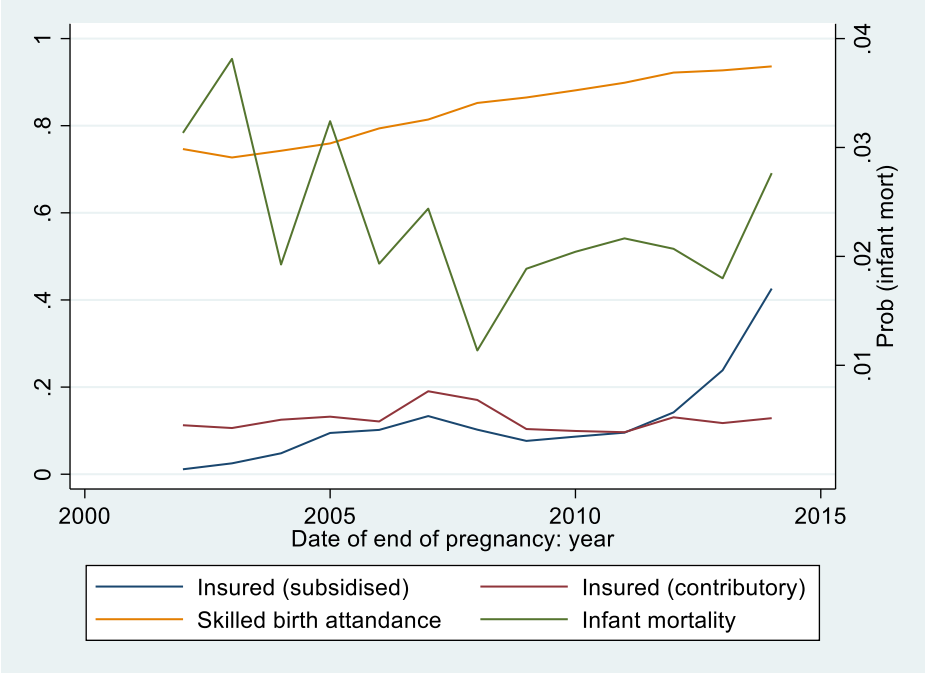


Figure 2 Trends in the probability of infant mortality, health care utilisation, and in the proportion of births covered by subsidised and contributory insurance

Notes: Left axis: proportion of births covered by each insurance type, in a given year, proportion of birth attended by skilled health professional; right axis: probability of infant mortality

Table 1 Panel A: Insurance status of the mother a year of birth, Panel B: Absolute infant mortality, by year of birth

Panel A: insurance status by year of birth				
Year of birth	**Uninsured**	**Subsidised insurance**	**Contributory insurance**	**Total**
2002	615	8	79	702
2003	729	21	89	839
2004	687	40	104	831
2005	620	76	106	802
2006	603	79	94	776
2007	582	115	164	861
2008	875	59	116	1,050
2009	782	73	99	954
2010	877	93	107	1,077
2011	821	97	98	1,016
2012	773	151	139	1,063
2013	680	252	124	1,056
2014	467	447	135	1,049
Total	9,111	1,511	454	12,076
Panel B: Outcomes by insurance status				
Infant mortality (%)	2.57	2.31	1.12	2.37
Assisted birth (%)	82.46	85.37	95.24	84.4

Table 2 Descriptive statistics by insurance status, before and after propensity score reweighting

	Uninsured	Subsidised HI				Contributory HI			
	mean	mean	SMD (raw)	SMD (PS weighted, ATE)	SMD (PS weighted, ATT)	Mean	SMD (raw)	SMD (PS weighted, ATE)	SMD (PS weighted, ATT)
Mother characteristics									
Age	27.14	28.01	14.0%	0.5%	-0.3%	29.10	34.4%	-9.4%	0.3%
Health (good)	0.87	0.83	-8.8%	3.4%	-4.5%	0.87	2.0%	0.5%	0.4%
Educ: primary	0.32	0.38	12.7%	5.9%	-0.6%	0.09	-59.0%	-1.6%	0.3%
Educ: secondary	0.26	0.28	4.5%	-2.0%	-3.3%	0.14	-30.6%	3.6%	-0.1%
Educ: senior	0.33	0.29	-8.6%	-1.1%	3.5%	0.41	17.1%	-2.2%	1.0%
Educ: higher	0.10	0.06	-15.4%	-5.3%	0.9%	0.36	66.6%	0.7%	-1.2%
Writes (Indonesian)	0.96	0.95	-2.8%	-1.0%	0.4%	0.99	18.6%	5.0%	-0.7%
Household characteristics									
Wealth quint 1	0.19	0.30	24.0%	2.6%	-0.9%	0.04	-51.4%	-0.3%	0.2%
Wealth quint 2	0.21	0.25	9.1%	4.2%	-3.4%	0.09	-34.5%	2.7%	0.3%
Wealth quint 3	0.22	0.21	-2.1%	0.3%	1.5%	0.17	-14.0%	-3.5%	0.2%
Wealth quint 4	0.20	0.16	-11.1%	0.1%	1.9%	0.30	22.0%	0.9%	-0.9%
Wealth quint 5	0.17	0.08	-27.5%	-8.7%	2.0%	0.41	55.3%	0.3%	0.4%
Raskin	0.50	0.72	45.6%	10.8%	-4.4%	0.26	-51.8%	-2.0%	-0.1%
Cash transfer	0.23	0.45	47.0%	7.7%	-5.2%	0.07	-46.1%	-2.3%	0.2%
Poor card	0.09	0.20	31.6%	5.3%	-2.6%	0.04	-18.5%	1.7%	0.6%
Rural	0.48	0.47	-2.3%	8.3%	-1.3%	0.28	-43.3%	-0.7%	0.7%
Disaster	0.23	0.28	10.2%	7.8%	-1.1%	0.24	1.5%	-3.4%	0.4%
Availability of health services in community									
Birth clinic	0.99	1.00	4.1%	4.3%	0.3%	0.99	0.4%	2.7%	-0.4%
Health centre	0.97	0.98	3.2%	5.0%	-0.7%	0.97	-1.4%	5.1%	-1.6%
Private practice	0.96	0.95	-1.8%	1.2%	-2.0%	0.94	-7.9%	2.2%	0.4%
Hospital	0.90	0.93	9.9%	-3.2%	0.3%	0.89	-4.6%	6.0%	-2.0%
Midwife	0.82	0.83	3.0%	12.4%	-0.7%	0.76	-15.0%	1.8%	0.3%
Child characteristics									
1st child	0.68	0.52	-33.7%	1.0%	0.9%	0.64	-10.0%	6.3%	-0.9%
2nd child	0.25	0.36	22.9%	-0.5%	-1.5%	0.30	9.8%	-3.0%	0.7%
>= 3rd child	0.06	0.12	20.1%	-1.0%	0.9%	0.07	1.5%	-7.0%	0.4%
Female	0.49	0.49	1.5%	-0.1%	1.0%	0.49	1.5%	0.5%	0.3%
Regions									
Sumatera	0.27	0.23	-8.3%	5.5%	0%	0.22	-9.8%	-8.9%	1.4%
Jakarta	0.06	0.04	-5.1%	-10.7%	2%	0.09	12.6%	1.0%	-1.0%
Jawa	0.44	0.44	-1.5%	-6.6%	0%	0.46	2.9%	16.3%	-2.7%
Bali,NTB,NTT	0.13	0.19	15.7%	2.4%	1%	0.10	-8.9%	-12.9%	1.5%
Kalimantan	0.05	0.02	-14.7%	6.3%	0%	0.07	5.9%	3.3%	0.3%
Sulawesi	0.05	0.08	10.4%	2.9%	-1%	0.06	4.3%	-6.6%	2.0%

