The Impact of the Indonesian Health Card Program: A Matching Estimator Approach

Meliyanni Johar

University of New South Wales, Australia

Australian School of Business Level 4, 417J, Sydney NSW 2052, Australia

Email: m.johar@unsw.edu.au

Tel: + 61 2 9385 6974

Abstract

This study evaluates the effectiveness of a pro-poor nation-wide health card program, which provides free basic health care at public health facilities in Indonesia. To quantify the effect of the program, it departs from the traditional regression-based approach in the literature. It employs propensity score matching to reduce the selection bias due to non-random health card distribution. The setting of the program and the richness of the data set support this strategy in providing accurate estimates of the program's effect on its recipients. The results indicate that, in general, the health card program only has limited impact on the consumption of primary health care by its recipients. This finding suggests the presence of other factors counteracting the generous demand incentive.

Keywords: Impact evaluation, health sector reform, Indonesia

JEL classification code: I1

Acknowledgements: The author wishes to thank Denzil Fiebig, Denise Doiron, Hajime Katayama, and two anonymous referees and the editor for their helpful comments and suggestions. This paper was presented at the Australian Health Economics Society Conference 27-28 September 2007 in Brisbane, Australia.

I. INTRODUCTION

Unequal access to health care has become prominent in the policy agenda of many countries worldwide. Interventions of various forms have been introduced by governments and non-government institutions in an attempt to minimise these asymmetrical situations, with targeted programs for the poor being the common theme. Some countries provide health goods directly (e.g., the United Kingdom), while others combine public provision with subsidised health insurance for the poor (e.g., Malaysia, Singapore). The common justification for subsidisation is that health care services are particularly costly for the poor, while they are more likely to face adverse health shocks (Wagstaff, 2005; Whitehead et al., 2001; Xu et al, 2003; Case et al., 2002). It is also believed that there are positive externalities from a healthy population.

In Indonesia, health care payments are largely out-of-pocket at the time of purchase or service provision, tending to be in cash. The health insurance market is underdeveloped, with less than 20% of the population covered by at least one form of health insurance (IFLS, 2000). As a result, many sick individuals with subsistence income and no insurance will not be able to obtain the necessary medical treatment, as more than half of the population lives on less than $2 per day (World Bank, 2002). Health care utilisation rates have been low and are unchanged, despite marked increases in the incidences of both communicable and non-communicable diseases. Restricted access to adequate health care has also been linked to critical health statistics, such as high under-five mortality (38 per 1,000 live births due to preventable factors) and the highest maternal mortality rate among Southeast Asian nations (WHO, 2006).

The health card program of 1994 is one of the government's major efforts to improve the nation's health conditions by promoting equality in access to primary health care. [1] It is a nation-wide project involving all public health facilities, including hospitals around the country. The program targets poor households and provides a full price subsidy to

[1] In 2000, the health card program was reformed to be contained within a social system (JPKM) managed by a state-owned insurance company. Under the reformed system, fund-holding entities were set up at district level to replace the central government in managing resource allocation and service provision in their areas. Nonetheless, the design of the health card program was unchanged.

medical expenses at public health facilities for all members of the household. Covered services include inpatient and outpatient care, diagnostic testing, contraceptive treatments, and children and maternal care. Health cards are distributed by the village heads based on a list of criteria that reflects household's welfare. Substantial regional heterogeneity makes a common rule for eligibility implausible, so it is decentralised and varies across communities. By design, there is no limit on the frequency of health card use, and households are to be reassessed annually.

This study measures the effectiveness of the program. To my knowledge, this is the program's first formal evaluation at the microeconomic level. So far, assessments are based on aggregate statistics on the number of health cards disseminated in one area and the recorded number of patients with health cards received by public health facilities (Widianto, 2007). Analyses at the aggregate level, however, are likely to be contaminated by other factors. Accurate evaluation of the program is necessary, particularly in the face of constrained resources to ensure that it is not wasteful. The outcome of this study will therefore provide valuable feedback to policymakers and suggest appropriate directions for future policy.

The study also contributes to the general health and development literature in at least three ways. First, it may extend our knowledge of the interaction between demand incentives, in the form of the health card program, and health care choice. We know very little about such interactions in developing countries. Most of the existing evidence in the literature relate to experiences in developed countries, which have very different environmental settings. In developing countries, for instance, transport and time costs associated with subsidised care are often non-negligible, and traditional medications are prevalent. Public health facilities are also often inadequate and rated poorly in quality by potential patients (see Filmer et al., 2000 for cross country review). In the presence of these costs and readily accessible alternatives, it is not clear whether a demand inducement will translate to increased consumption. In the literature, there are numerous studies on targeted programs in relatively poor countries, especially African countries, finding mixed conclusions about households' responses (Newman et al., 2002; Castro-

Leal et al., 2000; Pradhan et al., 1998). However, evidence from middle-income developing countries like Indonesia is relatively scarce (Gertler and Molyneaux, 2000; Pradhan et al., 2007).

Second, the study departs from a typical health study in its estimation technique. It is well recognised that the central issue in program evaluation studies is selection bias. Traditionally, instrumental variable (IV) regression is used to correct for sample selection bias arising from unobserved heterogeneity. However, selection on observables may be more appropriate in the case of the health card program, which relies on the village heads to select recipient households. In contrast to the IV technique, propensity score matching (PSM) supposes that households are different in observable ways. Moreover, in the presence of longitudinal data, PSM can be combined with the difference-in-differences (DID) technique to make a powerful estimator, PSM-DID, that is robust to inherent unobservables. PSM-DID has gained popularity in labour economics in producing reliable causal inference (Smith and Todd, 2000, 2005; Heckman et al., 1997), but it has not been widely used in the health field. Studies have also argued that PSM-DID provides estimates of program effects under more reasonable conditions than IV regression (Smith and Todd, 2005; Heckman et al., 1997, 1998; Augurzky and Schmidt, 2001).

Third, this is a demand-side study that manages to incorporate supply-side variations into the analysis. Supply variables are commonly excluded due to lack of data, yet it is known from the literature that supply influences demand either directly through additional arrangements or indirectly through increasing health awareness among patients. Data have supported this hypothesis by finding a positive correlation between health care utilisation rates and physician density (McGuire, 2000; Dow et al., 2000). Therefore, analyses without supply side factors may be invalid due to omitted variable bias.

