Discussion Paper 3

Propensity Score Matching Method in Quasi-Experimental Designs

An Approach to Program Evaluation of INHP-III

SAMBODHI

Sambodhi Research & Communications Pvt. Ltd. is a premier consulting and advisory institution catering to research and allied services to the social sector with belief that design and development of state-of-art knowledge ware products and provision of knowledge-based service are quintessential for expediting the developmental change process. Sambodhi, as the name signifies and means unexcelled complete enlightenment, state of universal knowledge or perfect wisdom. Registered under the Companies Act, Sambodhi, provides both customized and syndicated services to the sector.

Sambodhi has proficiency in developing knowledge based products encompassing sectoral know-hows, data management & analytic solutions and disseminating knowledge through capacity building services to a variety of organizations.

Propensity Score Matching Method in Quasi-Experimental Designs: An Approach to Program Evaluation of INHP-III

Gakhar, Kaushal Deep[1]; Kaur, Navneet[2]; and Kapur, Vidhu[3]

Abstract

The experimental designs are generally considered as the robust evaluation methodologies as there is random assignment. These are possible in clinical trials or in pilot phase of the project but during the development phase due to ethical issues and resource constraints; use of true experimental designs are not feasible in majority of development interventions as use of experimental design entails creation of treatment and comparison group thereby providing benefits to some and excluding others. It is unethical at program-level to provide the benefits to few and leave others and thus, there is difficulty in construction of both treatment and comparison at baseline. This makes attribution of observed outcomes and impacts to program intervention very difficult. The task gets more difficult when there are no baseline studies available. PSM offers one such alternative for addressing the concerns comparison and attribution**.** This paper is based on the case of Endline Evaluation of INHP-III where the Quasi-Experimental Design was employed using the PSM technique to construct the ideal comparison match for the treatment groups.

Key Words:
PSM, Counterfactual, Treatment, Comparison, Quasi-experimental, Matching Groups

[1] Manager- Research, Sambodhi,kaushal@sambodhi.co.in
[2] Senior Manager – Research, Sambodhi,navneet@sambodhi.co.in
[3] Senior Executive – Research, Sambodhi,vidhu@sambodhi.co.in

Background

A woman's nutritional status has important implications for her health as well as the health of her children. India has made momentous progress in social and economic development in recent times but is lagging in women and child health indicators. Infant and young child malnutrition has profound negative consequences on the health and development of a child and thus of society. Child malnutrition contributes to more deaths than any other health condition, globally accounting for or contributing to about six million of the 10.9 million deaths of under five children each year .

Under this context Integrated Nutrition and Health Project began in 1997. It aimed to significantly improve the health and nutritional status of women of childbearing age and children under two years of age. With focus on child health and nutrition interventions, this project supported the Government of India's (GOI) Integrated Child Development Services (ICDS) scheme and the Reproductive and Child Health (RCH) program of the National Rural Health Mission (NRHM). INHP in its implementation phase promoted a set of simple interventions to influence neonatal outcomes, including antenatal care, delivery care, early and exclusive breastfeeding and child feeding practices.

INHP III was implemented in 711 blocks of 75 districts from eight states viz. Andhra Pradesh (AP), Chhattisgarh (CG), Jharkhand (JH), Madhya Pradesh (MP), Orissa (OR), Rajasthan (RJ), Uttar Pradesh (UP) and West Bengal (WB). The primary target groups of INHP were pregnant women and lactating mothers and children under two years of age. The project was implemented in three phases. During the third phase of the project (March 2006 – September 2009), the project adopted a dual strategy of replication in new areas and phase-out in existing Primary Project Areas (PPA) in a systematic manner to ensure the sustainability of the process. The project's approach for the phase-out was quite complex in a sense that the project under went gradual phase-out from the current blocks. By September 2007, the project was phased-out from 341 of the 711 blocks and another 237 blocks had been phased-out in September 2008. The remaining 133 blocks were phased-out in September 2009. In the Replication Areas the focus was standardization and scaling up successful practices with the support of government systems and to ensure ownership of these processes by the system.