Notes: educ: education, quint: quintile, SMD: standardised mean difference, PS: propensity score, ATE: average treatment effect, ATT: average treatment effect among the treated, IPW: inverse probability of treatment weighting

4.2 Results: Average treatment effects

Table 3 displays the linear regression estimates for the effect of health insurance on infant mortality and assisted birth. For subsidised health insurance, the unadjusted (naïve) results for infant mortality are small and insignificant, but adjustment for observed confounders leads to an increase in the magnitude of the estimates, in line with the expected sign of bias, given the worse socioeconomic profile of mothers enrolled into subsidised insurance, compared to the uninsured ones. Even after covariates adjustment, year and province dummies, there is no evidence that the estimated effects of subsidised insurance on infant mortality are different from the null. For the contributory health insurance, the estimated coefficients move in the opposite direction, and decrease after covariate adjustment; however, there is strong evidence that resulting point estimates still indicate a large protective (i.e. infant mortality-reducing) insurance effect of 0.9 percentage point ($p < 0.01$). When investigating the assisted birth outcome, the regression estimates show an evidence of increase in assisted births among mothers with contributory health insurance, with an estimated increase of 3 percentage points. At the same time, there is a somewhat weak evidence of an increase for those with subsidised health insurance (2 percentage points) ($p<0.1$). (Results of the regression analyses with 1) pooling the two insured and uninsured populations, using a categorical insurance variable, and 2) using an 'any insurance' treatment variable are reported Appendix Table 1.)

Table 2 and Appendix Figure 1 describe the covariate balance achieved after inverse probability weighting using the estimated propensity scores for both treatment groups compared to the control group, and contrasts these to the unweighted balance. Using weights that aim to recreate the distribution for the treated (ATE, ATC and ATT weights), the balance improves for each covariate, and standardised differences stay above 10% for only a few covariates, and the ATE weights showing somewhat worse balance than the ATT and ATC weights. Appendix Table 2 displays the distributions of the estimated propensity scores. While there is a good overlap between the propensity score distributions for both insurance types, there is a large mass around zero for the uninsured, implying that many of those who did not get the insurance were unlikely to get it based on their observed covariates.

Table 3 (Panel A) reports the IPW regression estimates for each estimand, for both outcomes. For infant mortality, there is no evidence of an effect of subsidised health insurance, while we found evidence of a somewhat larger (when compared to the regression estimates) effect of the contributory health insurance. Here, estimated ATE and ATC are larger than the ATT, indicating that the uninsured would have benefitted more from the insurance than those who were actually insured. This pattern repeats with the assisted birth outcome, for both insurance types: the benefit in terms of increased access is larger among the untreated than among the treated, while these estimated effects are significant ($p<0.01$) for the contributory health insurance, and not significant for the subsidised health insurance. As the populations of uninsured and those with the two health insurance types widely differ in terms of observed characteristics, we expect that the differences between the estimated ATTs, ATC, and ATEs can be explained with effect modification due to the observed covariates. We explore this in the next section.

Table 3 Linear regression results for the effect of health insurance on Infant mortality and assisted birth outcomes

Panel A: Subsidised HI				
	Unadjusted 1	Unadjusted 2	OLS1	OLS2
Infant mortality				
Estimate	-0.0025	-0.0020	-0.0056	-0.0055
(SE)	(0.0047)	(0.0052)	(0.0052)	(0.0053)
Observations	10,622	10,622	10,622	10,622
Skilled birth attendance				
Estimate	0.0291***	-0.0295***	0.0163	0.0183*
(SE)	(0.0104)	(0.0110)	(0.0108)	(0.0108)
Observations	9,834	9,834	9,834	9,834
Panel B: Contributory HI				
	Unadjusted 1	Unadjusted 2	OLS1	OLS2
Infant mortality				
Estimate	-0.0126***	-0.0130***	-0.0088**	-0.0093**
(SE)	(0.0034)	(0.0035)	(0.0038)	(0.0039)
Observations	10,565	10,565	10,565	10,565
Skilled birth attendance				
Estimate	0.1279***	0.1225***	0.0237***	0.0294***
(SE)	(0.0079)	(0.0080)	(0.0080)	(0.0081)
Observations	9,732	9,732	9,732	9,732
Year dummies	N	Y	Y	Y
Covariates	N	N	Y	Y
Province dummies	N	N	N	Y

Notes: HI: Health Insurance, OLS: Ordinary least squares, *** p<0.01, ** p<0.05, * p<0.1

Table 4 Estimates of average treatment effects. Panel A: Inverse probability of treatment weighted regression Panel B: Causal Forests

Infant mortality	**Panel A1: IPW-Regression**	
Subsidised health insurance	**Estimate (SE)**	
ATE	-0.0026 (0.0058)	
ATC	-0.0058 (0.0055)	
ATT	-0.0026 (0.0052)	
Contributory health insurance		
ATE	-0.0147***(0.0033)	
ATC	- 0.0157***(0.0033)	
ATT	-0.0101**(0.0041)	
Assisted birth	**Panel A2: IPW-Regression**	**Panel B: Causal Forests**
Subsidised health insurance	**Estimate (SE)**	**Estimate (SE)**
ATE	0.0206 (0.0136)	0.016 (0.0115)
ATC	0.0231 (0.0149)	0.016 (0.012)
ATT	0.0120 (0.0111)	0.011 (0.0093)
Contributory health insurance		
ATE	0.0584***(0.0159)	0.055 (0.0109) ***
ATC	0.0639***(0.0176)	0.060 (0.012)***
ATT	0.0239***(0.0070)	0.024 (0.0058)***

Notes: SE: standard error, ATE: average treatment effect, ATT: average treatment effect among the treated, ATC: average treatment effect among the controls, IPW: inverse probability of treatment weighting, *** p<0.01, ** p<0.05, * p<0.1

4.3 Results: heterogeneous treatment effects

Figure 3 presents the distribution of the estimated causal individual-level treatment effects: the histograms of the point estimates (left panel) and the ranked estimates (solid line) with their 95% confidence intervals. For the subsidised health insurance, the estimated confidence interval (CIs) includes zero for the whole distribution, while for contributory health insurance, a small fraction of CIs exclude zero. The formal test for treatment effect heterogeneity indicates the presence of heterogeneity for the contributory health insurance (p=0.003), but not for the subsidised health insurance (p=0.69).

