Impact assessment of agricultural research in West Africa: an application of the propensity score matching methodology

Sabine Liebenehm

Leibniz University of Hanover, Faculty of Economics and Management, Germany
E-Mail: liebenehm@ifgb.uni-hannover.de

Hippolyte Affognon

International Livestock Research Institute, Nairobi, Kenya
E-Mail: h.affognon@cgiar.org

Hermann Waibel

Leibniz University of Hanover, Faculty of Economics and Management, Germany
E-Mail: waibel@ifgb.uni-hannover.de

Contributed Paper prepared for presentation at the International Association of Agricultural Economists Conference, Beijing, China, August 16-22, 2009

Abstract

Ex-post evaluation of agricultural research is aimed to empirically provide evidence of past investments' effectiveness. This paper is intended to measure the immediate impact of livestock research activities on cattle farmers' knowledge about trypanosomosis and its curative and preventive control strategies. According to the quasi-experimental design of the intervention, it is shown that its impact will be adequately estimated by propensity score matching (PSM). Based on data collected according to a knowledge, attitude and practice (KAP) questionnaire in the region of Kénédougou that is common to Mali and Burkina Faso, results indicate a significant gain in farmers' know-how due to participation in livestock research activities.

Keywords: African animal trypanosomosis (AAT); knowledge, attitude and practice (KAP); propensity score matching (PSM)

1 Introduction

At traditional livestock systems in tropical Africa, cattle in particular, provide besides meat and milk, transport, draft power and manure for crop production and hence, contribute to a nutritious and diverse diet. Moreover the value of cattle involves the benefit in savings and security (Steinfeld, 1988). African animal trypanosomosis (AAT), one of the most severe cattle diseases in sub-Saharan Africa, imposes a serious constraint on the livelihood of cattle farm households (Swallow, 1999; Budd, 1999, Affognon, 2007).

Research by the International Livestock Research Institute (ILRI) has developed technologies for integrated disease control based on the principle of rational drug use (Grace, 2005). One example is ILRI's research on trypanocide resistance "Improving the management of trypanocide resistance in the cotton zone of West Africa", in the region of Kénédougou from June 2003 to May 2004. So far little is known on the impact of these technologies on improving farmers' knowledge and capacities to achieve a better level of disease control. Therefore, the aim of this study is to analyse the effect of this livestock research project on farmers' knowledge and practices change of AAT that involves better diagnosis, as well as curative and preventive control strategies. As in many natural resource management projects, part of the project design has been the extension to deliver the technology to farmers (Zilberman & Waibel, 2007). Concrete information about correct disease diagnosis and management practices was provided to cattle farmers by researchers, veterinary and para-veterinary services (Affognon, 2007). The central hypothesis of this study is that the research project triggered change in farmers' behaviour, which in turn enhanced their performance in managing the disease.

Generally, in order to infer the impact of an intervention on individual outcome, it is necessary in project design to create a suitable comparison group among a large group of non-participants, which is identical to the participating group, except in the attitude of treatment assignment (Caliendo & Kopeinig, 2005; Raitzer & Kelly, 2008). Given that all farmers in the research villages got access to information, there arises a problem of selective placement. Voluntarily participating farmers might be for example more productive than those who did not attend the activities (Godtland *et al*, 2004). To overcome the problem of selection bias, this paper applies the propensity score matching (PSM) approach – "the second best alternative to experimental design" (Baker, 2000: 5). With this method a meaningful counterfactual can be formulated and causality of potential outcomes, i.e. the difference in knowledge test scores between the treatment group and the control group of the trypanocide resistance research project, can be established (Caliendo & Kopeinig, 2005). This outcome variable is grouped into four different categories to capture the effect of the research activities, on both the change in knowledge and practices:

1) Knowledge about trypanosomosis itself, comprising signs, causes, possibility of animal re-infection after being cured and animals' susceptibility to the disease;
2) Curative treatment knowledge and actual control actions in case of trypanosomosis occurrence, including the quality and quantity of trypanocides for treatment;
3) Preventive treatment knowledge and actual preventive strategies applied, involving also cattle husbandry and medical management comprising trypanocide expiry date, storage and source of medicines.
4) Finally, the total knowledge score sums all points from the three categories above.

Following the procedure applied in integrated pest management for crops knowledge categories are calculated in percentage of the maximum possible score (Godtland *et al*, 2004).