Since INHP was the first-of-its-kind large scaled intervention focused on nutrition and health outcomes, it was important to consolidate gains made over the programme period from time to time. The final evaluation of the programme was designed in a way to account for the complex project design adopted for the implementation across states and districts. The main objective of the endline evaluation conducted in 2009 was to measure change in the impact indicators over the project period and attribute that change to the project activities

Need for Attribution and Counterfactual

Evaluation approaches for development programs have evolved considerably over the decades, encouraged by rapidly expanding scope of impact evaluation researches in the sector. The basic purpose of any evaluation study is to assess the following three things:

Whether there is a change (from the situations at the start of the project)?

Is it because of the project?

What contributed towards the change?

The purpose of INHP end-line evaluation was to detect whether the change in outcome indicators is due to project; and understand how "changes" in the input and processes have affected output and outcome indicators over the period of implementation of the program. To truly identify the change due the project, there is a need to compare the observations on key indicators for individual in the project areas to that of individuals in non-project areas. This is called the counterfactual. The counterfactual question can be defined as: "What would have happened to those who, in fact, did receive treatment, if they had not received the treatment (or the converse)?".

An important problem of causal inference is estimation of treatment effects in observational studies. In a situations (like an experiment) where a group of units is exposed to a well-defined treatment, but (unlike an experiment) no systematic methods of experimental design are used to maintain a control group. Thus, the estimation of counterfactual could be biased because of problems such as self-selection or some systematic judgment by the researcher in selecting units to be assigned to the treatment. This paper discusses the use of propensity score-matching methods to correct for sample selection bias due to observable differences between the treatment and comparison groups even in the absence of comparison at the baseline. This paper compliments the existing literature on propensity score-matching methods by focusing on matching method adopted in the context of a different data set. Previous papers include Daheja & Sadek (1999), Heckman et al. (1996, 1998) and others.

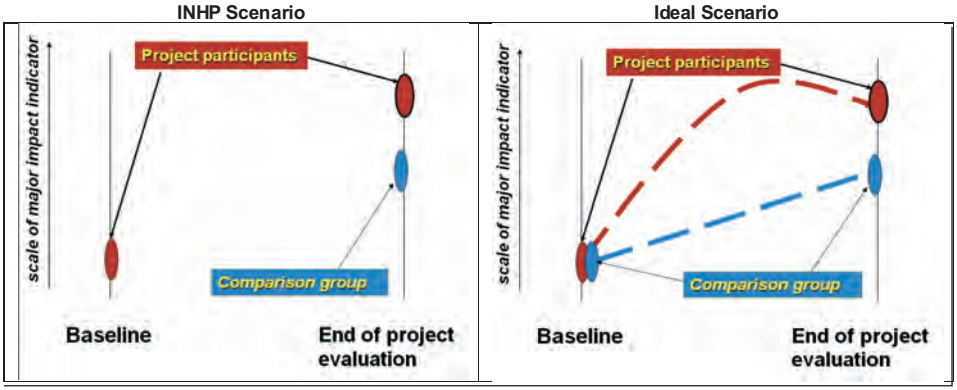
Evaluation Methodology

A comprehensive study design was followed for the final evaluation of INHP-III ensuring comparison not only to assess the change but also, to assess the efficacy of phase-out and also, to attribute change to the intervention carried out. In order to assess the effect of phase-out, a separate sample from blocks phased-out in September 2007 was drawn out for comparison with the survey conducted in these blocks in September 2007. The main sample for 2009 survey was drawn in a manner that proportionately represented blocks phased-out at different junctures (that is, stratified by year of phase-out), and then drawing additional sample from blocks phased-out in September 2007 to make up the full sample needed for that stratum.

The study adopted a quasi-experimental design wherein a comparison group is selected in addition to treatment group. Though the evaluation was carried out in seven states, it is important to point that due to adequacy problem of finding a comparison sample in three states i.e. Jharkhand, Chhattisgarh and Andhra Pradesh, the quasi-experimental design was employed in four states i.e. Rajasthan, Orissa, West Bengal and Uttar Pradesh wherein comparison sample were available. The design for assessments in comparison areas was similar to that in the main program areas.

For attributing change, equivalent samples from comparison areas in four states (Orissa, Rajasthan, Uttar Pradesh, West Bengal) is taken that has been compared to the "main" stratum of 2009. This was achieved by a drawing these sample carefully in a manner as similar to the main sample as possible.