The data are derived from the Indonesian Family Life Survey (IFLS) in 1993 (IFLS1), 1997 (IFLS2) and 2000 (IFLS3) by RAND. The IFLS is the only large sample longitudinal study on Indonesian households. It is a very rich data source collected at the household and individual levels, supporting the empirical strategy of this paper. Most studies in various literatures on Indonesia use the annual national household survey (*Susenas*) by the Central Bureau of Statistics, an ongoing large sample survey, but repeated at cross-sections. As an alternative data source, the IFLS is also nationally representative, using the same sampling frame as that used in *Susenas*. Moreover, it contains community and facility surveys, which provide information about health care providers in communities where households live. Details about the data set are documented in Thomas et al. (2001).

The results show that the health card program has only limited impact on health care consumption. The presence of a health card may allow the younger household members to receive curative treatments and females in the households to enroll in a contraception program, but it has no significant positive effect on the other household members. Considering that the program is fairly generous in its design, this finding may suggest that either a price subsidy to general health care is an unsuitable form of intervention to increase health care utilisation or certain factors counteract the demand inducement, or both. The policy implications are that program redesign or redirection of resources may yield a larger impact.

II. METHODOLOGY

In this study, the terminologies "treatment" and "treated" are not used to describe the health card program and households that received the concession, respectively, as they are typically used in program evaluation studies. This distinction is made to avoid confusion among households with health cards that receive medical treatment and those that do not. Thus, PSM-DID is used to calculate the average exposure effect on the recipient units (AER). The exposure in study is the availability of a health card in the household, and the recipient units are the members of households with health card. The following paragraphs outline the assumptions required by the method.

Let t_0 indicate the period before the health card program was introduced (pre-exposure) and t_1 indicate the post-exposure period. The first assumption behind PSM-DID is conditional independence:

$$Y^c_{t_1} - Y^c_{t_0} \perp d \mid P(X) \quad (1)$$

where Y^c is health care consumption by members of households without health card (control units), d is an indicator variable for whether or not the household has a health card and X is a vector of strictly exogenous variables that are unaffected by health card availability or anticipation of receiving health card (Heckman et al., 1998). This condition states that health card distribution is ignorable with respect to the outcome variables if all factors that are influencing the allocation and are related to health care utilisations are included in X. As X may be large in dimension, matching is performed on a single dimension or propensity score $P(X)$, which summarises information given by X (Rosenbaum, 2002). The second requirement is the presence of a common support. Unlike in regression-based methods, matching explicitly tests for the fulfilment of this condition, thereby avoiding "off-support" inferences. By off-support we mean the attempt to establish the exposure effect when recipient and control units in the data are actually incomparable, leading to biased AER estimates (Heckman et al., 1997; Cobb-Clark and Crossley, 2003).

To find a matching-pair for each recipient unit, the method searches for a unit in the control group with the closest propensity score to the recipient unit. Recipient units where we are unable to find a match lie outside the region of the common support and are thus excluded in the calculation of AER. Given the common support restriction, the AER is calculated as follows:

$$AET = \sum_{i \in D} \left[(Y_{it_1} - Y_{it_0}) - \sum_{j \in C} W_{ij} (Y_{jt_1} - Y_{jit_0}) \right] N_D^{-1}. \quad (2)$$

D and C are the pool of recipient and control units, respectively. W_{ij} is the weight placed on a control unit j for a recipient unit i, and N_D is the size of the recipient sample. The weighting function is user-specified, making it more amenable to heterogenous program effects. Kernel weights (Gaussian) are chosen using all control units within the region of common support, with the weights reflecting their relative closeness to the recipient unit-to-match (Dehejia and Wahba, 2002; Heckman et al., 1997; Galiani et al., 2005). The propensity scores are estimated using a logit model. The standard errors of the AER estimates are given by bootstrapping with 200 replications.

Bootstrapped standard errors are used because the analytical standard errors are not available and bootstrapping is not inconsistent with kernel matching (Abadie and Imbens, 2006). It also accounts for both sampling errors in the propensity score estimates and errors due to multiple matches for a single recipient unit.

Equation (1) suggests that PSM-DID generates an AER that is robust to both heterogeneity in observables and unobservables, as long as they are time-invariant. First-differencing eliminates the effects of households' innate taste or preferences on both exposure status and health care choice that are difficult to deal with in the absence of a longitudinal data (Smith and Todd, 2000). Meanwhile, double-differencing will eliminate aggregate effects such as macroeconomic effects and differences in survey questionnaires.

In the health literature, several studies have attempted to take advantage of the matching approach, but very few of them are convincing. To name the three most recent studies for developing countries: Galiani et al. (2005) estimates the effect of the privatisation of water system on children's mortality in Argentina; Trujillo et al. (2005) analyse the impact of subsidised health insurance in Columbia; and Wagstaff and Yu (2007) examine the performance of a World Bank investment in China. All of these studies reported some positive effects of the studied program. The China project, for instance, reduces households' out-of-pocket health expenses, particularly on drugs. However, the study lacks common support and many observations have to be dropped to make the counterfactual and the program beneficiary samples comparable. As a result, the AER is estimated by tolerating observations that are just off-support.

III. DATA AND VARIABLES

The data used in this study are derived from the Indonesian Family Life Survey (IFLS) in 1993 (IFLS1), 1997 (IFLS2) and 2000 (IFLS3) collected by RAND in collaboration with several local universities. The IFLS is a nationally representative longitudinal study covering 13 out of 27 Indonesian provinces where 83% of the population resides. In its first wave, the study sampled 7,224 households consisting of 22,000 individuals. The subsequent waves consist of recontacted IFLS1 households (origin households) and their split-offs, which are new households consisting of (target) member(s) of the origin household. Recontact rates are considerably high (95%), and in the latest wave, the study has involved 10,435 households and over 37,000 individuals. Nevertheless, as non-random attrition is a common concern in any longitudinal study, the IFLS provided sampling weights that reflect the initial probability that a household is sampled and its appearance in the subsequent wave(s). These sampling weights are used in estimation. The reliability of the IFLS data set had been formally documented in Thomas et al. (2001).

