Impact Evaluation of the P.I.P.P.I. Programme: a First Application of Counterfactual Analysis

Francesca Santello[1 a] , Sara Colombini[1,2], Marco Ius[1], Paola Milani[1]
[1]FISPPA Department, University of Padua, Italy
[2]Department of Economics, University of Modena and Reggio Emilia, Italy

Received 5 February 2018, accepted 30 March 2018

Abstract. The research focuses on the Italian Programme of Intervention for Prevention of Institutionalization (P.I.P.P.I.), the implementation of an integrated inter-professional, institutional and service method of intervention, at national level, with families at risk of child neglect. To prove the effectiveness of the programme with the (quasi) experimental evidence of Impact Evaluation, a set of non-participant families under the care of standard services are compared to participants of the 4th edition of the programme. Since a non-random selection process intervenes in the professionals' choice of the families to be included in the intervention, specific statistical techniques have been applied to minimize the selection bias resulting by comparing participants with non-participants. Even if environmental conditions are difficult to change, statistically significant effects on children's total risk of out-of-home placement and developmental needs satisfaction are estimated. Also, the effects on parents' response to child's needs are positive and, in general, professionals' support to parents becomes less important after they have participated in P.I.P.P.I.

Citations: Francesca Santello, Sara Colombini, Marco Ius, Paola Milani. Impact Evaluation of the P.I.P.P.I. Programme: a First Application of Counterfactual Analysis – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **1(24)** 2018 – Pp. 13-21.

Keywords: child neglect; P.I.P.P.I. programme; impact evaluation; causal effect; counterfactual analysis; non-experimental study; propensity score matching; difference-in-differences estimator.
Short title: Quasi-experimental impact evaluation.

Introduction

Since the goal of an intervention is to improve a certain targets' condition or behaviour, to obtain indications about programme effectiveness is of crucial importance. Impact evaluation consists of statistical methods designed to assess the effect of a programme on the outcome variables to which the programme is addressed. The evaluator's task is to determine whether the observed changes in beneficiaries' outcomes can be related to the intervention in a "causal" sense (counterfactual approach [1,2]).

In Italy, impact assessment studies are still not widespread for cultural and feasibility reasons. This gap is even bigger when specific areas of the social field are considered. The present study is one of the first attempt to apply the counterfactual approach in the evaluation of a programme involving social, educational and health services for children and their families: the Italian experience of an intervention contrasting child neglect - *the Programme of Intervention for Prevention of Institutionalization* (P.I.P.P.I.).

In Italy the research on the phenomenon of child neglect is only at an early stage: the empirical knowledge about interventions effectively capable to support vulnerable parenting and, more specifically, to prevent child out-of-home placement is not yet developed. The intervention currently implemented by territorial services to support vulnerable families and their children is fragmented and not systematically organized across the country. There is not an ordinary and a codified practice of work with families, which explains the difficulties above mentioned in conducting research studies.

With the P.I.P.P.I. programme, the Italian government has invested for the first time in the national history of social policies on a comprehensive programme involving services located all along the peninsula with continuity and uniformity. A total of about 6,000 professionals and 2,000 families in more than 200 territories have been overall engaged across editions, from 2011 until now. Therefore, evaluation plays a key role for policy makers and all stakeholders responsible for the programme to obtain indications for their decisions.

In addition to providing an assessment of the programme effectiveness, the statistical analysis here reported will allow to achieve a deeper understanding of the children's and families' characteristics correlated with child neglect.

[a]Corresponding author, email: francesca.santello@gmail.com

This work is organized as follows:

a) description of the programme P.I.P.P.I. at national and individual family level (section 1.1) with a focus on data collection tools (section 1.2);
b) explanation of the evaluation strategy and statistical methods (section 2);
c) empirical results, separately with the aim of describing vulnerable families (section 3.1) and to achieve an estimation of the programme impact (section 3.2);
d) final discussion presenting results in synthesis, caveats and work in progress.

1. P.I.P.P.I. programme

1.1 The main features

Run through a collaboration between the Italian Ministry of Labour and Social Affairs and the University of Padua, the P.I.P.P.I. programme aims at contrasting child neglect in families where children are at risk of out-of-home placement. In line with the European 2030 Agenda for Sustainable Development, the programme experiments actions intended to break the cycle of social disadvantage (REC 2013/112/UE) by promoting positive parenting (REC 2006/19/UE).