For sampling, a two-stage cluster design was adopted with block as the first strata. Anganwadi center (AWC) were the first primary sampling unit and mothers of children aged 0-5 months and 6-23 months were the secondary sampling unit. In order to assess and quantitatively measure changes in key project indicators i.e. change in nutrition status, the study selected a sample size of 733 for each age-group in each state for each stratum. At the level of primary sampling unit, requisite number of AWCs was spread across the sampled blocks in proportion to the number of AWC in the blocks.



Quasi-Experimental Design with Propensity Score Matching in INHP-III Evaluation

Evaluation measures the impact that a project has on its beneficiaries. It typically does this by comparing outcomes between treatment and comparison group, both before and after a project has been implemented. Difference-in-difference (DID) is one of the commonly applied estimators in evaluation research, which compares outcomes for two groups for two time periods. In the present evaluation research absence of comparison group in the baseline debilitated the application of DID. Thus, in order to attribute change to intervention, propensity score matching (PSM) method is used. Essentially, propensity score matching aided selection of a comparison group comprising individuals who did not in fact receive the project benefits, but who, given their observable characteristics, had the same probability of receiving the project as individuals in the treatment group. The project's impact is then estimates by the difference in outcomes between the treatment and comparison group.

Propensity score matching (PSM) approach was used for constructing a comparison group and for reducing the existing differences between the treated and the comparison group. PSM is commonly used to estimate causal treatment effects. The idea of the matching is to find, in a group of non-treated, those individuals who are similar to the treated individuals in all relevant pre-treatment characteristics. That being done, differences in outcomes between the treated and the adequately selected comparison group are more likely to be attributed to the project intervention. Substantive literature evidences have shown that the experimental estimators using PSM provides reliable and low-bias estimates of programme impact .

In an ideal scenario, the most reliable method for measuring programme impact is experimental design method wherein a comparison group is constructed by randomly allocating the programme to a subset of eligible households. However selection of comparison group may lead to bias in estimates of programme impact. Randomly selecting treatment and comparison group raises serious ethical concerns in many settings. Thus, for such kind of evaluation it is important to construct a counterfactual comparison which eliminates the obvious selection bias.

The propensity score was originally proposed as a method for producing balance of many covariates between two groups. PSM employs a predicted probability of group membership - e.g. treatment vs. comparison group-based on observed predictors, usually obtained from logistic regression to create a counterfactual group.

An important point to note here is that matching of cases put together heavy demands on data. Thus, matching is a suitable approach to evaluation only when informative data available. Besides the condition of data quality cannot be ignored.

The **propensity score** or the **probability of participating** in the program (being treated), is a function of the individual's observed characteristics

$$P(X) = Prob(D = 1|X)$$

Where, D indicates participation in project

X is the set of observable characteristics

To measure the effect of a program, we maintain the assumption of selection on observables i.e., assume that participation is independent of outcomes conditional on X_i

$$E (Y|X, D = 1) = E (Y|X, D = 0)$$

if there had not been a program.

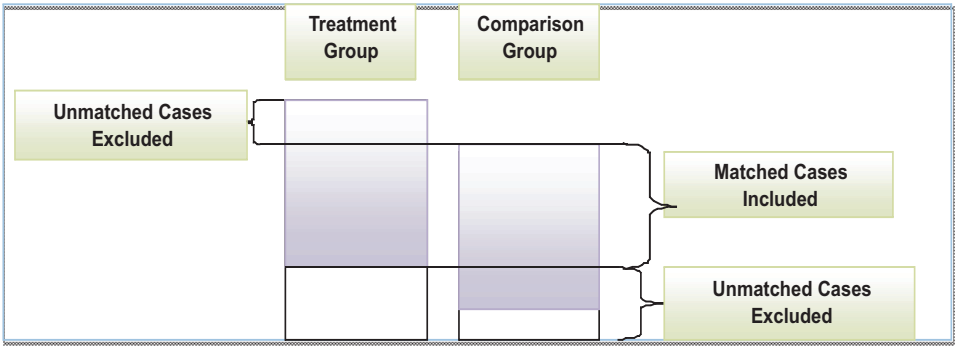
This is false if there are unobserved outcomes affecting participation