As the health card program began in 1994, there are two potential post-exposure periods: 1997 and 2000. However, because there are only a small number of households with health card in 1997 (2% of households in IFLS2), I define 1997 as another control period together with 1993, excluding the few households with health cards in 1997 from the sample, and let 2000 be the only post-exposure period. As households may appear in only 2 out of the 3 survey waves, 2 different samples are created. Sample restriction to households that appeared in all 3 waves will be inefficient, as it excludes almost 4,000 households from the analysis. In addition, maintaining 2 different samples provides a natural robustness test against model specification and sample specificity. Define Sample 1 as a sample consisting of households appearing in IFLS1 and IFLS3 and Sample 2 as a sample consisting of households with information in IFLS2 and IFLS3. The overlap proportion between the two samples is about 60%.

The first part of the empirical strategy computes the propensity score that a household has a health card. The main advantage of having a panel data set is that information on pre-exposure conditions is available and can be used to calculate the probabilities. A valid propensity score requires the covariates to be strictly exogenous. Lagged variables may be a way of achieving this (Caliendo and Kopeinig, 2008). For Sample 1, this means that the matching equation is a function of conditions in 1993, while for Sample 2 it is a function of conditions in 1997. Split-off households are excluded, as their present characteristics may be very different than they were when they were members of the original households (about 11% of all households in IFLS2 and IFLS3). The final samples with complete information consist of 5,262 households in Sample 1 and 4,580 households in Sample 2.

Given the propensity scores, the AER is calculated for only adult individuals ($\geq$ 15 years old) in the households. Children's health would be tricky to interpret: as they get older, the intensity of regular check-ups naturally falls. Members of a household would have identical propensity scores as they have the same exposure (health card) status. To account for individual differences in the health care utilisation pattern, household members are categorised according to their relation to the household head in the post-exposure period: the household heads, the spouse of the household heads, children of the household head and a pool of other household members (parents, in-laws, step children, grandparents, nieces/spouses, uncles/aunts and servants). This way of dividing the data is chosen because household hierarchy may convey extra information about intra-household information sharing and resource allocation, as there is only one card per household. Furthermore, it may be a convenient way to group individuals based on common characteristics. For example, the spouse of the household head tends to be a female, aged between 30–50 years old and does not tend to be participants in the formal (taxed) labour market. Throughout the analysis, it is important to keep in mind that the sample of children does not imply young individuals. Indeed, the average age of the children sub-sample is close to 22 years in the post-exposure year. The sub-sample of children is also much smaller in Sample 1, as young adults in 2000 are small children in 1993.

Dependent variables

Information about health card availability is obtained from the household heads. About 20% of households in each sample are recipient households. To measure health care utilisation, I consider the number of inpatient and outpatient visits received by individuals at public and private facilities. Public facilities include health centers and their subsidiaries, hospitals and village midwives. Private facilities are private hospitals and physicians. Traditional practitioners (e.g., religious healers) are excluded. For inpatient care, the reference period is 12 months to the survey, and for outpatient care, this period is 4 weeks. Outpatient care is further categorised according to the purpose: curative or preventative. From the list of purposes in the survey, visits for “treatment” and “medication” only are classified as curative-type services, whilst other purposes such as medical check-ups, vaccination and an assortment of other treatments are classified as preventative-type care. The dependent variables are the first-difference of these health care consumption measures.

In the samples, health care consumption levels are low. This picture has been recognised and is consistent with the *Susenas* data set. The important revelation is that the low utilisation rates persist to the post-exposure period amid the health card program. About 80% of members of recipient households are non-users (not consuming health care in any period) of public health care for outpatient services, and 93% are non-users of private health care. Meanwhile, inpatient cases are very rare for any households (1–3%), which may be explained by high opportunity costs of inpatient days due to the absence of compensation payments for many workers. The fact that service costs are different at public and private facilities is also reflected in the data with users of public (private) facilities tending to always use public (private) facilities and members of control households tending to consume more private services, which are typically relatively expensive. Only 2 – 3% of users of health care consumed more private care in the post-exposure year, while reducing utilisation of public health facilities. The sample correlation between utilisation of public and private facilities is about 0.03 in the pre-exposure period and essentially zero in the post-exposure year.

Tables 1A and 1B present health care consumption patterns in the post-exposure year by household hierarchy for Samples 1 and 2, respectively. These figures are data that a typical cross-sectional evaluation study would analyse. The absolute *t*-statistics for differences in two sample means are reported in parentheses. In general, this snap-shot picture suggests that the program has had some success in achieving its objective. When receiving outpatient care, members of recipient households paid more visits to public facilities than members of control households. Visits were mainly for curative services for children, while visits were for preventive services for spouses. As the initial utilisation levels were low, the increases are quite marked. With regard to inpatient care, the two samples provide different pictures. This discrepancy may be explained by the rarity of hospitalisation cases in each year. Hence, extra caution should be taken when dealing with the inpatient care variables.

Before doing the formal evaluation of whether the differences in health care utilisation are due to the health card program, Table 2A–B reports simple DID estimates from the two samples. Double-differencing eliminates almost all the previously observed differences in health care consumption between recipient and control individuals. Analyses that miss the pre-exposure information may therefore suggest misleading conclusions.

Explanatory variables

Conditions that explain both eligibility and health care demand are relevant covariates. Eligibility is determined by household's welfare conditions, while health care demand is influenced by demographics, socioeconomic status, health care supply and environmental conditions. Fortunately, the IFLS data set is sufficiently rich to capture these variables.

The first set of variables consists of measures of household's welfare. These include household compositions and conditions, value of non-business assets and ownership of at least one form of health insurance. In addition, I include household head characteristics, as these are observed by the village heads and hence influence health card eligibility. For example, unemployment of the household head is listed as one of the eligibility criteria.

The second set of variables measures the extent of the household's knowledge of health care facilities. Knowledge is important because it reflects the accessibility of a certain type of health facility and the consideration to receive medical treatment at this facility. Furthermore, it has been suggested that the main way social relationships influence health care demand is through their effect on health knowledge (Andersen, 1995). Social networks disseminate references and updates on new products and help locate the appropriate health care providers (Weerdt and Dercon, 2006). As such, households with strong social networks tend to be knowledgeable. This link suggests that we may reduce the omitted variable problem associated with social network – which is hard to quantify – by including health knowledge variables as covariates. In the IFLS, spouses were asked whether they know the whereabouts of public hospitals, private hospitals, health centres, private practices, nurses or midwives and traditional practitioners that the family attends. They respond to these questions on behalf of the entire household.