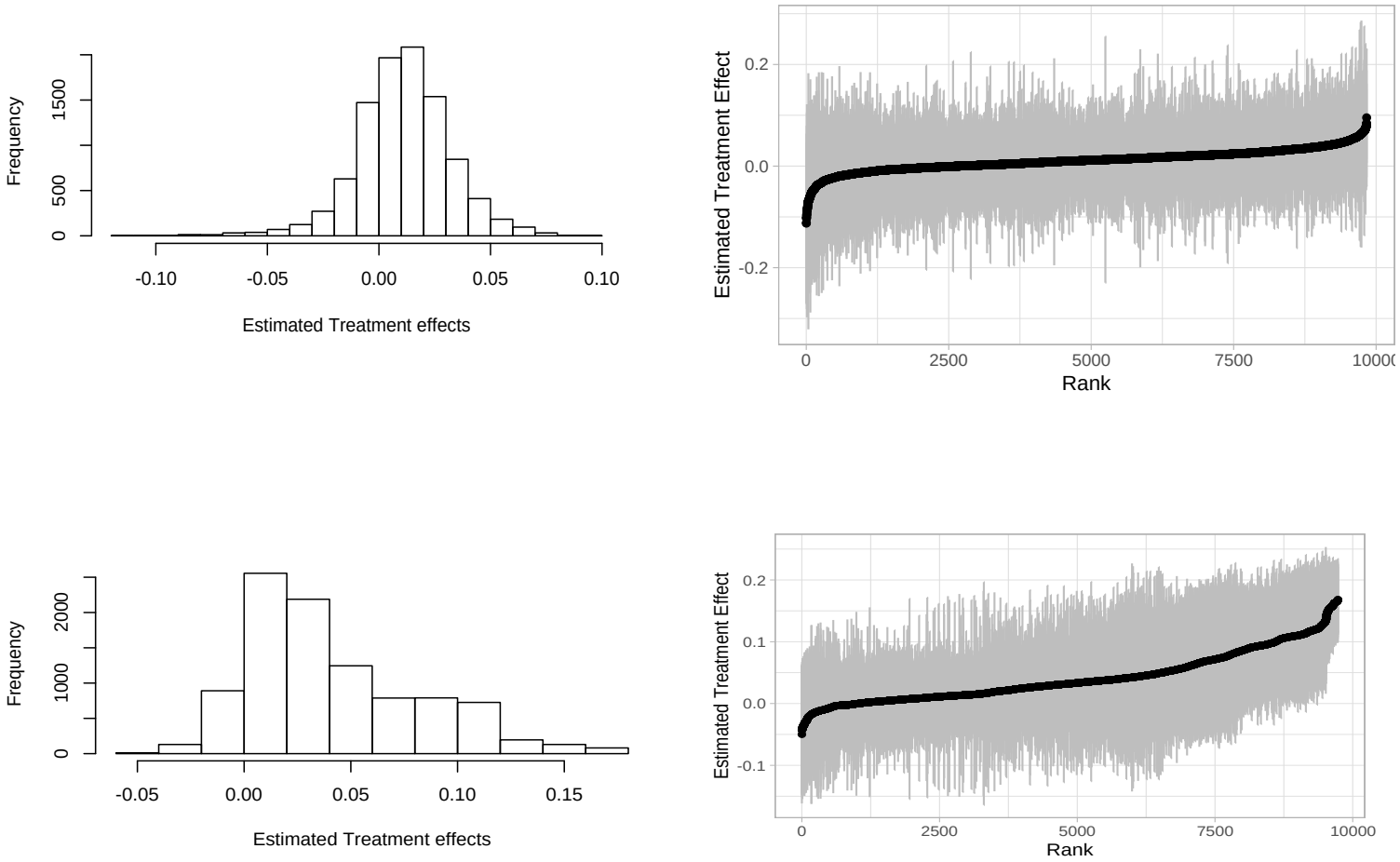


Figure 3 Estimated (conditional) individual level treatment effects. Top panel: subsidised health insurance; bottom panel: contributory health insurance. Left panel: distribution of the point estimates. Right panel: point estimates (black curve) and 95% confidence intervals, ranked

Notes: IPWRA: inverse probability weighted regression adjustment, CF: causal forests, ATE: average treatment effect, ATT: average treatment effect among the treated, ATC: average treatment effect among the controls

Table 5 presents the ranking of covariates in terms of their importance in predicting treatment effect heterogeneity, in terms of utilisation of skilled assistance when giving birth. For the contributory health insurance scheme, these largely overlap with the pre-specified socioeconomic covariates: education, wealth quintiles, and the rurality of the household. The most important variable associated with the estimated heterogeneous effect was the indicator for East Java province: a relatively industrialised region of Indonesia. For the subsidised health insurance scheme, the most influential variables were mother's age and the birth order of the child, followed by being in receipt of cash transfers and possessing a poor card.

Panel B of Table 4 and Figure 4 present the average treatment effects reported by the causal forest estimator, which closely correspond to the IPW-regression estimates. Figure 5 and Table 6 presents the conditional average treatment effects among the controls for pre-specified subgroups (top panel) as well as subgroups constructed based on the variables suggested by the final causal forests variable importance (right panel). We detect large differences in subgroup ATCs for contributory health insurance corresponding to subgroups suggested by variable importance: there is a strong gradient in terms of wealth quintiles in the estimated subgroup effects, and there are also considerable differences in ATC reported between those with different education levels, and between rural and urban communities. The differences in the subgroup effects, while showing a similar direction, are much less pronounced for the subsidised health insurance, and there is no evidence in support of a subgroup ATCs being different from zero. Among the subgroups discovered by the causal forest algorithm, for the subsidised scheme we found some evidence ($p<0.05$) of treatment effect for the subgroup with the third or higher birth order. None of these results were found to be sensitive to the choice of tuning parameters for the causal forest algorithm, which were selected outside of the cross-validation algorithm (number of trees used for tuning, number of trees used for the final Causal Forests). We present the selected tuning parameters in the Appendix Table 3.

As a final, exploratory analysis, Appendix Table 2 compares the characteristics of mothers when they are grouped based on the estimated individual level CATEs, using the median value as the cut-off. It appears that mothers who benefitted relatively more from the subsidised health insurance are older, more likely to be in lower wealth quintiles, and more likely to have received cash transfer or rice subsidy, than those in the lower half of the treatment effect distribution. Those benefitting most from contributory health insurance are also more likely to belong to the lower wealth quintiles, less likely to have had higher levels of education, and twice as likely to have received subsidies, compared to those in the lower half of the distribution. There is no difference in the availability of health services among the two groups.