In the following, procedures of sample selection and data collection are provided. Thereafter, the methodology of PSM including a sensitivity analysis is described and based on its implementation results are discussed. Finally, conclusions are drawn.

2 Survey design

In order to measure the impact of the research activity on farmers' knowledge the project villages in the region of Kénédougou, common to south-eastern Mali and south-western Burkina Faso, were revisited from October to December 2007. The household head, i.e. the decision maker, who is responsible for livestock production and animal health management, was asked to take a specific knowledge test about trypanosomosis and its control. All farm households in the respective villages were selected for the survey if they possessed cattle, at least one animal. The test was originally developed in French. Trained interviewers conducted the survey in the respective local language, i.e. *Bambara* in Mali and *Djoula* in Burkina Faso, and in turn filled in the questionnaire in French. Questions were applied in open-ended manner, followed by option lists and the use of picture cards as visual support. Before the start-up of the survey the questionnaire had been pre-tested in each country in order to ensure common acceptance and comprehensibility. After each interrogation the questionnaire was revised to diminish the presence of potential errors when translating on the spot. In total, data from 508 cattle farmers were collected. To identify project participants, farmers were asked if they attended former research activities by ILRI.

3 Methodology

Matching on the probability of participation, given all observable treatment-independent covariates X solves the problem of selection bias. The propensity score of vector X can be defined as:

$$P(X) = \Pr(Z = 1 \mid X), \qquad (1)$$

where Z denotes the participation indicator equalling one if the individual participates, and zero otherwise. Given that the propensity score is a balancing score, the probability of participation conditional on X will be balanced such that the distribution of observables X will be the same for both participants and non-participants. Consequently, the differences between the groups are reduced to only the attribute of treatment assignment, and unbiased impact estimates can be produced (Rosenbaum & Rubin, 1983). The counterfactual group can be identified if potential outcomes, i.e. knowledge test score, Y_1 (Y_0) of participants (non-participants) are independent of participation, conditional on observables *X*:

$$Y_0, Y_1 \perp Z \mid X, \forall X \ . \tag{2}$$

This conditional independence assumption indicates that the selection is exclusively based on the vector of observables *X* that determines the propensity score (Rosenbaum & Rubin, 1983; Caliendo & Kopeinig, 2005). Additionally, in order to ensure randomized selection the common support condition needs to be applied:

$$0 < P(X) < 1 . \tag{3}$$

It guarantees individuals with identical observable characteristics a positive probability of belonging both to the participation group and the control group (Rosenbaum & Rubin, 1983; Heckman *et al*, 1999). Simultaneous adoption of both assumptions (2) and (3) ensures that participation is strongly ignorable and implies that:

$$Y_0, Y_1 \perp Z \mid P(X) . \tag{4}$$

As long as outcomes are independent of participation given observables, then they also do not depend on participation given propensity score. Therefore, the multidimensional matching problem is reduced to a one-dimensional problem. The distribution of potential outcomes will be balanced among participants and counterfactuals (Rosenbaum & Rubin, 1983; Heckman *et al*, 1997; 1998).

Building on theses underlying assumptions, unbiased impact estimates can be derived by the following three steps.

Firstly, based on the definition of propensity score in equation (1), the probability of participation can be derived by binary response models. Following Todd (1995), who finds that various methods to predict propensity score produce similar impact estimates, for computational simplicity a logit model will be applied here. The propensity score can then be defined as:

$$P(X) = \Pr(Z = 1 \mid X) = F(\beta_1 x_1 + ... + \beta_i x_i) = F(X\beta) = e^{X\beta}, \quad (5)$$

where *F(·)* produces response probabilities strictly between zero and one.

In accordance with chosen characteristics that capture relevant observable differences between participants and non-participants, Table 1 reports the results from the logit model, while the estimated coefficients are expressed in terms of odds of *Z=1*. Having tested different model specifications, the summary statistics show that the present model is statistically significant. The goodness of fit test achieves a Pearson Chi-square with a high probability value. Moreover, the area under the Receiver Operating Characteristic (ROC) curve proves a fairly accurate classification performance. Hence, the chosen observables adequately explain the probability of participation. Examining single observables, it is shown that the dependency ratio of the household, cattle herd size, farming experience, perception of drug resistance and the country of origin in particular significantly influence the participation decision. Considering a marginal change in number of cattle, the probability of participation would increase by 0.4% (*ceteris paribus*). Farming experience yields an even higher marginal effect, because more experience in both crop and livestock production enhances the participation probability by about 14.7%. Likewise, the probability increases by 5.4% when farmers observe their cattle falling sick with AAT. When farmers perceive the treatment to be ineffective, which indicates resistance, the probability of participation is affected even more strongly. Finally, the participation probability is about 37% higher for individuals living in Mali than for Burkinabes.