The final set of control variables deal with variations in the quantity and quality of health care providers. The availability of community-level data and facility surveys in the IFLS makes this study one of the few demand-side analyses that can jointly account for variations in the supply-side factors. In cross-sectional studies, the inclusion of supply variables as covariates may be implausible due to the simultaneous determination between demand and supply. However, with panel data, information on past supply conditions are available and reasonably assumed to be given. To measure quality, I consider the availability of full-time health workers (e.g., GPs, dentists and nurses) and the availability of services in a facility. Furthermore, both demand and supply of health care may be affected by exogenous health shocks to the community. To control for this, I create two indicator variables for minor and major health shocks in the last three years prior to the reference period. A minor (major) health shock refers to any health-related epidemics such as outbreak of diseases or a flood affecting less than (at least) 50% of the local population. In Sample 1, 31 out of 312 IFLS communities (10%) experienced at least one recent major shock, and 132 communities (42%) experienced at least one recent minor shock. In Sample 2, the corresponding numbers are 41 (13%) and 100 (32%) communities.

In addition to these covariates, a dummy variable for urban area and provincial dummy variables are included to control for regional heterogeneity. This is essential because Indonesia's population is very unevenly distributed. About 70% of the population lives on Java Island, which has a land area of only 7% of the country's total dry land. As a result, development stages vary substantially across regions (Lanjouw et al., 2001). Environmental conditions such as soil fertility and rainfall activity, as well as staple food (which affects health conditions), vary across regions. Finally, households sampling weights are used to take into account attrition in the survey.

Table 3 reports selected summary statistics by exposure status. All variables are characteristics as at a pre-exposure period. It can be seen that characteristics are similar in the two samples due to overlaps, but each sample provides a unique time interval to the post-exposure year. Sample 1 provides a longer interval of 7 years, while Sample 2 provides a 3-year gap. Recipient households tend to be headed by poorly educated heads, consist of more elderly members and fewer working-age adult members, have low values of asset and expenditure, and are less likely to have health insurance and live in relatively sub-standard quality houses with non-tile flooring and limited access to directly piped water for drinking and cooking. These households also tend to be home owners, which may indicate long-term residents. Hence, at least on average, the program appears to have successfully reached its target.

IV. RESULTS

Figures 1 and 2 summarise the quality of the matching in Samples 1 and 2, respectively. It can be seen that both samples feature considerable regions of overlap, especially for Sample 2, and there is no household in both samples with either 0 or 1 for a propensity score. This result is encouraging, considering that the health card program is a targeted program. The common support condition fails if households with certain characteristics either always or never receive health cards. Imposing the common support condition results in an exclusion of a small number of households (less than 3% in both samples), and both samples readily pass the balancing tests as described by Becker and Ichino (2002).

The Appendix reports results from the matching equation. Significant covariates suggest that recipient households are different from control households. In general, health cards are more likely to be distributed to households with inferior housing conditions, few assets and low income. Household composition variables and the characteristics of the household head have the expected signs but are not significant. Nonetheless, they are retained in the matching equation (Rubin and Thomas, 1996). With regard to knowledge variables, the likelihood of holding a health card increases with knowledge about the locations of public health centers and decreases with knowledge of private hospitals. This result may be explained by the tendency of private hospitals to be located in well-off neighbourhoods (with few poor households) where there is demand for them.

Meanwhile, the relationship between health card eligibility and supply variables is less clear. This makes sense. The inclusion of supply variables as covariates in the matching equation is justified on the basis that they influence health care utilisation, rather than health card eligibility. It is unreasonable to expect that changes in supply variables directly affect a household's welfare state, which, in turn, determines its eligibility for a health card. The quantity of public health facilities in the community, for instance, is governed by the primary health system, which promises at least one form of public provider within a defined distance. Meanwhile, many of the quality measures have a negative sign, suggesting that recipient households tend to reside in areas with inferior health services. However, only a few of them are statistically significant.

Compared to all samples, control households in the matched samples have relatively low income, more knowledge about public facilities and traditional healers and less knowledge about private facilities, indicating increased compatibility with recipient households. Unlike in the case of covariate-matching, PSM does not generally yield identical means for all covariates for the matched pairs because the propensity score is a summary measure. Instead, it places a larger emphasis on balancing covariates that are the key predictors of the exposure status, as found in the matching regression.

Table 4A–B reports the AER estimates for different household members in the two samples. These are the main results. Differences between estimates in these tables compared with those in Table 2A–B reflect the confounding effects of observed covariates on health care consumption patterns. In general, the conditional estimates in Table 4A–B are smaller in absolute magnitude than the unconditional estimates in Table 2A–B, suggesting that the overall effect of observables on the exposure status is positive.

This direction is consistent with the fact that village heads disseminate health cards based on a household's observed characteristics.

For outpatient care, in Sample 1, none of the AERs are statistically significant at conventional levels. For household heads and spouses, the direction of the exposure effects for all types of care at public facilities however is positive. For children and other household members, some of the estimates have unexpected signs, but they may be unreliable due to small sample sizes. Using a Monte Carlo experiment, Zhao (2004) , finds that the PSM method does not perform well in small samples compared to other methods of matching, as the variance of the estimated program effect is too large.

In Sample 2, the results persistently show that health card availability has no significant effect on heads' health care consumption. The AER for every type of health care in the head sub-sample is very small. On the other hand, having a health card significantly lowers spouses' visits to private providers to receive preventative-type treatments. Further investigation reveals that this result is largely driven by the spouses' demand for contraception at public facilities. I will explore the effect of health cards on acquiring contraception next. On the other hand, spouses' visits to public providers for curative-type treatments are higher after holding a health card. This result is statistically significant for a positive alternative.