Table 5: Covariate importance in explaining treatment effect heterogeneity (10 highest importance covariates)

	Subsidised HI		Contributory HI	
Ranking	**Variable importance measure**	**Variable**	**Variable importance**	**Variable**
1	0.126	Birth order >=3	0.127	Province East Java
2	0.085	Birth year 2012	0.123	Higher education
3	0.084	Age >=31	0.083	Wealth quantile 4
4	0.075	Past covariates imputed	0.069	Province South Kalimantan
5	0.066	Cash transfer	0.066	Rural community
6	0.065	Poor card	0.060	Wealth quantile 5
7	0.063	Birth year 2014	0.055	Province West Sumatra
8	0.062	Birth order =2	0.049	Private practice in community
9	0.054	Province West Nusa Tenggara	0.048	Senior education
10	0.046	Natural disaster	0.045	Province Banten

Notes: The variable importance measure is based on the count of the proportion of splits on the given variable

Table 6: Estimated conditional average treatment effects, for pre-specified subgroups and subgroups discovered by the causal forest algorithm

	Subsidised HI		Contributory HI	
	CATC	**SE**	**CATC**	**SE**
All (uninsured)	0.0162	0.0125	0.0596***	0.012
Pre-specified subgroups				
1-2nd quint	0.0313	0.0252	0.1299***	0.0379
3-4th quint	0.011	0.0147	0.0351***	0.0121
5th quint	0.0055	0.025	0.0009	0.0117
No/primary educ	0.0206	0.0204	0.0985***	0.0285
Secondary educ	0.0232	0.0166	0.0408***	0.0153
Higher educ	0.0093	0.0328	0.008	0.0132
Rural community	0.0182	0.0209	0.0844***	0.0246
Urban community	0.0143	0.0142	0.0375***	0.0095
Discovered subgroups				
<=31y	0.0115	0.0149		
>31y	0.0309	0.0225		
1st-2nd child	0.0132	0.0133		
3rd+ child	0.0602**	0.030		
Java			0.1324***	0.0213
Rest of Indonesia			0.0482***	0.0137

Notes: HI: health insurance, CATC: conditional average treatment effect among the controls, SE: standard error quint: quintiles, educ: education, *** p<0.01, ** p<0.05, * p<0.1

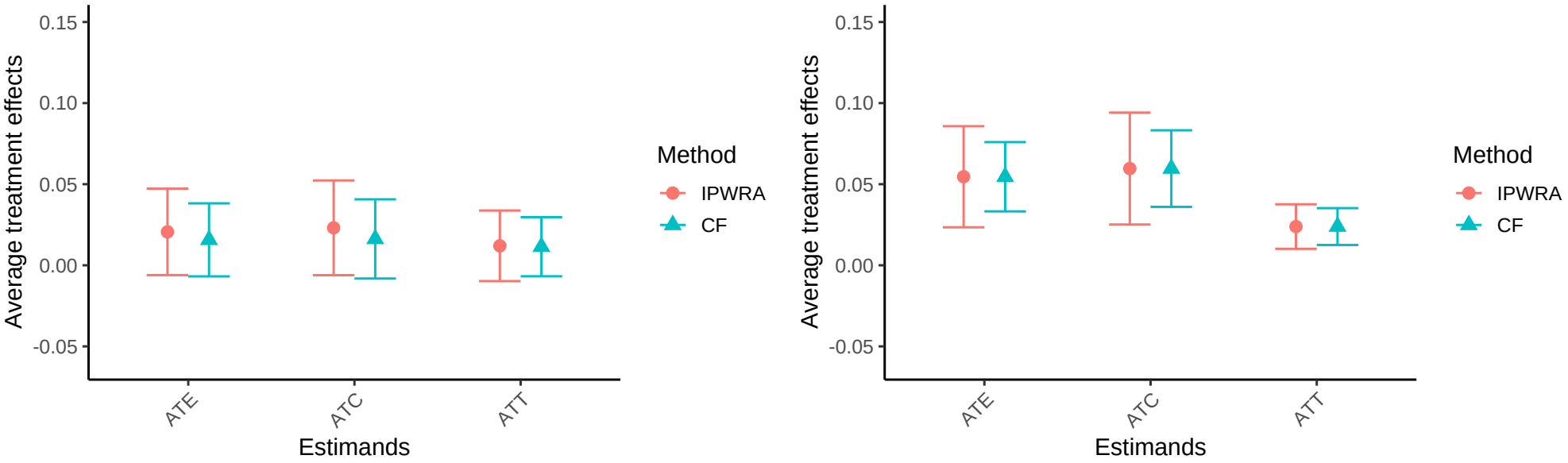


Figure 4 Estimated average treatment effects. Left panel: subsidised health insurance; right panel: contributory health insurance

Notes: IPWRA: inverse probability weighted regression adjustment, CF: causal forests, ATE: average treatment effect, ATT: average treatment effect among the treated, ATC: average treatment effect among the controls