Pre-requisites for propensity score matching

While working with PSM it is important to consider the following:

- Large and comparable sample size (in both treatment and comparison group) so as to get substantial number of appropriate matches
- Rich data on as many observable characteristics as possible
- Use of same set of question in treatment and comparison group
- Administration of questionnaire in treatment and comparison group at the same time
- Group overlap must be substantial
- Hidden bias may remain because matching only controls for observed variables (to the extent that they are perfectly measured)
- Requires good quality data

Unless these conditions are satisfied practical constraints exist in the use of matching technique. Small sample size pose problem for propensity score matching by reducing the common support region. With large sample even if the treatment and comparison groups are different, it is possible to estimate the impact of treatment on the treated, while with smaller samples treatment effects can only be retrieved for a sub-set of the treated population (Shahdish, Cook & Campbell, 2002). As the number of characteristics used in the match increases, the chances of finding a match is reduced. It is easy to see that including even a relatively small number of characteristics can quickly result in some participants remaining unmatched.



In order to have robust estimate for PSM the sample size should be sufficient enough as while running PSM there is probability of 25-35% sample loss. This is to ensure that even after such loss of sample the remaining sample size is adequate to establish the statistical significance of the results. The loss of sample while performing PSM is clearly evident in INHP-III evaluation. An example of reduction in sample size for one of the indicator is as illustrated in the table below.

	Sample Difference	
	Before matching	**After matching**
Project	1880	1,370
Comparison	1844	1,686
Total	3,724	3,056

Secondly, if the two groups do not have substantial overlap, then substantial error may be introduced. For example in a bid to treat only the best cases in project and control group, the result may be regression toward the mean (Bryson, Dorsett, & Purdon, 2002).

Thus, the primary condition that needs to be met in order for PSM to be feasible is to have a comparable group available to provide matches.

Further, it is critical to build a set of matching indicators strongly, essentially socio-demographic indicators which are not outcome indicators, at the start of intervention in order to generate comparable estimates during evaluation. It is to be noted that matching indicator should not be the outcome indicator as this may affect the intervention results.

Also, the treatment and comparison groups were subjected to the same set of questions. This aids the tracking of indicators and impact at the end of intervention.

The INHP-III evaluation was designed in a way to fulfil the above requirements for PSM. A comprehensive household survey was designed and questionnaires were prepared to meet these requirements. The variables included in the questionnaire capture many of the determinants of participation that are typically unobservable to the researcher. This helped to reduce a potentially significant source of bias in PSM estimators.

Further towards fulfilment of the pre-requisites of PSM to develop robust PSM estimates, matching for selection of comparison group has been done at district and tehsil / block level.

[4] Sub-district level administrative unit. A typical district may have 4-5 tehsils.

Method of Selection of Comparison Group in INHP-III Evaluation

Sample of comparison districts has been drawn carefully in a manner as similar to the main sample as possible. First, a set of non-INHP districts were matched with treatment districts in each of the four states, using census characteristics and available information about the state of public health services (mainly, vacancies in ICDS posts). From all available ICDS blocks in these districts, samples were drawn in a manner similar to the method followed in the project areas. Detailed steps for selection of comparison group are described below.

I. At the first level for selection of comparison area matching was done at the district-level. The districts were selected based on the demographic characteristics and geographical distance. The scores of all the selected project districts and comparison districts are calculated based on the socio-demographic characteristics. The socio-demographic indicators from Census of India 2001 are used for matching districts for comparison. The socio-demographic indicators used for matching are:

- Rural Population %,
- SC Population %,
- ST Population %,
- Female Literacy %,
- Work Population (M) %,
- Work Population (F) %,
- Bathroom Facility in house %,
- Type of Latrine within house (Pit, Water Closet, Other) %,
- Source of lighting (Electricity) %,
- Type of Census Houses (Permanent, Semi Permanent) %,
- Source of Drinking Water (Tap) %

After running the matching of indicators, the districts which are near to the project districts in total score are selected as comparison districts. Further, the districts which are near to the project districts in terms of geographical distance are finally selected to be included in the study.

II. The second-level of matching is done at the level of tehsil (wherever tehsil data is available). The comparison tehsils are also selected on the basis of socio-demographic characteristics. Thus, the tehsils with the comparable scores with project tehsils are selected.