Health card availability especially benefited the children of the household head. In particular, it allows the children to pay more frequent visits to public health facilities to cure illness and receive medications. Relative to the low initial consumption level, the availability of a health card increases children's health care consumption by more than 80%. Their consumption of preventative-type treatments at public facilities also increases but is not statistically significant. Meanwhile, health card availability is associated with fewer visits to private providers, but to a smaller magnitude than the overall increase in consumption of public health care. It is not statistically significant. For other household members, consistent with the results from Sample 1, having a health card has no meaningful effect on health care consumption.

Samples 1 and 2 provide consistent results with regard to inpatient care. In no case does health card availability increase consumption of inpatient care by members of recipient households. This is not unexpected, given that inpatient treatment is a rarity for Indonesian households.

Contraceptive enrolment

The effect of having a health card on fertility decisions is also important, as the health card can cover both maternal care and contraceptive treatments. The former encourages a larger family size, while the latter delays or prevents pregnancy. However, there are reasons to believe that the effect of having a health card on encouraging a larger family is small. These include uncertainty in the continuity of health card availability and the influence of the ongoing government's family planning campaign encouraging just 2 children per family. To investigate this matter, I consider contraceptive take-up during the study period and future plans to use contraceptive device for females who do not use contraception in the post-exposure year. In the IFLS, contraceptive booklets are forwarded to all females aged 15–54 year olds (eligible females). For the majority of households, there is only one eligible female in the household: the spouse of the household head. There are about 3,799 eligible females in Sample 1 and 3,518 females in Sample 2. Table 5 reports the AER estimates for eligible females in the two samples. 20% of them are members of recipient households.

Figures under the (D) columns reveal the average proportions of new contraceptive enrolments by eligible females in recipient households within 7 and 3 years' time, respectively. Figures under the (C) columns are similarly defined for eligible females in the control households. The mean for Sample 2 is lower due to enrolments prior to 1997.

The results are unsurprising given that the reinforcement to enroll is heightened by supply expansion in public facilities around 1997. The AER estimates suggest that eligible females take advantage of both health card coverage and the supply expansion to start using contraception. This result is consistent with Jensen (1996), which finds that Indonesian women's contraceptive behaviour is highly sensitive to the presence of subsidised facilities. Results from the RAND HIE have also suggested that preventative-type services are particularly price-sensitive because preventative-care is a luxury good as opposed to a normal good and is highly substitutable. However, it is noteworthy that enrollment does not guarantee continuity of use. The enrollees may fail to meet their next treatment if a health card is unavailable. Meanwhile, a health card has no effect in altering attitude towards contraception among eligible females who, for various reasons (e.g., religion), choose not to use contraception.

Robustness Check

This section concerns the sensitivity of the results to assumptions made as part of the analysis and explores several problems that may lead to biased results.

First, program effects may be heterogeneous, and classification by household hierarchy does not sufficiently capture individual heterogeneity. For example, the age range for household heads is 80 years. Other ways to slice the data are therefore considered, including by age and gender. The AERs are re-calculated, and overall the results by age groups reflect the previous results with positive effects found among the younger cohorts (who are likely to be in the children sample in the earlier division by household hierarchy), and negative effects on preventative care at private facilities for 30-49 year olds, the age range for most spouses. The AERs based on gender group are also calculated, and the results for males and females closely reflect the results for heads and spouses.

Second, the estimated AERs may be contaminated by the effect of relevant macroeconomic changes. So far, it is assumed that macroeconomic effects are homogenous across units and differenced away. Thus, all households without health cards are a potential comparison group. Nevertheless, it is often the case that macroeconomic movements affect different groups of the population differently. One way to investigate this possibility is to restrict the potential comparison group to households that share similar characteristics with the recipient households. Income level is commonly used to guide this division. However, in the face of substantial regional heterogeneity, this separation is complex (see Lanjouw et al., 2001; Booth, 1993 for discussion).

An alternative way may be to restrict the comparison group to households who reside in communities that issued health cards. Because health cards apply nationally, no restriction is made on the recipient households. In effect, this restriction eliminates exclusive neighborhoods consisting of only rich households. The resulting restricted sample contains about 70% of households in each sample. The propensity scores are recalculated, and households are re-matched. The matching results are summarised in Figures 3 and 4.

As with the unrestricted samples, there are considerable regions of overlap, but this time the upper bounds of the estimated propensity scores are closer to 1. There are also fewer matched households with propensity score less than 0.2 than there are in the unrestricted samples. All of these results confirm that the sample restriction produces a pool of more similar households. In general, the estimated health card effects are larger in these samples, as the means for the new control samples are smaller. This trend may reflect the larger effect of macroeconomic crisis on health care consumption pattern of the poor. However, the main thrust of the results maintains that the health card program only has limited effects on the health care consumption of its beneficiaries.

Third, knowledge of the whereabouts of health facilities may be correlated with health behaviour and health care utilisation, thereby violating the exogeneity assumption. To entertain this possibility, the matching equation is re-estimated without the knowledge variables, and the resultant distribution of the propensity scores is plotted in Figures 5 and 6. The scores from the two matching equations track one another quite closely, particularly in Sample 1. Comparable distributions of the propensity scores in turn imply that the resulting common support regions and weights for the control units are comparable. In terms of the results, the AER estimates change slightly in magnitude. Taking their statistical significance into account, the qualitative conclusion does not change. Only the AER estimate for children's outpatient care (for curative treatment) at public facilities in Sample 2 is significantly positive.

The last two tests concern the validity of the first assumption of PSM: that there is no selection on unobservables. Matching and balancing removes biases due to non-overlapping support and differences in the propensity score distributions of the recipient and control households. However, they do not remove bias resulting from unobservables. Recently, Becker and Caliendo (2007) suggest an indirect check for this assumption by considering how large the effect of the unobservables or "hidden bias" needs to be in order to reverse the results found by PSM. To illustrate the rationale behind the test, consider a binary outcome (which is a function of observed covariates x and unobservables u) for a matched pair *i* and *j*, and let P_i and P_j be the probability that each unit receives health card. The odds ratio that they hold a health card is:

$$\frac{P_i/(1-P_i)}{P_j/(1-P_j)} = \frac{P_i(1-P_j)}{P_j(1-P_i)} = \frac{\exp(\beta x_i + \gamma u_i)}{\exp(\beta x_j + \gamma u_j)}, \qquad (3)$$

which becomes $\exp[\gamma(u_i - u_j)]$ if a matched pair has comparable observable covariates. In other words, theoretically, the matched pairs differ only by a factor γ and their unobservables. PSM requires either $\gamma = 0$ or $(u_i - u_j) = 0$. An important result from Rosenbaum (2002) is that, supposing that the above hypothesis is false, then the odds ratio that one of the matched pair receives health card can be bounded by the following bounds:

$$\frac{1}{\exp(\gamma)} \le \frac{P_i(1-P_j)}{P_j(1-P_i)} \le \exp(\gamma). \qquad (4)$$

Under the null hypothesis, $\gamma = 0$, *i* and *j* have the same probability of receiving the concession.