Table 1 Logit model to predict the probability of participation conditional on selected observables

Dependent variable: Participation (*Z=1*) **Covariates *X***		**Odds ratio**	**Marginal effects**
Household size		1.013	0.003
Dependency ratio		0.532*	-0.152*
Number of children at school		1.048	0.011
Age of household head		0.996	-0.001
Formal education of household head		0.907	-0.024
Quadratic term of education of household head		1.006	0.002
Number of cattle in household		1.012**	0.004**
Mixed farming experience of household head		1.843***	0.147***
Number of means of transport		1.043	0.01
Perception of drug resistance dummy (1 = Resistance)		2.264***	0.182***
Perception of disease dummy (1 = AAT)		1.256	0.054
Country dummy (1 = Burkina Faso)		0.208***	-0.371***
Observations	508		
Summary statistics			
Log-Likelihood	-295.82584		
Pearson Chi2(495)	520.28 Prob>Chi2	0.2086	
Pseudo R-squared	0.142		
Area under ROC curve	0.7462		

Note: *p<0.1, **p<0.05, ***p<0.01

Source: own survey

The resulting predicted probability of participation is plotted in Figure 1. While the propensity score is more or less equally distributed for participants, the distribution of non-participants is skewed to the right. In other words, there are more non-participants than participants with a probability of participation less than 50%. Therefore, the application of the common support condition (assumption 3) will be essential for impact estimations in the following.