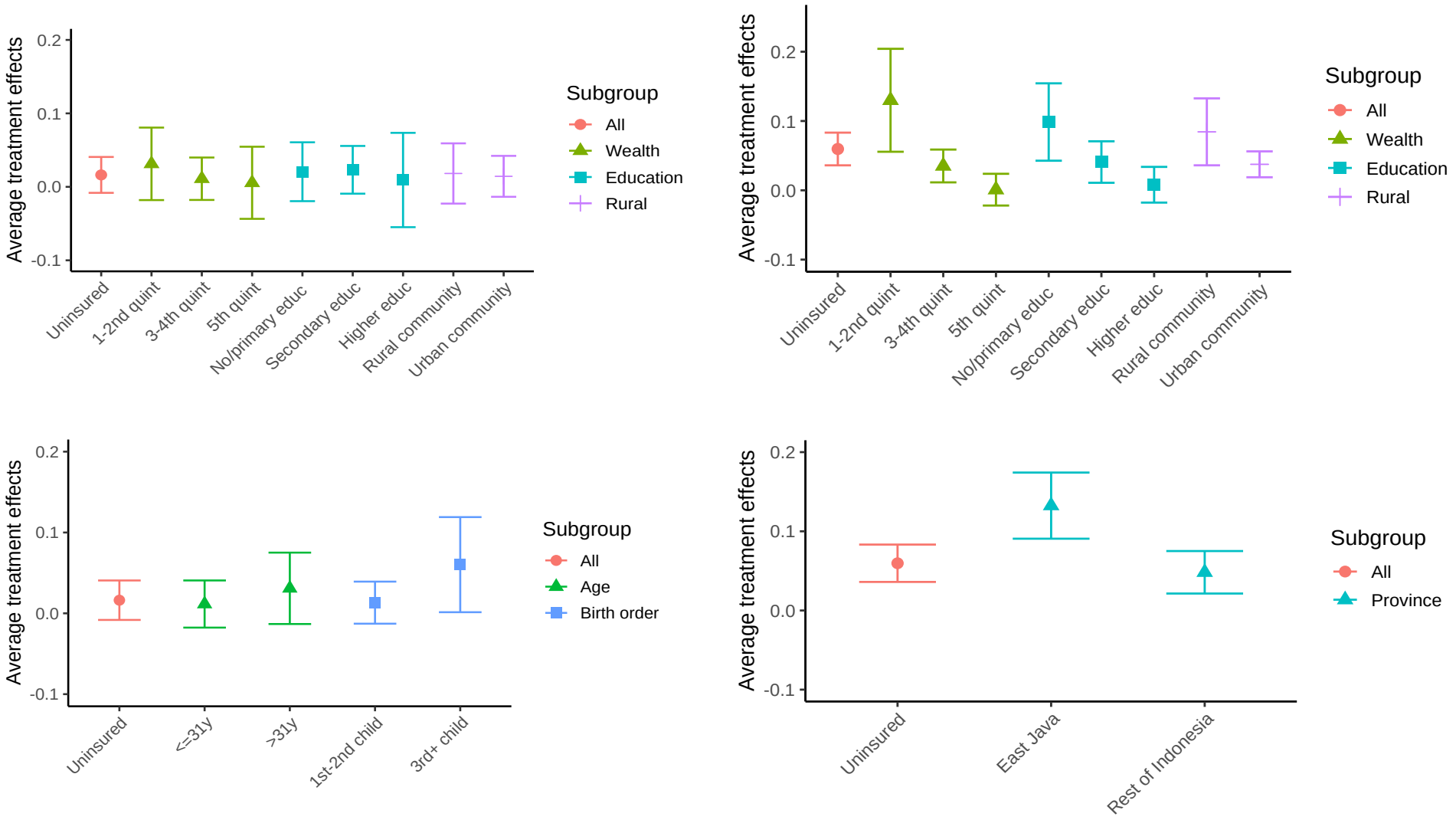


Figure 5 Estimated subgroup average treatment effects. Left panel: subsidised health insurance; right panel: contributory health insurance. Top panel: pre-specified subgroups. Bottom panel: subgroups discovered by the causal forest algorithm

5. Discussion

In this paper, we estimated the average and heterogeneous effects of two main Indonesian health insurance types on infant mortality, and on maternal health care utilisation at the time of delivery. We find that enrolment in contributory health insurance reduced infant mortality on average by 0.9 ($p < 0.05$) percentage points, corresponding to a sizeable 30% reduction from the average infant mortality rate (i.e. infant deaths per 1000 live births) over the period. By contrast, we found no evidence of non-zero effect of subsidised health insurance. Our findings for the health care utilisation outcome may help explain these results: contributory insurance increased the expected probability of having a birth attended by a healthcare professional, but there was no evidence of such an effect for the subsidised scheme. Our findings mirror the previous evidence that found small to negligible impact of subsidised health insurance schemes on health services utilisation (Johar 2009, Sparrow et al. 2013, Erlangga et al. 2019a), and they are also consistent with the findings of Anindya et al. (2020) that found that the JKN programme significantly improved the utilisation of skilled birth attendance, in a population that pooled the recipients of subsidised and contributory insurance.

We delved deeper into this, by examining the heterogeneity in the effects: for both insurance schemes, the estimated causal effects on health care utilisation among the uninsured appear to be higher than among the insured (that is, on average, we expect those uninsured in the study period would benefit from being insured more than the expected benefit estimated amongst those who are insured). Indeed, we found that the benefits, in terms of increased access to assisted birth, are relatively higher among the more vulnerable subgroups, for both insurance types – reflecting the findings of Anindya et al. (2020) - and these differences are more pronounced for the contributory health insurance. While pre-specified socioeconomic variables ranked high in terms of being associated with treatment effect heterogeneity, we found further variables that according to the variable importance of the causal forest algorithm were more strongly associated with treatment heterogeneity: for example, women residing in certain provinces (e.g. East Java) would have benefited more than other subgroups, had they been insured (contributory vs remaining uninsured). For subsidised health insurance, none of the pre-specified population subgroups had a significant causal treatment effect. However, the variable importance of the CF algorithm suggested the child's birth order explains significant treatment heterogeneity, with the subgroup of children who were third born or higher having the highest causal average treatment effect, with a 95% CI that excluded zero.

A significant limitation of our analysis is that we could not conduct a subgroup analysis for the IMR outcome due to the very rare nature of the outcome variable (approximately 300 events out of 12,000). Furthermore, because we use household survey data, covariate information was collected at discrete time points, which were assumed to provide valid baseline measurements for births that happened closer or further away from the survey dates. This measurement error can also lead to a downward bias in the estimated coefficients.

This paper is the first study to characterise the effect heterogeneity of a health policy intervention by employing causal forests. We highlighted the role of such causal machine learning approaches in exploring the heterogeneity of policy impacts according to observed covariates, beyond what can be captured by pre-specified subgroup analyses. Our study also highlights a crucial challenge when using this approach to estimate treatment heterogeneity in an observational framework: the need to adjust for observed confounding (indicated by large baseline imbalances in individual, community and household characteristics). The close agreement between the average treatment effects reported by a parametric IPW regression model and causal forests provides some reassurance that the adjustment for observed confounding was successful in both cases. Nevertheless, it may be the case that despite the variety of methods we employed, we still may not have captured all unobserved confounding, leading to biased estimates. First, the eligibility assessment was complex, based on geographical and household level criteria that are not fully reflected in our dataset. Second, take-up of the insurance was ultimately voluntary, leaving the possibility that those who are somewhat better off were less likely to opt for subsidised insurance, due to the perceived stigma and potentially lower quality of services, compared to those to be obtained in the private sector.

Our study highlights two potential further avenues for research: First, generalised random forests can be used to report conditional average treatment effects even in the presence of unobserved confounding, where instrumental variables are available (Athey et al. 2019a). Second, the estimated individual level treatment effects can be used to formulate so-called 'optimal policy rules': treatment assignment mechanisms that maximise a pre-specified welfare function of the decision-maker (Athey and Wager 2020). For the Indonesian context, such optimal policy rules could inform health policymaking in the country by guiding the re-design of the eligibility criteria for subsidised health insurance, which may be needed given the fiscal challenges brought by the move towards Universal Health Coverage. Beyond the specific case of Indonesia, the method may be used to help target policy efforts towards where the greatest potential benefits can be realised, thereby helping to pinpoint where adaptation of policy may be needed, be it in terms of eligibility or otherwise. In doing so, this will enable researchers to move policy impact evaluations beyond simple binary judgements on whether something 'works' or not, towards matters of for whom policies 'work' and how these can be improved.