III. Third-level of comparison is done at the block level based on socio-demographic characteristics (wherever data is available); vacancy Positions of ICDS staff (CDPOs & Supervisors) at block level; and type of block (rural, urban, tribal). The vacancy positions at block level are calculated for both the project and comparison blocks and the blocks with similar vacancy positions are selected matching accordingly whether it is urban, rural or tribal block.

IV. The final matching is done at Household Level using the Propensity Score Matching.

Performing the Propensity Score Matching

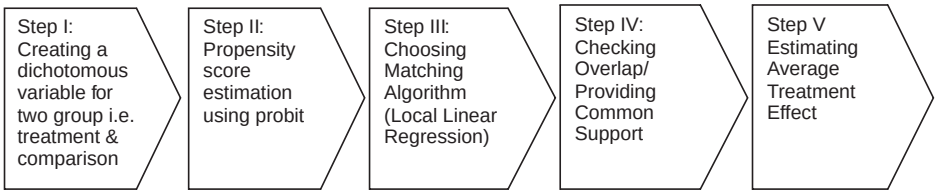
Running PSM for programme estimation involves several steps. Before proceeding to matching it is first necessary to ensure that any combination of characteristics observed in treatment group may also be seen in the non-treatment group, ensuring the existence of potential matches. This is called Common Support Requirement (CSR). Most commonly, and as adopted in the present evaluation, the CSR is enforced by discarding those in the treatment group who have probability of participation that lies outside the range evident in the non-treatment group. The programme effect cannot be estimated for these discarded individuals.

It is established that quality of the match is improved by ensuring that matches are formed only where the distribution of the density of propensity scores overlap between treatment and comparison observations, or where the propensity score densities have common support. However, by adopting CSR approach treatment individuals near the cut points of propensity score faces potential comparison individuals with propensity score which is higher or lower than that of treatment observation. To account for this problem a probit model is estimated for programme participation. For each programme outcome, propensity score is estimated using a probit model including both determinants of participation in the programme and factors that affect the outcome.

After establishing the common support region and obtaining a set of propensity scores to be used in creating the match, "balancing properties" of the data were tested. This was done by testing distribution (mean) of propensity scores in treatment and comparison observations. Thus, all impact results presented in this study are based on specifications that passed the balancing test.

A comparison group was constructed from the non-treatment group using kernel and local linear matching method was used. Kernel and local linear matching was developed from the non-parametric regression methods. All treated subjected were compared with a weighted average of all controls. Present evaluation used a modified version of PSMATCH2 command (Leuven and Sianesi, 2003) using STATA 10.0 module to compute the average treatment effect (ATE) across treatment and comparison area.

The broad steps followed for the computation of the comparison estimates are described below:



****

The steps below provide the details on performing the propensity score matching for one PTT indicator of INHP. The same procedure was followed for other PTT indicators

****

Indicator: Percentage of children 12-23 months fed at least three times in a separate bowl in addition to the breast milk

Step 1: Propensity score generation

The estimation of propensity score is dependent on two options:

Variables to be included for the estimation: Generally the variables unaffected by participation should be included in the model. The same set of questionnaire containing the variables was administered in both treatment and control area. The variables used in study for pscore estimation are socio economic variables like type of house, electricity connection, ownership of radio, animals etc.

Model to be used: The probit model is used for estimation in present evaluation

The syntax below is used to calculate the pscores:

```
pscore arm vulnerab htype lawc electric radio  cowbuffa goneouts migratio antycard fdsecuri nregawrk q1301i q1514ad q1514a1d q1507b q1528a q1529a we5 q1503a q1301a1 q1301d1 q1301c q1301d q15277_0 q15277_a q1537e dis51 dis52 dis53 dis54 dis55 dis56 dis57 dis58  dis59,pscore(pps1) comsup blockid(psblk1)

/*arm: dichotomous variable,variables from vulnerab to q1537e are SES variables used for matching,variables from dis51 to dis59 are district varibables */

**************************************************
Algorithm to estimate the propensity score
**************************************************
```

The tables below presents the frequency of the intervention and comparison area for the IND16