At national level, the aim of the programme is harmonizing practices and models of intervention addressed to families in situation of child neglect through training, documentation and systematic evaluation shared throughout the entire country. Child neglect is defined as a significant deficiency or a failure to respond to the needs of a child recognized as fundamental on the grounds of the current scientific knowledge [3-4]. Considered as a complex social problem, in accordance with the bio-ecology of human development [5-6], responding to children's needs requires a collective action, which is not only a matter of helping parents, but also of promoting the responsibility of public institutions and social networks to develop relationships for families in their communities. For this reason, for each family in P.I.P.P.I., the programme provides the simultaneous functioning of four "specific activities" involving different ecosystem levels:

1) home-care intervention,
2) parents' and children's groups,
3) natural family helpers,
4) cooperation between school, families, social and health services.

These actions are integrated in a shared care plan, co-constructed for and with each child and family, by a multidisciplinary team (MT) of professionals following a *Participative and Transformative Evaluation* approach (PTE) [7].

At the individual child's level, P.I.P.P.I. proceeds in four stages, over a period of about 18 months. The intervention starts at T0 with the *PreAssessment* (phase 1), in which the MT completes a questionnaire for a wide range of households to decide which one to invite to be included in the programme. Once the family signs its agreement to enter the programme, a deeper assessment and a planning stage follow (phase 2). This second step implies the definition of the MT of professionals (teachers, psychologists, social and health workers, voluntary family helpers, etc.) who, together with the family and the child themselves, negotiate and realize the specific individual micro-plans of interventions.

To achieve the specific purposes of the micro-plans, the realization of the programme provides the four activities above mentioned (phase 3) throughout the duration of the intervention. At the end of the programme (T2), a follow-up assessment is repeated with the same instruments already used at T0 (phase 4). This *PostAssessment* re-evaluates the actual conditions of the family and the child and helps the MT to decide whether the family should continue with P.I.P.P.I. or may move back to the standard care of social welfare support, the more or less intense the interventions are.

1.2 Data collection instruments

To drive theories and methods into practice, P.I.P.P.I. provides several tools that are used to support the care process within the PTE perspective. The main tools used by professionals to give voice to children and families and foster their participation, but also to facilitate communication within the MT, are the multidimensional model of the Child's World (CW), *RPMonline* and the *Pre-PostAssessment*. The data collected through these instruments are then used to evaluate the outcomes of families and children at both individual and aggregate level.

RPMonline is a web tool designed by the University of Padua to achieve a shared Assessment (*Rilevazione*) and care Plan (*Progettazione*) and to Monitor (*Monitoraggio*) the effectiveness of the measures taken to support families and children [8]. It is based on the CW model (the so-called "Triangle"), which consists in the Italian adaptation of the British Assessment Framework [9-10], a triangular representation of the three fundamental dimensions in considering a child:

1) the child's development needs;
2) the parental responses to meet these needs;
3) the environmental factors which may influence the response to child's needs.

Each CW dimension is also divided into a number of sub-dimensions, for a total of 17 variables. This instrument has been designed to get all points of view in the team, included children and families as well; thus, it facilitates negotiation between the different perspectives about the plan. The corresponding information is updated regularly to assess progress. Professionals can indeed document, for each sub-dimension, the contents of the qualitative analysis from the points of view collected and negotiated in the team, and record the corresponding quantitative synthesis through a score defined on a Likert scale (from 1="serious problem" to 6="strength"). This forms the CW Questionnaire [11], which must be neces-

sarily completed at times T0 and T2.

The *PreAssessment* is a questionnaire used by professionals before the beginning of the implementation (T0) to choose the families to be included in the programme. The same instrument, but in the version of *PostAssessment*, is completed by professionals at the end of the intervention (T2). The *PreAssessment* is used as a guide for a shared decision-making following the path of the PTE: the MT compares and negotiates the different points of view of the members to come to an agreement about the opportunity for a family to access the programme. The questionnaire, built as a checklist, has been created specifically for the P.I.P.P.I. implementation and was inspired by the works of Braconnier and Humbeeck [12]. It consists of five following sections:

1) vulnerability conditions and social interventions accessed by the family and the children;
2) family history;
3) risk and protection factors associated to each side of the CW, defined as scores on a six-point Likert scale (from 1=few factors to 6=many factors);
4) quality of the family and social services' relationship;
5) overall risk evaluation (from 1=absence of risk to 6=high risk), to be completed considering the previous sections.