In practice, the test is based on a non-parametric Mantel and Haenszel (MH, 1959) test statistic Q_{MH} (1959) for binary outcomes with the null hypothesis that, given random sampling, the outcome variables are not affected by exposure status. As demonstrated in

Aakvik (2001), the test involves comparing the number of recipient units who benefited from the program exposure and its expected number if the program has no effect:

$$Q_{MH} = \frac{|Y_1^D - \sum_{s=1}^{S} E(Y_{1s}^D)| - 0.5}{\sqrt{\sum_{s=1}^{S} \mathrm{var}(Y_{1s}^D)}}, \tag{5}$$

where Y_{1s}^D $(= N_{Ds} Y_{1s} / N_s)$ is the number of positive outcome for the recipient sample in stratum *s* of the sample, $N_{Ds} + N_{Cs} = N_s$, and $Y_{1s}^D + Y_{1s}^C = Y_{1s}$, where Y_{1s} is the total number of positive outcome in stratum *s*. Under the null hypothesis, there is no exposure effect, and a positive outcome is equally likely for recipient and control units. The strata reflect the estimated propensity score.

When the outcome variables are not binary in nature, as they are in this case, they may be transformed such that a successful outcome is coded as 1, and otherwise it is coded as 0 (Becker and Caliendo, 2007). The test may then be interpreted as a check on a simpler assessment of the program concerning only the presence of any intended effect. In this case, a successful outcome is an increase in health care consumption. Now, because the concession only applies at public facilities and the dependent variables are one-sided under the binary definition, we focus on utilisation of public facilities.

Let $\Gamma = \exp(\gamma)$. Rosenbaum (2002) shows that, for a given Γ and $u \in \{0,1\}$, the MH test statistic is bounded by two known distributions, which move apart from each other, reflecting increased uncertainty about the test statistics in the presence of hidden bias. The size of Γ would reflect the extent to which the matching results depart from the assumption held under the null hypothesis. The bounds are calculated using the routine of Becker and Caliendo (2007), and matched pairs are found by radius matching with radius 0.01. The test may be unsuitable for kernel matching, which uses the entire sample as the matching pair. For a given Γ, the upper bounds adjust the MH test statistics downwards when the AERs are overestimated, and the lower bounds adjust them upwards when the AERs are underestimated.

For the health card program, the direction of the hidden bias is not obvious. Due to the double-differencing involved in PSM-DID, we know that the hidden bias must come from time-varying unobserved heterogeneity. However, we must not confuse this source of bias with selection by observables, which we have explored so far and found to be positive in direction. They need not have the same pattern. Indeed, the true effect may be the reverse of the estimated effect if the hidden bias dominates in the opposite direction. Nevertheless, given that most AERs (at public facilities) have positive signs, which indicate success, underestimation is somewhat less concerning than overestimation. In other words, if individuals with low value of unobservables are overrepresented in the recipient samples, the true effects will be larger and more significant than estimated, and this is not undesirable. Hence, we shall focus on overestimation cases.

The tests suggest that, at small levels of Γ, health card availability no longer has a significant effect on spouses and children in recipient households (Sample 2). Under $\Gamma = 1$, significant exposure effects for spouses and children are found, but the positive results may be reversed if spouses and children in recipient households are allowed to differ by 10% or more in terms of unobserved characteristics with their counterpart in control households. In other cases where the exposure effects at $\Gamma = 1$ are insignificant, the conclusion that the health card program has no effect is robust to a hidden bias that would increase Γ to at least 1.3 (up to double the odds of being a recipient unit for other household member sub-samples). In short, these checks suggest that it is unlikely that the result of no positive effect is reversed.

The last test involves estimating a parametric model. PSM-DID is non-parametric and makes no distributional assumption about the error terms in the program exposure and the outcome equations. On the other hand, distributional assumptions about these error terms allow them to be correlated. A joint estimation of the two equations can identify the correlation coefficient. In this case, I assume a bivariate normal distribution and estimate probit-family models. The dependent variables are made binary, as in the Rosenbaum test, and health card availability is instrumented using the set of covariates used in the matching equation. The coefficients of the exposure status are reported in Table 6. The common support restriction is imposed to isolate the effect of hidden bias. It can be seen that, except for the children sample in Sample 1 (which is very small), the assumption of zero correlation between the error terms, controlling for the observable heterogeneity, is plausible. The inclusion of extensive covariates may have teased out the heterogeneity in unobserved factors. In sub-samples that are large, all of the average (local) exposure effects are statistically zero.

Discussion

The conclusion that the health card program has little or no effect may be counterintuitive at first. By design, the program offers a generous subsidy to those who were formerly restricted in access to formal health care. For instance, experiences from developed countries would predict excess utilisation by the program beneficiaries. Consequently, generous programs are typically equipped with mechanisms that discourage excessive usage or other rent-seeking behaviours (Manning et al., 1987; Million et al., 2003 provide contrasting evidence from developed countries). However, this result is not new for developing countries (Filmer et al., 2000), and there are several explanations that may rationalise the result.

First, Poterba (1994) argues that government intervention through a price subsidy is ill-suited when the price elasticity of demand for the subsidised good is low, or when there is large uncertainty and divergence in this elasticity across units. From Indonesia's own health experiment study (Indonesia Resource Mobilization Study), Gertler (1995 in Lanjouw et al., 2001) finds that the demand for health services in general at Indonesian public health facilities fit these conditions: it is inelastic and varies greatly with income. Furthermore, the standard economic demand-supply theory predicts a small change in quantity for a given price change when the demand schedule is inelastic.