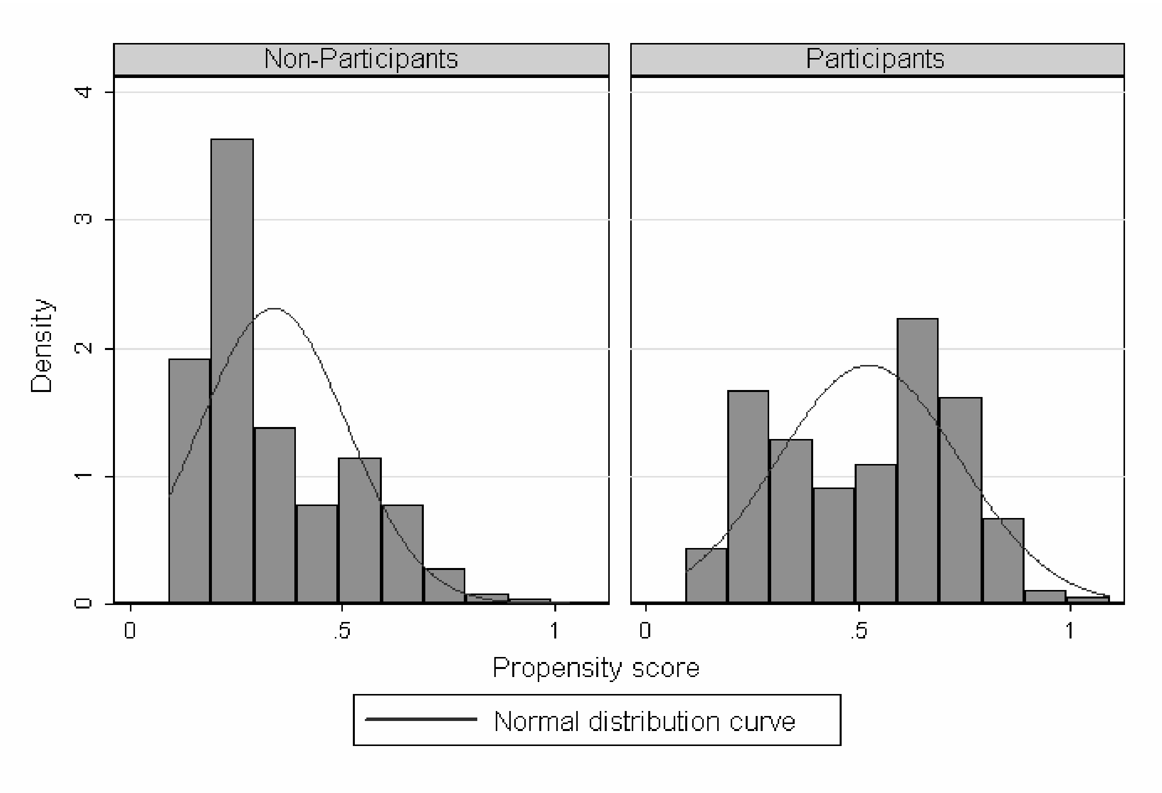

Figure 1 Histogram of propensity score for non-participants and participants

Source: own survey

The parameter of interest here is the average treatment effect on the treated (ATT) (Krasuaythong 2008). Applying the composite assumption (4) the true ATT, based on PSM, can be written as:

$$ATT_{PSM} = E_{P(x)}\{E(Y_1 \mid Z = 1, P(X)) - E(Y_0 \mid Z = 0, P(X))\}, \qquad (6)$$

where $E_{P(X)}$ represents the expectation with respect to the distribution of propensity score in the entire population. The true ATT indicates the mean difference in knowledge test scores between participants and non-participants, who are identical in observable characteristics and adequately weighted by a balanced probability of participation.

An adequate match of a participant with his/her counterfactual is achieved, as long as they are identical in their observable characteristics. In order to obtain such matched pairs Caliendo and Kopeinig (2005) report that there are different matching methods, with trade-offs in terms of bias and efficiency. Therefore, three different matching estimators are described in the following in order to associate the outcome of participating units to the outcome of their controls. In general, the matching estimator of the ATT is given by:

$$ATT_{PSM} = \frac{1}{N_1}\sum_{k=1}^{N_1}(Y_{1k} - \sum_{l=1}^{N_0} w_{kl} Y_{0l}), \tag{7}$$

where N_1 (N_0) is the total number of participants (non-participants), Y_{1k} represents the outcome of participant *k*, Y_{0l} is the average outcome of the matched counterfactual *l* to the participant *k* weighted by w_{kl} (Heckman *et al*, 1998). To begin with the most straightforward method, nearest neighbour matching (NNM) involves selection of the non-participant with the propensity score closest to that of the respective participant. A nearest neighbour will be matched only once without replacement. This one-to-one matching will cause no concern as long as the distribution of propensity scores of the two groups is similar. However, if the nearest neighbour is far away, i.e. scores are substantially different, poor matches will be obtained. The average outcome of the matched control will be equally weighted ($w_{kl}^{NN} = 1$). Hence, the impact estimator is the average difference in knowledge score between participants and controls (Smith & Todd, 2005). Secondly, radius matching (RM) involves all neighbours within a maximum propensity score distance (caliper), a priori defined, and thus corresponds to the common support assumption. Additionally, poor matches through too-distant neighbours are avoided (Dehejia & Wahba, 2002; Smith & Todd, 2005). Thirdly, kernel-based matching (KM), a non-parametric matching estimator that includes all individuals of the underlying sample of non-participants and weights more distant observed characteristics among both groups down (Heckman *et al*, 1997; 1998). Hence, kernel-based matching on all control units indicates a lower variance (Caliendo & Kopeinig, 2005). The kernel-based estimator of the ATT is therefore the mean difference in outcomes, while the matched outcome is given by a kernel-weighted average of outcomes of all non-participating units. The weight is based on the distance between participants' (*k*) and non-participants' (*l*) propensity score:

$$w_{kl}^{KM} = \frac{K\left[(P(X_k) - P(X_l))/h\right]}{\sum_{l=1}^{N_0} K\left[(P(X_k) - P(X_l))/h\right]}, \qquad (8)$$

where kernel K(·) refers to a probability density function and bandwidth *h* controls the smoothness (Heckman *et al*, 1997). Here, the choice of kernel is not as important as the choice of the bandwidth. If the smoothing parameter chosen is too high, the underlying features of the distribution will be veiled, while a too-small bandwidth will increase the variance between the true and estimated density function (Silverman, 1986).