Finally, the rich Questionnaire on Socio-Demographic Information is completed at T0 and constantly updated in *RPMonline*.

2. Evaluation strategy

In the P.I.P.P.I. programme, the simple pre-post comparison of results indicates a significant improvement in all outcome variables considered on families and children from the beginning to the end of the programme [13]. The question which counterfactual analysis aims to answer is how much of this positive change can be attributed to the participation in the intervention: what would have happened to children and families in P.I.P.P.I. if they did not take part in the programme but have been instead followed by the child protection mainstream activities?

To answer this question, it is necessary to compare the outcomes observed for the participants in P.I.P.P.I. with the results that would have been observed, on the same subjects, in the absence of the programme ("counterfactual" situation [14]). Denoted by Y the outcome variable which the programme likely affects, impact evaluation aims to determine how much Y changes in mean on the "treated" group at T2 as a result of the participation in the programme, the so-called *Average Treatment Effect*:

$$ATE = E(Y^1_{T2} - Y^0_{T2}|D = 1), \quad (1)$$

where D={0,1} is the indicator of exposure to "treatment" and variables Y^1/Y^0 refer to outcomes respectively with (1) / without (0) the programme.

Since it is not possible to observe the targets in the alternative situation of no participation in the programme, the counterfactual outcome of the benficiaries ($Y^0_{T2}|D$=1) must be derived from the same outcome observed in a convenient comparison group of subjects not exposed to the intervention, the "control group" ($Y^0_{T2}|D$=0). By doing so, however exits the risk of confronting two groups of subjects systematically different from each other independently on the intervention and therefore attributing to it differences that would have been anyway observed (difference on the right-hand side of Eq.(3):

$$A\hat{T}E = E(Y^1_{T2}|D = 1) - E(Y^0_{T2}|D = 0), \quad (2)$$

$$A\hat{T}E = ATE + E(Y^0_{T2}|D = 1) - E(Y^0_{T2}|D = 0). \quad (3)$$

This a priori difference, known in the literature as "selection bias" [15], derives from the selection (self-selection) process by which the treated have been chosen (they chose) to participate in the intervention. Only in the ideal situation where assignment is random, the selection bias is null by construction. Because of practical and organizational reasons, as well as ethical motivations, in the experimentation of P.I.P.P.I. it is not possible to randomly decide which vulnerable families to include in the programme.

However, in the fourth implementation of P.I.P.P.I., for some local areas involved in the programme, the same data collected on the target families were also available for a comparison sample of vulnerable families under the standard care of social services (control families). This paper shows the comparison of the target families with the control ones, appropriately cleared from the selection bias by means of specific statistical techniques.

To estimate ATE, a two-steps evaluation strategy is applied. The average outcome at the end of the programme in the counterfactual setting of no participation in the intervention is firstly approximated by what is observed in the control group "re-sampled" thorough the Matching statistical procedure (step 1). Such technique allows to make a comparison fixed the conditions and the characteristics which potentially determine the selection bias [16-17]. Then, the combined use of the *Difference-In-Differences* (DID) estimator (step 2) allows to further control for the residual differences still exiting in the starting conditions between the two groups [18-19].

The first step implies to pair each treated family to a subgroup of control families as closely as possible with respect to a certain set of pre-treatment characteristics, X. The evaluator therefore assumes that two families with similar characteristics, one of which treated and the other not, would have presented the same outcomes if they both did not participate in the programme (Selection on observables assumption [20]):

$$E(Y^0_{T2}|D = 1, X) = E(Y^0_{T2}|D = 0, X). \quad (4)$$

Since the practical difficulty to find matches when there are many variables to control for, matching has been applied by

means of the propensity score [21], i.e. the conditional probability of receiving the treatment given the X:

$$p(X) = Pr(D = 1|X). \quad (5)$$

Only the need to obtain an estimate of the propensity score conditionally on which all covariates have approximately identical value is important in estimating the propensity score (Balancing property [22]). In the present evaluation study, various matching techniques have been tested to find the appropriate comparison group for the families participating in P.I.P.P.I. (see Ref.[23] for a detailed review of the alternative matching methods).