Second, relatively stable health care consumption may be due to the fact that households are not selected exclusively on the basis of their health conditions. Healthy individuals require little or no health care services. However, the problem of serious underutilisation of health care, particularly by the poor, is well known. Poor individuals often do not consider themselves sick when others facing a similar condition do because they have suffered from the illness for a long time and thus have adapted to these adverse health conditions (Akin et al., 1998). As a price subsidy, the health card program does not make it compulsory for its holders to consume health services. Like any other households, those holding health cards can choose whether or not to get medical treatment, from which provider they will receive the treatment, and the level of treatment to be consumed.

A related issue to the features of the program is the variation in the information dispersed by the village heads when distributing the cards. Due to the decentralised nature of the program and the lack of official materials in at least the first few years, such as brochures explaining the program, households may receive different levels of information about the objective and the scope of the program. For instance, beneficiary households may understand that the health card can be used in the event of illness, even though they are uncertain about its use in other circumstances such as for preventative-type services and non-emergency procedures. There is a body of literature predicting that recipient households will not try to economise health card availability if they think such behaviour is abusing public sympathy (Miguel and Gugerty, 2005; Fehr and Gachter, 2000; Besley et al., 1993). The reason is the fear of social sanctions, such as isolation and shame. The credibility of this theory has been highly praised in developing countries where there is great physical proximity between neighbours and where households tend to be long-term and less mobile residents.

The third explanation is an inadequate public health system. There have been reports that the qualities of public health facilities are deteriorating and their drug supplies are inadequate (Frankenberg et al., 1999; Waters et al., 2003). Low quality may explain the underutilisation of public health facilities and households' decision to choose to self-medicate or seek alternative treatment and ignore the availability of a health card.
For policymakers, the finding of this study has several important implications. If the rationale behind the introduction of the program were to encourage health care utilisation, then the results presented here suggest that a price subsidy at the point of purchase, such as the health card, is not the best way to encourage Indonesian poor to increase utilisation. Alternative strategies complementing the price subsidy, such as educational

programs, may be needed. There are also suggestions that people are more responsive to incentives for specific health care services as opposed to a general coverage, or to programs that are targeted to individuals as opposed to households. The limited effect of the health card program in increasing utilisation highlights the presence of other non-negligible costs that counteract the demand inducement. The limited supply condition may be the primary deterrent. Naturally, the program could have had objectives other than increased utilisation. The availability of the health card protects households from (catastrophic) health expenses. One could argue that the introduction of the health card has been successful on this particular dimension.

As a final comment, the result highlights that experiences of developed countries might not be generalised to developing countries. For example, moral hazard behaviour associated with generous programs is a common concern in developed countries. However, it may not matter to the same extent in developing countries. The cost of risky behaviour or reduced care to individuals is very high due to the absence of compensation payments and insurance. In addition, policymakers in developing countries work in a less favourable environment, where resources and public knowledge may be very limited. Informal health care (treatments that may not have scientific accreditation), such as traditional healers and home-grown remedies are popular and highly accessible, and many households consider them to be acceptable substitutes for formal health care.

V. CONCLUSION

It has been suggested that poor households have poor health because they lack access to adequate health care. The intention of the health card program has been to protect the health status of the Indonesian poor by allowing them access to primary health care. The program is generous in providing full coverage for a wide range of primary health care services at public health facilities. The finding in this study however, shows that, in general, households do not exploit the presence of a health card to increase health care utilisation. One area where the program has seemingly encouraged increased utilisation is in contraceptive take-up by eligible females in the households. In this case, however, the demand reinforcement was paralleled with an expansion of family planning services in the public health facilities.

The limited effect of the health card program was possibly foreseeable. Changes in price would not result in big quantity changes in the demand schedule for general health care in public facilities if this demand is not price-sensitive to begin with. The planned demand expansion cannot be realised if the supply is limited. If the government wishes to encourage health care use, at least in the short-run, priority should be placed on expanding and improving the primary health system.

References:

Aakvik A. Bounding a matching estimator: the case of a Norwegian training program. Oxford Bulletin of Economics and Statistics 2001; 63(1); 115-142.

Abadie A, Imbens G W. On the failure of the bootstrap for matching estimators. NBER Technical Working Paper 2006 no.325.

Akin J S, Guilkey D K, Hutchinson P L, Mcintosh, M T. Price Elasticities of Demand for Curative Health Care with Control for Sample Selectivity on Endogenous Illness: An Analysis for Sri Lanka. Health Economics 1998; 7(6); 509-531.

Andersen R M. Revisiting the behavioral model and access to medical care: does it matter? Journal of Health and Social Behaviour 1995; 36; 1–10.

Augurzky B, Schmidt C. The Propensity Score: A Means to An End. IZA Discussion Paper 2001 no. 271.

Becker S O, Caliendo M. Sensitivity Analysis for Average Treatment Effect. The Stata Journal 2007; 7(1); 71-83.

Becker S O, Ichino A. Estimation of Average Treatment Effects Based on Propensity Scores. The Stata Journal 2002; 2; 358-377.

Besley T, Coate S, Loury G. The Economics of Rotating Savings and Credit Associations. American Economic Review 1993; 83(4); 792–810.

Booth A. Counting the Poor in Indonesia. Bulletin of Indonesian Economic Studies 1993; 29(1); 53-83.

Caliendo M, Kopeinig S. Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys 2008; 22(1); 31-72.

Case A, Lubotsky D, Paxson C. Economic Status and Health in Childhood: the Origin of the Gradient. American Economic Review 2002; 92(5); 1308-1334.

Castro-Leal F, Dayton J, Demery L, Mehra K. Public Spending on Health Care in Africa: Do the Poor Benefit? Bulletin of the World Health Organization 2000; 78(1); 66-74.

Cobb-Clark D A, Crossley T. Econometrics for Evaluation: An Introduction to Recent Developments. Economic Record 2003; 79(247); 491-551.

Dehejia R H, Wahba S. Propensity Score Matching Methods for Nonexperimental Causal Studies. Review of Economics and Statistics 2002; 84(1); 151-161.