References

Abadie A, Cattaneo MD. Econometric methods for program evaluation. *Annual Review of Economics*. 2018 Aug 2;10:465-503.

Achadi EL, Achadi A, Pambudi E, Marzoeki P. A study on the implementation of Jampersal policy in Indonesia. 2014. *Open Knowledge Repository*. https://openknowledge.worldbank.org/handle/10986/20740

Acharya A, Vellakkal S, Taylor F, Masset E, Satija A, Burke M, Ebrahim S. The impact of health insurance schemes for the informal sector in low-and middle-income countries: a systematic review. *The World Bank Research Observer*. 2013 Aug 1;28(2):236-66.

Agustina R, Dartanto T, Sitompul R, et al. Universal health coverage in Indonesia: concept, progress, and challenges. *Lancet*. 2019;393(10166):75-102. doi:10.1016/S0140-6736(18)31647-7

Aji B, De Allegri M, Souares A, Sauerborn R. The impact of health insurance programs on out-of-pocket expenditures in Indonesia: an increase or a decrease? *International Journal of Environmental Research and Public Health*. 2013 Jul;10(7):2995-3013.

Anindya K, Lee JT, McPake B, Wilopo SA, Millett C, Carvalho N. Impact of Indonesia's national health insurance scheme on inequality in access to maternal health services: A propensity score matched analysis. *Journal of Global Health*. 2020 Jun;10(1).

Aryeetey GC, Westeneng J, Spaan E, Jehu-Appiah C, Agyepong IA, Baltussen R. Can health insurance protect against out-of-pocket and catastrophic expenditures and also support poverty reduction? Evidence from Ghana's National Health Insurance Scheme. *International Journal for Equity in Health*. 2016 Dec 1;15(1):116.

Athey S, Imbens GW. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives.* 2017 May;31(2):3-2.

Athey S, Tibshirani J, Wager S. Generalized random forests. *The Annals of Statistics*. 2019a;47(2):1148-78.

Athey S, Wager S. Estimating treatment effects with causal forests: An application. *arXiv preprint* arXiv:1902.07409. 2019b.

Athey S, Wager S. *Policy Learning with Observational Data*. 2020. https://www.econometricsociety.org/system/files/15732-4.pdf

Bai CE, Wu B. Health insurance and consumption: evidence from China's new cooperative medical scheme. *Journal of Comparative Economics*. 2014 May 1;42(2):450-69.

Baicker K, Finkelstein A. The effects of Medicaid coverage—learning from the Oregon experiment. *The New England Journal of Medicine*. 2011 Aug 25;365(8):683.

Breiman L. Random forests. *Machine Learning*. 2001 Oct 1;45(1):5-32.

Chen Y, Jin GZ. Does health insurance coverage lead to better health and educational outcomes? Evidence from rural China. *Journal of Health Economics*. 2012 Jan 1;31(1):1-4.

Cheng L, Liu H, Zhang Y, Shen K, Zeng Y. The impact of health insurance on health outcomes and spending of the elderly: evidence from China's new cooperative medical scheme. *Health Economics*. 2015 Jun;24(6):672-91.

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. T*he Econometrics Journal*, Volume 21, Issue 1, 1 February 2018a, Pages C1–C68.

Chernozhukov, V., M. Demirer, E. Duflo and I. Fernandez-Val Generic machine learning inference on heterogeneous treatment effects in randomized experiments, *National Bureau of Economic Research*. 2018b.

Chou SY, Grossman M, Liu JT. The impact of national health insurance on birth outcomes: a natural experiment in Taiwan. *Journal of Development Economics*. 2014 Nov 1;111:75-91.

Currie J, Gruber J. Saving babies: The efficacy and cost of recent changes in the Medicaid eligibility of pregnant women. *Journal of Political Economy*. 1996 Dec 1;104(6):1263-96.

Dow WH, Schmeer KK. Health insurance and child mortality in Costa Rica. *Social Science & Medicine*. 2003 Sep 1;57(6):975-86.

Erlangga D, Ali S, Bloor K. The impact of public health insurance on healthcare utilisation in Indonesia: evidence from panel data. *International Journal of Public Health*. 2019b May 1;64(4):603-13

Erlangga D, Suhrcke M, Ali S, Bloor K. The impact of public health insurance on health care utilisation, financial protection and health status in low-and middle-income countries: A systematic review. *PloS One*. 2019a Aug 28;14(8):e0219731.

Fink G, Robyn PJ, Sié A, Sauerborn R. Does health insurance improve health?: Evidence from a randomized community-based insurance rollout in rural Burkina Faso. *Journal of Health Economics*. 2013 Dec 1;32(6):1043-56.

Hainmueller J, Mummolo J, Xu Y. How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis*. 2019 Apr;27(2):163-92.

Harimurti P, Pambudi E, Pigazzini A, Tandon, A. 2013. The Nuts & Bolts of Jamkesmas, Indonesia's Government-Financed Health Coverage Program for the Poor and Near-Poor. UNICO Studies Series;No. 8. World Bank, Washington DC. World Bank. https://openknowledge.worldbank.org/handle/10986/13305 License: CC BY 3.0 IGO.

Hidayat B, Thabrany H, Dong H, Sauerborn R. The effects of mandatory health insurance on equity in access to outpatient care in Indonesia. *Health Policy and Planning*. 2004 Sep 1;19(5):322-35.

Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis.* 2007;15(3):199-236.

Johar M. The impact of the Indonesian health card program: a matching estimator approach. *Journal of Health Economics*. 2009 Jan 1;28(1):35-53.

Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*. 2007;22(4):523-39.

Lagomarsino G, Garabrant A, Adyas A, Muga R, Otoo N. Moving towards universal health coverage: health insurance reforms in nine developing countries in Africa and Asia. *The Lancet*. 2012 Sep 8;380(9845):933-43.

Lawn JE, Cousens S, Zupan J, Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: when? Where? Why? *The Lancet*. 2005 Mar 5;365(9462):891-900.

Mackenbach JP. Tackling inequalities in health: the need for building a systematic evidence base. *Journal of Epidemiology & Community Health*. 2003 Mar 1;57(3):162.

Mahendradhata Y, Trisnantoro L, Listyadewi S, Soewondo P, Marthias T. et al. The Republic of Indonesia Health System Review. *Health Systems in Transition*, Vol-7 No.1. WHO Regional Office for South-East Asia. 2017. https://apps.who.int/iris/handle/10665/254716

Mensah J, Oppong JR, Schmidt CM. Ghana's National Health Insurance Scheme in the context of the health MDGs: An empirical evaluation using propensity score matching. *Health Economics*. 2010 Sep;19(S1):95-106.

Moreno-Serra R, Smith PC. Does progress towards universal health coverage improve population health? *The Lancet*. 2012 Sep 8;380(9845):917-23.