Finally, in consideration of the quasi-experimental design of the trypanocide resistance study, it might be possible that unobservable factors like farmers' intrinsic motivation and specific abilities, as well as preferences, had affected the participation decision. This problem of hidden bias is circumvented by the following bounding approach. Within the logit model to estimate propensity score (equation 5), the probability of participation *F(·)* needs to be complemented by a vector *U* containing all unobservable variables and their effects on the probability of participation captured by γ:

$$P(X) = \Pr(Z = 1 \mid X) = F(X\beta + U\gamma) = e^{X\beta + U\gamma}. \qquad (9)$$

Rearranging the odds ratio of two individuals (*m* and *n*) who are identical in observable characteristics, the resulting relative odds of participation is given by:

$$\left(\frac{P(X_m)}{1-P(X_m)} * \frac{1-P(X_n)}{P(X_n)} \right) = \frac{e^{\beta_n x_n + \gamma_n u_n}}{e^{\beta_m x_m + \gamma_m u_m}} = e^{[\gamma(u_m - u_n)]} . \quad (10)$$

As long as there is no difference in U between the two individuals, or if the unobserved variables exerted no influence on the probability of participation, the relative odds ratio becomes one, and the selection process is random. Sensitivity analysis now examines how strong the influence of γ on the participation process needs to be, in order to attenuate the impact of participation on potential outcomes (Rosenbaum, 2002). For the sake of simplicity, it is assumed that the unobservable variable is a binary variable taking values zero or one (Aakvik, 2001). The following bounds on the odds ratio of the participation probability of both individuals are applied:

$$\frac{1}{e^{\gamma}} \leq \frac{P(X_m)(1-P(X_n))}{P(X_n)(1-P(X_m))} \leq e^{\gamma} . \quad (11)$$

Both individuals have the same probability of participation, provided that they are identical in X, only if $e^{\gamma} = 1$. Consequently there will be no selection bias on unobservable covariates. If $e^{\gamma} = 2$, one of the matched individuals may be twice as likely to participate as the other agent (Rosenbaum, 2002). If e^{γ} is close to one and changes the inference about the treatment effect, the impact of participation on potential outcomes is said to be sensitive to hidden bias. In contrast, insensitive treatment effects would be obtained if a large value of e^{γ} does not alter the inference about treatment effects (Aakvik, 2001). In this sense, e^{γ} can be interpreted as a measure of the degree of departure from a study that is free of unobservable selection bias (Rosenbaum, 2002). An appropriate control strategy for hidden bias is to examine the sensitivity of significance levels. Here, several values of e^{γ} bounds are calculated on the significance level, and hence, the null hypothesis of no effect of participation on potential outcomes respectively on knowledge score, is then tested. Therefore, the question arises at which critical impact level of the unobservable variables the inference about the treatment effect on knowledge will be undermined, as indicated by the loss of significance (DiPrete & Gangl, 2004).

In sum, unbiased impact estimates of a quasi-experimental study design can be obtained in three steps: (i) chose a binary response model with appropriate observable characteristics to predict the probability of participation; (ii) estimate the performance difference between treatment and control group according to selected matching methods that minimize the difference in observables of both groups; and (iii) analyse the effect of unobservable influences on the inference about impact estimates. Based on the implementation of these steps, the following results can be obtained.

4 Results

Based on the predicted propensity score an appropriate counterfactual group that is as similar as possible to the participating group is matched now. Table 2 shows the impact estimators obtained from the three different matching algorithms. Ensuring that observations are ordered randomly and that there are no large disparities in the distribution of propensity score (Figure 1), one-to-one matching yields the highest and most significant average treatment effects on the treated in all four outcome categories. The nearest neighbour estimate of the average total knowledge gain due to participation is about 3.16%. Since this method produces relative poor matches due to the limitation of information, the attention should be focused on the other two matching algorithms. Here, the estimated impacts of participation in research activities on knowledge score are lower regarding the respective categories.