Kernel and Stratification Matching turned out to be preferable to improve the balancing of observables and also to obtain a comparison sample representative of the entire treated group of families. While the Kernel Matching considers all the members of the control group for each treated unit, with a weight inversely proportional to the distance between the two observed propensity scores, Stratification Matching divides the support of the propensity score into blocks and within each block, where the propensity score can be assumed approximately constant, compares treated and non-treated subjects.

At the second step, for each pair of readjusted comparison groups by means of matching, the DID estimator has been applied. Conditionally on X, this estimation strategy assumes the same trend in the average outcomes when non-participation in the programme is experienced for treated and controls:

$$E(Y^0_{T2} - Y^0_{T0}|D = 1, X) = E(Y^0_{T2} - Y^0_{T0}|D = 0, X). \quad (6)$$

Therefore, DID estimates the impact of the programme by the difference between (matched) participants and non-participants in the pre-post difference in outcomes:

$$ATE = E_X[ATE|X], \quad (7)$$

where $ATE|X$ is equal to

$$E(Y^1_{T2} - Y^0_{T0}|D = 1, X) - E(Y^0_{T2} - Y^0_{T0}|D = 0, X). \quad (8)$$

Fig. 1 represents the chart explaining how the DID estimator works. In addition to the simple difference of the pre-post change in the average outcomes, the DID estimator has been computed via the regression of the pre-post differences on the indicator of intervention participation plus other control variables. Since Kernel Matching turns out to be a re-sampling, the final estimators can be calculated by weighting the units of the baseline groups.

In Stratification Matching, where the support of the propensity score is divided in blocks satisfying the balancing property, ATE is alternatively computed by the weighted mean of the conditional ATE estimates, averaged over the blocks using the relative proportion of treated units (Blocking with regression [24]). All estimation procedures have

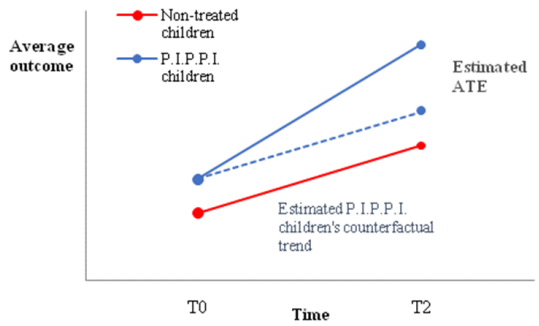

Fig. 1. The DID estimator.

been implemented using the statistical software *Stata*, which provides a specific package to perform propensity score matching [25].

3. Empirical results

Only 10 territories of the 46 overall involved in the fourth edition of the programme were selected to take part in the evaluation study, plus other two participating in the previous edition. The choice of the local areas to be included in the research was driven by a preliminary "territorial analysis" to test the starting conditions as well as the hypothesis of homogeneous treatment. To assure a good baseline comparison group for treated, practitioners were invited to follow the same inclusion criteria used for targets in the choice of the families to include in the control group. The same quantitative data, with the same tools and timelines, were collected in both the compared groups. The instruments were completed by professionals in a team also for the controls.

To facilitate the understanding of the instruments to operators who were not trained in the programme, professionals trained in P.I.P.P.I. (the "coaches") were present during the compilation with the task of supporting in the use of the tools and furthermore of conducting the assessment of the child and his/her family.

The number of families and children in each comparison group separately for local area is reported in Table 1: a total of 107 children in 97 families participating in P.I.P.P.I. vs. 146 children in 143 families in the care of standard services.

By directly comparing participants with non-participants, significant statistically differences are observed for some pre-treatment variables. This confirms the presence of selection bias, which prevents a direct comparison between the baseline groups to assess the programme effectiveness.

3.1. Propensity score and matching

To estimate the propensity score (i.e. the probability of entering the programme), a Logit model has been assumed. The specification of the model was driven by the need to obtain an estimate of the propensity score which satisfies the Balancing property and, at the same time, gives an accurate approximation to the conditional probability of participation in the pro-

Table 1. Families and children participating in the evaluation study.

Region	Territory	Treated		Non-treated	
		Families	Children	Families	Children
Emilia Romagna	Correggio-Guastalla	9	9	-	-
	Ferrara	8	8	-	-
	Modena	-	-	14	14
	Reggio Emilia	-	-	15	15
Lombardia	Bergamo	10	12	13	13
	Mantova	9	10	15	15
	Milano	9	10	15	15
Piemonte	Alessandria	9	11	14	14
	Fossano	9	9	15	15
	Torino	18	22	14	17
Veneto	Venezia	6	6	15	15
	Vicenza	10	10	13	13
Total		97	107	143	146

gramme.