The treatment is arm

arm	Freq.	Percent	Cum.
control	1,844	49.52	49.52
intervention	1,880	50.48	100.00
Total	3,724	100.00	

Step 2: Estimation of the propensity score using probit model

The propensity scores are estimated using the probit model.

```
Probit regression                 Number of obs   =      3056
                                  LR chi2(25)     =    182.36
                                  Prob > chi2     =    0.0000
Log likelihood = -2010.7113       Pseudo R2       =    0.0434
```

arm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
vulnerab	-0.1070392	0.0513159	-2.09	0.037	-0.2076164	-0.0064619
htype	0.3615038	0.0569902	6.34	0	0.2498051	0.4732024
lawc	0.0168012	0.0591502	0.28	0.776	-0.099131	0.1327335
electric	0.0122348	0.0589784	0.21	0.836	-0.1033607	0.1278304
radio	0.1621624	0.0609376	2.66	0.008	0.042727	0.2815979
cowbuffa	-0.0113319	0.0487351	-0.23	0.816	-0.1068509	0.0841872
goneouts	-0.0898709	0.0651188	-1.38	0.168	-0.2175014	0.0377595
migratio	-0.0560598	0.0688487	-0.81	0.416	-0.1910008	0.0788812
antycard	-0.072669	0.1152601	-0.63	0.528	-0.2985746	0.1532366
fdsecuri	0.5237831	0.0850879	6.16	0	0.3570139	0.6905523
nregawrk	0.1005627	0.0541778	1.86	0.063	-0.0056239	0.2067492
q1301i	0.1080402	0.047722	2.26	0.024	0.0145067	0.2015737
q1514ad	-0.0325223	0.0688591	-0.47	0.637	-0.1674836	0.102439
q1514a1d	0.1296817	0.0596932	2.17	0.03	0.0126852	0.2466783
q1507b	0.0048781	0.0015254	3.2	0.001	0.0018884	0.0078677
q1528a	0.0103052	0.0030486	3.38	0.001	0.0043301	0.0162803
q1529a	-0.000991	0.0037721	-0.26	0.793	-0.0083842	0.0064022
we5	0.034716	0.0591881	0.59	0.558	-0.0812905	0.1507226
q1503a	0.1966934	0.0909773	2.16	0.031	0.0183812	0.3750056
q1301a1	0.0531768	0.0830794	0.64	0.522	-0.1096558	0.2160093
q1301c	0.2925762	0.1691184	1.73	0.084	-0.0388897	0.6240421
q1301d	-0.0901668	0.1279439	-0.7	0.481	-0.3409323	0.1605988
q15277_0	-0.000202	0.0005834	-0.35	0.729	-0.0013454	0.0009414
q15277_a	0.0005516	0.0005663	0.97	0.33	-0.0005583	0.0016615
q1537e	0.0341639	0.0678985	0.5	0.615	-0.0989147	0.1672425
_cons	-1.116122	0.3286403	-3.4	0.001	-1.760245	-0.4719987

Step 3: Choosing the matching algorithm

The identification of neighborhood and common support is very important in PSM, because the PSM estimated differ on the way the neighborhood for each treated individual is defined and the common support is handled. There are various matching techniques but we have used the local linear matching as it uses the weighted average of all the individuals in the comparison group to estimate the counterfactual outcome as compared to other techniques like near neighbor matching or interval matching that uses on few observations from comparison group.

Step 4: Providing common Support

The overlap between the treatment and comparison is identified by providing the common support. Implementing the common support condition ensures that any combination of characteristics observed in the treatment group can also be observed among the control group . There are two approaches to identify the overlap:

Maxima Minima Comparison

Estimating Density Distribution in treatment and comparison

The second method is used to identify the region of common support with the trim of 5% because it excludes the points which have estimated density zero but also the additional 5% of the points with the low positive density and also the common support can be easily identified with density distribution even if the distribution is thin as compared to the maxima minima comparison.