Table 2 Estimated impact of trypanocide resistance research activities on farmers' knowledge using different matching algorithms

	Knowledge score in % of maximum scores		Average treatment effect on the treated
	Participants	Non-participants	
Nearest neighbour matching	Using the single closest neighbour		
Knowledge score on disease	25.3	22.93	2.37***
Knowledge score on control	23.54	19.29	4.25***
Knowledge score on prevention	16.01	13.0	3.01***
Total knowledge score	20.81	17.65	3.16***
Observations	211	211	
Radius matching	Using all neighbours within a caliper of 0.01		
Knowledge score on disease	25.04	23.22	1.82**
Knowledge score on control	23.17	19.27	3.9***
Knowledge score on prevention	15.79	13.18	2.6***
Total knowledge score	20.54	17.81	2.73***
Observations	194	294	
Kernel-based matching	Using a biweight kernel function and a smoothing parameter of 0.06		
Knowledge score on disease	25.28	23.37	1.91**
Knowledge score on control	23.55	19.91	3.64***
Knowledge score on prevention	16.03	13.18	2.85***
Total knowledge score	20.81	18.03	2.78***
Observations	210	293	

Note: *p<0.1, **p<0.05 and ***p<0.01

Source: own survey

Following the radius matching algorithm, considering only all neighbours within a caliper of 0.01, the difference in total knowledge scores in percentage of maximum score achieved, is about 2.73%. Moreover, the estimated treatment effect in the category of curative control knowledge and action even accounts for 3.9% at a significance level of 1%. Weighting the average outcome of the matched control with a biweight kernel function and a smoothing parameter of 0.06, like recommended by Silverman (1986), produces also the highest impact estimate due to participation in the category of curative know-how and actual executed control strategies. Similarly to the radius matching estimator in the total score category the kernel-based matching algorithm produces a significant average treatment effect on the treated of 2.78% at the 1% level.

Consequently, it can be confirmed that livestock research activities generate in fact a significant gain in farmers' knowledge on trypanosomosis and improve both curative and preventive strategies.

In order to control for unobservable influences, Table 3 compares the sensitivity of treatment effects on different knowledge scores among the three introduced matching algorithms. Overall, robustness results produced by Rosenbaum's bounds are quite similar. Kernel-based matching produces the most robust treatment effect estimates with respect to hidden bias, especially for preventive knowledge and action, as well as for total knowledge. Matched pairs might differ by up to 100% (e^{γ}=2) in unobservable characteristics, while the impact of participation on preventive treatment knowledge, as well as on total knowledge, would still be significant at a level of 5% (*p*-value = 0.023 and *p*-value = 0.0144, respectively). The same categories of knowledge score are robust to hidden bias up to an influence of e^{γ}=2 at a significance level of 10% following the radius matching approach.

Table 3 Sensitivity analysis with Rosenbaum's bounds on probability values

	Upper bounds on the significance level for different values of e^{γ}				
	e^{γ}=1	e^{γ}=1.25	e^{γ}=1.5	e^{γ}=1.75	e^{γ}=2
Nearest neighbour matching	Using the single closest neighbour				
Knowledge score on disease	0.0001	0.0072	0.0871	0.327	0.6324
Knowledge score on control	<0.0001	0.0031	0.0494	0.2284	0.5151
Knowledge score on prevention	<0.0001	<0.0001	0.0018	0.0211	0.1009
Total knowledge score	<0.0001	<0.0001	0.004	0.0074	0.0465
Radius matching	Using all neighbours within a caliper of 0.01				
Knowledge score on disease	0.0005	0.0255	0.1884	0.505	0.785
Knowledge score on control	<0.0001	0.0009	0.019	0.1149	0.3267
Knowledge score on prevention	<0.0001	<0.0001	0.0015	0.0171	0.0832
Total knowledge score	<0.0001	<0.0001	0.0007	0.0099	0.0545
Kernel-based matching	Using a biweight kernel function and a smoothing parameter of 0.06				
Knowledge score on disease	0.0001	0.012	0.1254	0.4131	0.7202
Knowledge score on control	<0.0001	0.0008	0.0194	0.1241	0.3555
Knowledge score on prevention	<0.0001	<0.0001	0.0001	0.003	0.023
Total knowledge score	<0.0001	<0.0001	<0.0001	0.0017	0.0144

Source: own survey

Also the less qualified matching algorithm, the nearest neighbour matching, is robust to selection bias on unobservable characteristics up to an impact level of e^{γ}=1.75 and e^{γ}=2, respectively. The estimated treatment effects on knowledge about trypanosomosis itself, as well as on the curative knowledge and action, are sensitive to hidden bias, at a smaller unobservable impact level of e^{γ}=1.5. Nevertheless, it has to be considered that these sensitivity results are worst-case scenarios, although they indicate information about uncertainty within the matching estimators of treatment effects (Rosenbaum, 2002).