Table 2 reports the results obtained from the estimation of the Logit model, where the average marginal effects on the probability to enter the programme are reported. These preliminary results of the counterfactual analysis allow to identify the key variables which come into play in the selection process of target families; it is in fact possible to outline more clearly the characteristics of the vulnerable families chosen by professionals to work with P.I.P.P.I., as compared to the vulnerable ones excluded [26].

The results, in general, refer to aspects which reduce the propensity score. It seems that vulnerable families owing to the presence of a traumatic and/or stressing event have been chosen to a greater extent. Risk conditions related to the perturbations of the family equilibrium because of conflict and absence of parents, stepfamily and adoption, as well as problems of disability or pathologies of parents are, instead, predictors of the non-participation in P.I.P.P.I.

Professionals tend to choose children in households with a background of transgenerational care; conversely, they seem to exclude those situations in which out-of-home placements has been experienced. The more parents have established a good relationship with professionals, the more families are likely to be involved in the intervention. Lastly, households whose vulnerability is the result of risky behaviours or conditions, such as alcohol or substance dependence, detention, degraded environment, etc., seem to be more frequently present among excluded families.

Fig. 2 shows the probability distribution of the estimated propensity score separately for the treated and the control group. On average, children in the treated group have a higher probability to be involved in the programme than the others.

In situations like this one, where the overlap condition is not fully satisfied, and the number of non-treated units is not much higher than the number of treated, Kernel and Stratification matching are recommended. By re-sampling observations through matching, the comparison groups are much closer: none of the differences observed before matching is statistically significant and the problem of selection bias seems to be lessened.

Table 2. Marginal effects of the pre-treatment variables on the propensity score

Variable	Estimate
Child's and family information	
Age	0.9
Female	0.8
Foreigner family	-10.5
Number of children in the family	-2.6
Type of family	
Both biological parents	-17.8
Single parent	-6.2
Single parent and other adults	-18.8
Other types	-
Previous out-of-home placements	-15.9 *
Years duration of the care process	-0.8
Bad relationship between family and social services	-11.1 **
Vulnerabilities	
Transgenerational care	19.1**
Perturbations of family equilibrium	-17.5**
Parents' disability or psychiatric disease	-17.2*
Traumatic and/or stressing event	14.7**
At risk behaviors, conditions	-10.7**
Child abuse or witness of violence	-10.2
Child's disability or psychiatric disease	-9.0
Child neglect	7.7
Social deprivation	7.5
Economic deprivation	-6.0
Parents' psychological disease	4.9
Child's psychological disease	0.0

** {*} Difference statistically significant
at the 95% {90%} level of confidence.
Average percentage marginal effects.
Robust standard errors clustered by territorial area.

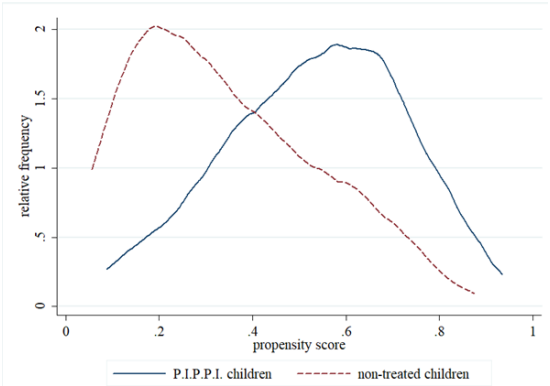

Fig. 2. Estimated probability distribution of the propensity score in the baseline treated and non-treated groups.

3.2. Estimated effects on outcomes

Table 3 reports the DID estimates of the ATE computed comparing the matched samples. The estimates refer to the average effect of the intervention in the group of the P.I.P.P.I. children on outcome variables measured at T2.

By "effect" we intend the difference at T2 between what is observed as a result of the intervention and what would have been observed in the counterfactual situation of care with standard services in mean on the P.I.P.P.I. child. The variables considered as outcomes are following.

1. The objective situation of the family at T2 in terms of greater parents' autonomy from professionals' support (dichotomous variable), defined "positive" (value=1) if the care process continues with lighter interventions or the family is no more in care because of an improvement in the situation.
2. The professionals' evaluation of the family situation at T2 (six-points Likert scores): all the sub-dimensions of the CW (17 variables); the number of protection and risk factors associated to each side of the CW – children's needs, parenting competences and environmental context (6 variables); an overall evaluation of the child's risk to be placed out-of-home.