The region of common support is [.30477797, .97358306]

Description of the estimated propensity score in region of common support

Estimated propensity score

Percentiles	Smallest	Largest		
1%	.3576768	.304778		
5%	.391907	.3099542		
10%	.4149267	.3130384	Obs	3056
25%	.4601646	.3179162	Sum of Wgt.	3056
50%	.534039		Mean	.5517541
75%	.6198257	.9599863	Std. Dev.	.1186278
90%	.735133	.9615634	Variance	.0140725
95%	.7848193	.9648018	Skewness	.6809354
99%	.8594617	.9735831	Kurtosis	3.040792

The balancing property is satisfied

This table shows the inferior bound, the number of treated and the number of controls for each block

Inferior of block of pscore	arm: control	arm: intervent	Total
.2	126	73	199
.4	535	459	994
.5	419	488	907
.6	208	,,, 375	583
.7	67	198	265
.8	15	93	108
Total	1,370	1,686	3,056

It is to be noted that if balancing property is not satisfied then step I has to be repeated using different set of variables. This has to be repeated until balancing property is satisfied.

****

End of the algorithm to estimate the pscore

Step 5: PSM estimate using PSMATCH2

psmatch2 arm, outcome(ind16) pscore(pps1) llr kerneltype(tricube) common trim (5)
There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.
(668 missing values generated)

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ind16	Unmatched	.341043891	.305109489	.035934402	.017028809	2.11
	ATT[5]	.345193508	.324741857	.020451651	.	.

psmatch2: Treatment assignment	psmatch2: Common support Off support	On support	Total
Untreated	0	1,370	1,370
Treated	84	1,602	1,686
Total	84	2,972	3,056

Step 6: Bootstraping to estimate Standard Error

The bootstrapping is done to estimate the standard error. The standard error estimated with bootstrapping is not due to the normal sample variation (Heckman ,Ichimura & Todd ,1998)but also, due to variance in estimation of propensity score, common support and order in which treated individuals are matched..

In boothstrapping sample is treated as population and the estimate of sampling distribution is done using the techniques like Monte Carlo technique by drawing the large number of resamples from original sample with replacement.

The syntax for bootstrapping is indicated below:
bootstrap r(att): psmatch2 arm, outcome(ind16) pscore(pps48) llr kerneltype(tricube) common
(running psmatch2 on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..... 50

Bootstrap results Number of obs = 3056
Replications = 50
command: psmatch2 arm, outcome(ind16) pscore(pps1) llr kerneltype(tricube) common
_bs_1: r(att)

	Observed	Bootstrap	Normal-based

[5]Average Impact of Treatment on Treated

```
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |   .0116496   .0198487     0.59   0.557    -.0272532    .0505524
```

Result

Table below provides the PSM estimate for some of the indicators from the endline assessment of INHP-III. The whole process of PSM is run each time for all the Performance Tracking Indicators of the project.

The findings show that 33.3% of new born were put to breastfed within 1 hour post-partum in the project group as compared to 24.3% in the comparison group. The estimate difference in percentage between project and comparison is established as statistically significant (p=0.0000<0.05). Similarly, the percentage of children 18-23 months who received at least two doses of Vitamin A is 25.8% and 11.3% for the project and comparison group, respectively, and the difference between project and comparison is found to be statistically significant (p=0.0000<0.05).

Indicator for 12-23 month old children who are fed at least three times in a separate bowl in addition to the breast milk show that 34.5% of the children in the age group of 12-23 months were fed at least three times in a separate bowl in addition to the breast milk in the project group as compared to 32.4% in the comparison group. The estimate difference in percentage between project and comparison is not statistically significant (p=0.3640>0.05).

Indicator (Source: Final Survey Report - INHP-III)	Project	Comparison	ATT	S.E	P value	Confidence Interval
% of newborns put to breast within 1 hour post-partum	33.3	24.3	9.00	1.4	0.0000	*(6.2,11.7)*
Percentage of children 18-23 months who received at least two doses of Vitamin A	25.8	11.3	14.5	2	0.0000	*(10.8,18.8)*
Percentage of children 12-23 months old, in program catchment area, receiving measles vaccine	70.2	65.3	4.9	1.7	0.0050	*(1.4,8.3)*
Percentage of children 12 – 23 months fed at least three times in a separate bowl in addition to the breast milk	34.5	32.4	2.1	2.2	0.3640	*(-2.3,6.4)*
% of children under 12 months of age exclusively breastfed till 6 months post-partum	30.7	27.1	3.6	1.1	0.0010	*(1.4,5.6)*

Conclusion

The use of this evaluation technique in INHP-III evaluation intended to attribute the impact of programme. To know the impact of the programme, observed outcomes are compared with the outcomes that would have resulted had the group not participated in the programme. Use of PSM ensured that the project and comparison share almost exactly the same characteristics, and selection bias has been mitigated in the new sample.