Moreno-Serra R, Smith PC. Broader health coverage is good for the nation's health: evidence from country level panel data. *Journal of the Royal Statistical Society. Series A,(Statistics in Society*). 2015 Jan;178(1):101.

Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. *arXiv preprint* arXiv:1712.04912. 2017 Dec 13.

O'donnell O, Van Doorslaer E, Wagstaff A, Lindelow M. Analyzing health equity using household survey data: a guide to techniques and their implementation. *The World Bank*; 2007 Oct 27.

Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V. Damned if you do, damned if you don't: subgroup analysis and equity. *J Epidemiol Community Health*. 2012 Jan 1;66(1):95-8.

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. 1994 Sep 1;89(427):846-66.

Robinson PM. Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*. 1988 Jul 1:931-54.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983 Apr 1;70(1):41-55.

Sachs JD. From millennium development goals to sustainable development goals. *The Lancet*. 2012 Jun 9;379(9832):2206-11.

Shrestha R. The village midwife program and infant mortality in Indonesia. *Bulletin of Indonesian Economic Studies*. 2010 Aug 1;46(2):193-211.

Sosa-Rubi SG, Galárraga O, Harris JE. Heterogeneous impact of the “Seguro Popular” program on the utilization of obstetrical services in Mexico, 2001–2006: A multinomial probit model with a discrete endogenous variable. *Journal of Health Economics.* 2009 Jan 1;28(1):20-34.

Sparrow R, Budiyati S, Yumna A, Warda N, Suryahadi A, Bedi AS. Sub-national health care financing reforms in Indonesia. *Health Policy and Planning*. 2017 Feb 1;32(1):91-101.

Sparrow R, Suryahadi A, Widyanti W. Social health insurance for the poor: Targeting and impact of Indonesia's Askeskin programme. *Social Science & Medicine*. 2013 Nov 1;96:264-71.

Strauss, J., F. Witoelar and B. Sikoki. The Fifth Wave of the Indonesia Family Life Survey (IFLS5) Overview and Field Report. 2016.

Strauss, J., F. Witoelar, B. Sikoki and A. Wattie The Fourth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report. 2009.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar. The Third Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report." (WR-144/1-NIA/NICHD). 2004.

Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Statistical Methods in Medical Research*. 2012 Feb;21(1):55-75.

Thabrany H. Private health sector in Indonesia: opportunities and progress. *J. Indonesian Med. Assoc*. 2001;5:1-3.

Tibshirani, J., S. Athey, S. Wager, R. Friedberg, L. Miner and M. Wright. grf: Generalized Random Forests (Beta). R package version 1.1.0.2018.

Trujillo AJ, Vecino Ortiz AI, Ruiz Gómez F, Steinhardt LC. Health insurance doesn’t seem to discourage prevention among diabetes patients in Colombia. *Health Affairs*. 2010 Dec 1;29(12):2180-8.

VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009 Nov 1;20(6):880-3.

Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*. 2018 Jul 3;113(523):1228-42.

Wagstaff A. Estimating health insurance impacts under unobserved heterogeneity: the case of Vietnam's health care fund for the poor. *Health Economics*. 2010 Feb;19(2):189-208.

Wang H, Yip W, Zhang L, Hsiao WC. The impact of rural mutual health care on health status: evaluation of a social experiment in rural China. *Health Economics*. 2009 Jul;18(S2):S65-82.

World health statistics 2019: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.

Yilma Z, Mebratie A, Sparrow R, Dekker M, Alemu G, Bedi AS. Impact of Ethiopia's community based health insurance on household economic welfare. *The World Bank Economic Review*. 2015 Jan 1;29(suppl_1):S164

Appendix A: Variables used in the principal component analysis to construct the wealth index

- Whether the household has electricity,
- Access to piped water,
- Types of stove,
- Toilet inside the house,
- Refrigerator,
- Television,
- House and land owned by household,
- Ownership of other house,
- Vehicles,
- Household appliances,
- Savings,
- Receivables,
- Jewellery.

Appendix B: Tables

Appendix Table 1: Regressions results for pooled analysis of health insurance variable

	Unadjusted 1	Unadjusted 2	OLS 1	OLS 2	Unadjusted 1	Unadjusted 2	OLS 1	OLS 2
Infant mortality								
Subsidised HI	-0.0025	-0.0024	-0.0058	-0.0058				
	(0.0047)	(0.0051)	(0.0052)	(0.0052)				
Contributory HI	-0.0126***	-0.0128***	-0.0084**	-0.0087**				
	(0.0034)	(0.0035)	(0.0038)	(0.0038)				
Any insurance					-0.0075**	-0.0078**	-0.0071**	-0.0072**
					(0.0032)	(0.0033)	(0.0034)	(0.0034)
Observations	12,076	12,076	12,076	12,076	12,076	12,076	12,076	12,076
Skilled birth attendance								
Subsidised HI	0.0291***	-0.0229**	0.0209**	0.0227**				
	(0.0104)	(0.0109)	(0.0106)	(0.0106)				
Contributory HI	0.1279***	0.1224***	0.0272***	0.0329***				
	(0.0079)	(0.0080)	(0.0079)	(0.0080)				
Any insurance					0.0767***	0.0518***	0.0241***	0.0277***
					(0.0074)	(0.0076)	(0.0071)	(0.0071)
Observations	11,202	11,202	11,202	11,202	11,202	11,202	11,202	11,202
Year dummies	N	Y	Y	Y	N	Y	Y	Y
Covariates	N	N	Y	Y	N	N	Y	Y
Province dummies	N	N	N	Y	N	N	N	Y

Notes: HI: Health Insurance, OLS: Ordinary least squares, *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2: Comparison of observed characteristics of mothers with low and high treatment effects