Dow W H, Gertler P, Schoeni R F, Strauss J, Thomas D. Health Care Prices, Health and Labor Outcomes: Experimental Evidence. RAND Working Paper 2000; California.

Frankenberg E, Beegle K, Thomas D, Suriastini W. Health, Education and the Economic Crisis in Indonesia. RAND working paper 1999, Santa Monica, USA.

Fehr E, Gachter S. Cooperation and Punishment in Public Goods Experiments. American Economic Review 2000; 90(4); 980–994.

Filmer D, Hammer J S, Pritchett L H. Weak Links in the Chain: A Diagnosis of Health Policy in Poor Countries. World Bank Research Observer 2000; 15(2); 199-224.

Galiani S, Gertler P, Schargrodsky E. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. Journal of Political Economy 2005; 113; 83–120.

Gertler P J, Molyneaux J W. The Asian Demographic Transition and Its Causes: The Impact of Targeted Family Planning Programs in Indonesia. Population and Development Review 2000; 26 (supplement); 61-85.

Gertler P J. Progress Report of the Indonesian Resource Mobilization Study. Research Advisory Committee, Development Economics Research Group, The World Bank 1995, quoted in Lanjouw P, Pradhan M, Saadah F, Sayed H, Sparrow R. Poverty, Education and Health in Indonesia: Who Benefits from Public spending? Policy Research Working Paper 2739, The World Bank 2001, p.23.

Heckman J, Ichimura H, Todd P E. Matching as an Econometric Evaluation Estimator: Evidence for Evaluating a Job Training Programme. Review of Economic Studies 1997; 64(4); 605-654.
------- Characterizing Selection Bias Using Experimental Data. Econometrica 1998; 66(5); 1017-1098.

International Labor Organization. Indonesia: Providing Health Insurance For the Poor. Social Security Extension Initiatives in South East Asia, International Labor Organization http://www.ilo.org/public/english/region/asro/bangkok/events/sis/download/paper25.pdf

Jensen E R. The Fertility Impact of Alternative Family Planning Distribution Channels in Indonesia. Demography 1996; 33(2); 153-165.

Lanjouw P, Pradhan M, Saadah F, Sayed H, Sparrow R. Poverty, Education and Health in Indonesia: Who Benefits from Public spending? Policy Research Working Paper 2739, World Bank 2001.

Manning W G, Newhouse J P, Duan N, Keeler E B, Leibowitz A, Marquis M S. Health Insurance and the Demand for Medical Care: Evidence from A Randomized Experiment. American Economic Review 1987; 77(3); 251–277.

Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease 1959; 22(4); 719-748.

McGuire T. Physician Agency. In: Culyer and Newhouse (Eds), Handbook of Health Economics vol.1A. North Holland: Amsterdam; 2000.

Miguel E, Gugerty M K. Ethnic Diversity, Social Sanctions, and Public Goods in Kenya. Journal of Public Economics 2005; 89(11); 2325-2368.

Million A, Riphahn R T, Wambach A. Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation. Journal of Applied Econometrics 2003; 18(4); 387-405.

Mortality Country Fact Sheet 2006. World Health Organization, viewed 5 September 2007, <http://www.who.int/whosis/mort/profiles/mort_searo_idn_indonesia.pdf>.

World Health Organization. Mortality Country Fact Sheet 2006. Viewed 5 September 2007, <http://www.who.int/whosis/mort/profiles/mort_searo_idn_indonesia.pdf>.

Newman J, Pradhan M, Rawlings L B, Ridder G, Coa R, Evia J L. An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund. The World Bank Economic Review 2002; 16(2); 241-274.

Poterba J M. Government Intervention in the Markets for Education and Health Care: How and Why? NBER Working Paper 1994 no. 4916.

World Bank. Poverty 2006. Viewed 5 September 2007, <http://siteresources.worldbank.org/DATASTATISTICS/Resources/table2-7.pdf>.

Pradhan M, Rawlings L, Ridder G. The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation. World Bank Economics Review 1998; 12; 457–482.

Pradhan M, Saadah F, Sparrow R A. Did the Health Card Program Ensure Access to Medical Care for the Poor during Indonesia's Economic Crisis? World Bank Economic Review 2007; 21(1); 125-150.

Rosenbaum P R. Observational studies (2nd Ed.). New York: Springer-Verlag; 2002.

Rubin D B, Thomas N. Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics 1996; 52; 249-264.

Smith J, Todd P. A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies. Swiss Journal of Economics and Statistics 2000; 136(3); 1-22.
----- Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators? Journal of Econometrics 2005; 125(1); 305-375.

Thomas D, Frankenberg E, Smith J P. Lost But Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey. Journal of Human Resources 2001; 36(3); 556-592.

Trujillo A, Portillo J, Vernon J. The Impact of Subsidised Health Insurance for the Poor: Evaluating the Colombian Experience Using Propensity Score Matching. International Journal of Health Care Finance and Economics 2005; 5(3); 211-239.

Wagstaff A. Health Systems in East Asia: What Can Developing Countries Learn from Japan and the Asian Tigers? World Bank Working Paper 2005 no. 3790.

Wagstaff A, Yu S. Do Health Sector Reforms Have Their Intended Impacts?: The World Bank's Health VIII Project in Gansu Province, China. Journal of Health Economics 2007; 26(3); 505-535.

Waters H, Saadah F, Pradhan M. The Impact of the 1997–98 East Asian Economic Crisis on Health and Health Care in Indonesia. Health Policy and Planning 2003; 18(2); 172-181.

Weerdt J D, Dercon S. Risk-sharing Networks and Insurance Against Illness. Journal of Development Economics 2006; 81(2); 337-356

Whitehead M, Dahlgren G, Evans T. Equity and Health Sector Reforms: Can Low-income Countries Escape the Medical Poverty Trap? Lancet 2001; 358, 833–836.

Widianto H. 2007, personal communication, 12 January, at the Ministry of Health, Jakarta, Indonesia.

Xu K, Evans D B, Kawabata K, Zeramdini R, Klavus J, Murray C J L. Household Catastrophic Health Expenditure: A Multicountry Analysis. Lancet 2003; 362(9378); 111-117.

Zhao Z. Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence. Review of Economic and Statistics 2004; 86(1); 91-107.