Large-scale experimental studies generally use Randomized Control Trials (RCT) to remove selection bias. RCT is accepted as the gold standard for determining intervention efficacy because the biases associated with other non-experimental or quasi-experimental designs can be avoided. However, in the absence of RCT, application of PSM can be established as a robust technique for evaluating programme in quasi-experimental design. PSM can be used for proving internal validity and reducing the selection bias, thereby bringing in the robustness in the evaluation design. By using PSM the internal validity is enhanced as it constructs the ideal comparison match for the treatment group at the household level.

References:

Akhter U. Ahmed, A. R. (2007). *Relative Efficacy of Food and Cash Transfers In Improving Food Security and Livelihoods Of The Ultra-Poor In Bangladesh.* International Food Policy Research Institute.

Bryson, A., Dorsett, R., & Purdon, S. (2002).Retrieved 07 08, 2010, from http://eprints.lse.ac.uk: http://eprints.lse.ac.uk/4993/1/The_use_of_propensity_score_matching_in_the_evaluation_of_active_labour_market_policies.pdf

D'Agostino. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat. Med. , 17*, 2265-2281.

Daheja, R., & Sadek, W. (1999). Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs. *Journal of the American Statistical Association* , 1053–1062.

Daheja, R., & Wabha, S. (2002). Propensity Score Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics , 84(1)*, pp. 151-161.

Handa, J. J., & Sudhanshu. (July, 2005). *An Assessment of Propensity Score Matching as a Non Experimental Impact Estimator: Evidence from Mexico's PROGRESA Program.* Office of Evaluation and Oversight.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an Economic Evaluation Estimator. *Review of Economic Studies , 65:2*, 261–294.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (November 1996). “Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method,”. *Proceedings of the National Academy of Sciences*, *93:23*, pp. 13416-13420.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (September 1998). Characterizing Selection Bias Using Experimental Data. *Econometrica* (pp. 1017–1098). 66:5.

Marco, C., & Sabine, K. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Discussion Paper , No.. 1588*.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects . *Biometrica , 70*, 41-55.

Shahdish, W. R., Cook, T. D., & Campbell, D. T. (2002). Construct Validity and External Validity. In *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Inference. Boston: Houghton Mifflin.

Wang-Sheng, L. (March, 2006). *Propensity Score Matching and Variations on the Balancing Test.* Melbourne: Melbourn Institute of Applied Economics and Social Research; The University of Melbourne.

Acknowledgements
The present paper highlights the evaluation technique used in the Evaluation of INHP-III, a programme implemented by CARE-India and supported by USAID. The authors thank CARE-India for their support in bringing out this script in the form of paper. We extend our sincere gratitude to Dr. Ramesh V. Babu, USAID; Mr. Mukesh Kumar, Programme Director, CARE-India; Mr. George Kurian, Programme Manager, CARE-India and Ms. Mercy Manoranjini, Monitoring and Evaluation Officer, CARE-India for their constant support in accomplishment of the study. Special thanks are due to Dr. Sridhar Srikantiah who has been the mentor during the study. We would also like to acknowledge the inputs and suggestions of Mr. Kultar Singh, Chief Executive Officer, Sambodhi at various stages compilation of this paper.

The term “Sambodhi” represents a state of universal knowledge and perfect wisdom. *En route* to the state, we strive at developing and disseminating state-of-the-art knowledge ware products and services through dynamic customized and syndicated processes.

South Asia
C – 126, Sector 2,
Noida – 201301,
Uttar Pradesh, India
Phone: 0120-4056400

South-East Asia
#132C, Street 135, Sangkat Thkov,
Khan Chamkarmorn, Phnom Penh,
Cambodia, Phone: +855 81738017

Sub-Saharan Africa
1st Floor,Acacia Estates
Building, Kinondoni Road,
Dar-e-Salaam, Tanzania
Phone: +255 787894173