	Subsidised health insurance			Contributory health insurance		
	Below median CATE Mean (SE)	**Above median CATE** Mean (SE)	**SMD[1]**	**Below median CATE** Mean (SE)	**Above median CATE** Mean (SE)	**SMD**
Age at birth <=23	0.34 (0.47)	0.26 (0.44)	0.167	0.22 (0.41)	0.35 (0.48)	0.299
Age at birth 23- <=27	0.29 (0.45)	0.19 (0.39)	0.246	0.30 (0.46)	0.20 (0.40)	0.245
Age at birth 27- <=31	0.23 (0.42)	0.20 (0.40)	0.067	0.23 (0.42)	0.21 (0.41)	0.063
Age at birth >31	0.14 (0.35)	0.35 (0.48)	0.501	0.25 (0.43)	0.24 (0.43)	0.004
1st wealth quintile	0.20 (0.40)	0.22 (0.41)	0.049	0.08 (0.27)	0.26 (0.44)	0.507
2nd wealth quintile	0.20 (0.40)	0.23 (0.42)	0.07	0.11 (0.31)	0.28 (0.45)	0.451
3nd wealth quintile	0.23 (0.42)	0.22 (0.42)	0.009	0.17 (0.37)	0.27 (0.44)	0.252
4th wealth quintile	0.20 (0.40)	0.19 (0.39)	0.014	0.33 (0.47)	0.10 (0.31)	0.561
5th wealth quintile	0.17 (0.38)	0.13 (0.34)	0.11	0.32 (0.47)	0.08 (0.27)	0.632
No/primary education	0.28 (0.45)	0.38 (0.49)	0.216	0.09 (0.29)	0.48 (0.50)	0.964
Secondary education	0.27 (0.44)	0.25 (0.44)	0.026	0.11 (0.31)	0.37 (0.48)	0.645
Senior education	0.36 (0.48)	0.29 (0.45)	0.145	0.55 (0.50)	0.13 (0.34)	0.992
Higher education	0.10 (0.30)	0.08 (0.27)	0.076	0.25 (0.43)	0.01 (0.12)	0.739
Poor card	0.13 (0.33)	0.08 (0.28)	0.141	0.08 (0.27)	0.09 (0.28)	0.033
Received cash transfer	0.22 (0.42)	0.31 (0.46)	0.205	0.12 (0.33)	0.31 (0.46)	0.46
Received subsidised rice	0.49 (0.50)	0.57 (0.49)	0.157	0.34 (0.47)	0.60 (0.49)	0.535
Writes in Indonesian	0.97 (0.17)	0.94 (0.24)	0.153	0.99 (0.11)	0.93 (0.25)	0.282
Public health clinic in community	0.99 (0.09)	0.99 (0.09)	0.009	0.99 (0.08)	0.99 (0.10)	0.049
Hospital in community	0.91 (0.29)	0.90 (0.31)	0.046	0.89 (0.31)	0.90 (0.29)	0.053
Private practice in community	0.96 (0.21)	0.96 (0.20)	0.021	0.94 (0.23)	0.97 (0.17)	0.121

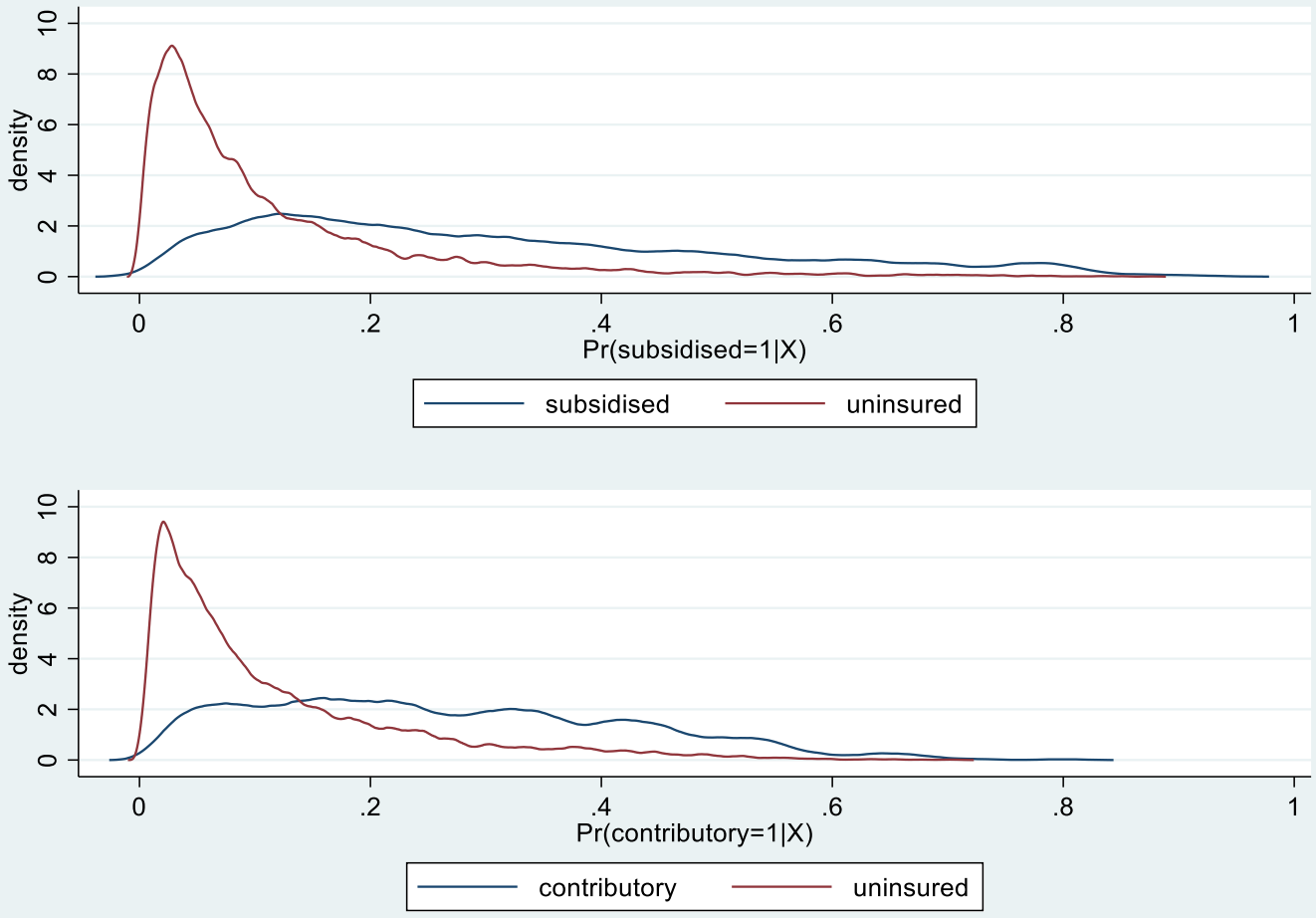
Notes: SE: standard error, SMD: standardised mean differences

Appendix Table 3: Tuning parameters in the causal forest analysis

Tuning parameter	grf package argument in causal_forest() function	Values (subsidised HI analysis)	Values (contributory HI analysis)
Fraction of the data used to build each tree	Sample. fraction	0.472	0.500
Number of variables tried for each split	mtry	21	21
Minimum number of observations in each tree leaf	min.node.size	1	5
The fraction of data used for determining splits	honesty.fraction	0.620	.0500
Prunes the estimation sample tree such that no leaves are empty	honesty.prune.leaves	TRUE	TRUE
Maximum imbalance of a split	alpha	0.091	0.05
Controls how harshly imbalanced splits are penalised	Imbalance.penanty	0.061	0

Appendix C: Figures

Appendix Figure 1: Overview of balance after propensity score weighting: standardised mean differences of covariates involved in the propensity score analysis



Appendix Figure 2: Estimated propensity score distributions

Notes: The top panel shows the estimated probability of being insured among those who were in fact uninsured, while the bottom panel shows the corresponding probabilities for those who were insured. The right (top and bottom) panels show the corresponding estimates for the probability of contributory insurance.