5 Conclusions

Propensity score matching (PSM) allows measuring the short-term impact of a natural resource management project on farmers' knowledge and practice of trypanosomosis control. Due to the quasi-experimental design of the intervention, with non-randomised selection of villages and farmers, PSM is effective to overcome the selection bias on observable characteristics of project participants and non-participants. PSM creates then reliable impact estimates, respectively treatment effects, when the predicted probability of participation given observable treatment-independent covariates is balanced among those who are identical in these observables. Hence, matched participants and non-participants can only be distinguished by their treatment attribute and unbiased performance differences can be obtained.

Using three different matching algorithms significant and robust differences between matched participants and non-participants regarding cattle farmers' knowledge were identified. Hence, it can be concluded that the gain in farmers' knowledge is attributable directly to participation in the research intervention. The strongest effect of the research intervention is on the curative knowledge of ATT and subsequent adequate control decisions. Moreover, significant advancements in preventive strategies are also observable. Overall, the research project has been effective to increase farmers' knowledge and practices. However, since this study serves as a baseline to evaluate the current level of farmers' disease management, further efforts are necessary to assess the program's impact on livestock and farm productivity.

References

Aakvik, A. 2001. Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*, **63**(1), 115-143.

Affognon, H. 2007. Economic Analysis of Trypanocide Use in Villages under Risk of Drug Resistance in West Africa. Doctoral Dissertation, Leibniz Universität Hannover, Germany.

Baker, J.L. 2000. *Evaluating the Impact of Development Projects on Poverty: a Handbook for Practitioners*. Washington D.C.: World Bank.

Budd, L.T. 1999. *DFID-funded Tsetse and Trypanosomiasis Research and Development since 1980*. Volume 2. London, UK: Department for International Development (DFID).

Caliendo, M. and S. Kopeinig 2005. Some Practical Guidance for the Implementation of Propensity Score Matching. Discussion Paper No. 1588. Bonn, Germany: Institute for the Study of Labor (IZA).

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

DiPrete, T.A. and M. Gangl 2004. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, **34**, 271-310.

Godtland, E., Sadoulet, E., de Janvry, A., Murgai, R. and O. Ortiz 2004. The Impact of Farmer-Field-Schools on Knowledge and Productivity: A Study of Potato Farmers in the Peruvian Andes. *Economic Development and Cultural Change*, **53**(1), 63-92.

Grace, D. 2005. Epidemiology and Control of Cattle Trypanosomosis in Villages under Risk of Trypanocide Resistance in West Africa. Doctoral Dissertation, Freie Universität Berlin, Germany.

Heckman, J.J., Ichimura, H. and P.E. Todd 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, **64**(4), 605-654.

Heckman, J.J., Ichimura, H. and P.E. Todd 1998. Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies*, **65**(2), 261-294.

Heckman, J.J., LaLonde, R.J. and J.A. Smith 1999. The Economics and Econometrics of Active Labor Programs. In *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card, pp. 1865-2097. The Netherlands: Elsevier.

Krasuaythong, T. 2008. Technologies Adoption in Vegetable Production in Northern Thailand. Doctoral Dissertation, Leibniz Universität Hannover, Germany.

Raitzer, D.A. and T.G. Kelly 2008. Assessing the Contribution of Impact Assessment to Donor Decisions for International Agricultural Research. *Research Evaluation*, **17**(3), 187-199.

Rosenbaum, P.R. 2002. *Observational Studies*. New York: Springer.

Rosenbaum, P.R. and D.B. Rubin 1983. The Central Role of Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**(1), 41-55.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman & Hall.

Smith, J. and P. Todd 2005. Does Matching Overcome LaLonde's Critique of non-experimental Estimators? *Journal of Econometrics*, **125**(1-2), pp. 305-353.

Steinfeld, H. 1988. *Livestock Development in Mixed Farming Systems. A Study of Smallholder Livestock Production Systems in Zimbabwe*. Germany: Vauk Kiel.

Swallow, B.M. 1999. Impacts of Trypanosomiasis on African Agriculture. International Livestock Research Institute (ILRI), Nairobi, Kenya.

Todd, P. 1995. Matching and Local Linear Regression Approaches to Solving the Evaluation Problem with Semi-Parametric Propensity Score. University of Chicago: Mimeo.

Zilberman, D. and H. Waibel 2007. *International Research on Natural Resource Management: Advances in Impact Assessment*. Oxfordshire: CABI.