Estimates are reported separately for type of matching and way of computation of the DID estimator. Type A estimations calculate the simple difference in the average pre-post changes; type B effects are derived from the estimation of a regression model of the pre-post differences on the treatment indicator plus other variables on which the differences in trends may depend: weighed regression for Kernel Matching and the simultaneous estimation of the conditional regressions within the blocks of the propensity score for Stratification Matching.

The results of the counterfactual analysis confirm the effectiveness of the intervention on most of the outcome variables considered. In particular, all methods agree in estimating a statistically significant impact on children's total risk of out-of-home placement (the overall risk evaluation of the *PostAssessment*) and the satisfaction of the child's developmental needs (the child side of the "Triangle"), as well as on parents' achievement of a certain degree of "autonomy" in their work with services (the actual situation of the family at T2). While results are robust for the variables regarding children and environment, the effects estimated on the parental outcomes are statistically weaker and not consistent with the estimation method applied.

By summarizing the empirical evidence, we observed the following statements.

1. All methods show a significant effect on the total risk evaluation: the overall child's risk of out-of-home placement decreases by 0.47-0.55 points on the Likert scale 1-6 thanks to the participation in the programme. In other words, if the P.I.P.P.I. child had not entered the programme, but had been instead followed with the ordinary practices of social services, he/she would have achieved an average score of total risk evaluation significantly higher at T2.

2. The effect estimated for the impact on the family's work with services at T2 is also robust and statistical significant with a high degree of confidence. On average, the situation of the family at T2 is more frequently satisfactory thanks to P.I.P.P.I.: the conclusion of the care process because the situation has improved or the continuation of the work with services characterized by lighter interventions are events significantly more likely after P.I.P.P.I.

3. Concerning the CW Questionnaire, the impact of the programme is particularly high on the side of the child's needs; only for the sub-dimension "Health & Physical Development" P.I.P.P.I. and the ordinary services do not differ significantly. Even with regard to risk and protection factors, those one related to the child's developmental needs appear to reduce and increase, respectively, as a result of the participation in the programme.

4. Only with respect to a few sub-dimensions of the family and the environmental dimensions of the CW, P.I.P.P.I. and the ordinary services differ significantly. On the family side, only the sub-dimensions "Play, encouragement & fun" and "Parents' self-realization" seem to be affected by the programme; on the environmental side, "Employment & income", "Housing" and "Relationship with school and other services" improve to a greater extent, thanks to P.I.P.P.I.

5. As far as environmental conditions are concerned, the correspondent levels of protection and risk factors recorded by professionals in the *Pre-PostAssessment* seem to be unaffected by the programme.

Conclusions

The simple pre-post comparison of the variables to which P.I.P.P.I. is targeted indicates a significant improvement in the situation of the families and their children from T0 to T2 [27]. By applying counterfactual analysis, the effectiveness of the

Table 3. ATE estimates

Variable	Kernel		Stratification	
	A	B	A	B
FAMILY-SERVICES RELATIONSHIP				
Parents' lower need from professional support	0,29 ***	0,84 ***	0,28 ***	0,34 ***
PROFESSIONAL EVALUATIONS				
CW scores				
Child's needs				
Health & Physical Development	0.19	0.20	0.16	0.14
Social Skills	0.58 **	0.69 ***	0.55 **	0.62 ***
Identity, Self Esteem & Social Presentation	0.57 **	0.61 **	0.53 *	0.53 *
Selfcare Skills	0.45 **	0.45 **	0.42 **	0.51 ***
Family & Peer Relationship	0.75 ***	0.82 ***	0.75 ***	0.73 ***
Learning	0.41 **	0.45 **	0.36 **	0.47 ***
Play & Free Time	0.75 ***	0.77 ***	0.75 ***	0.85 ***
Parental behaviour				
Basic Care	0.15	0.15	0.18	0.22
Emotional Warmth	0.30	0.33	0.34	0.38 *
Guidance & Boundaries	0.41	0.45	0.40	0.45 *
Play, Encouragement and Fun	0.65 **	0.66 **	0.65 **	0.68 **
Parents' Self-Realisation	0.74 **	0.76 **	0.70 **	0.83 ***
Environmental conditions				
Support from others	0.38	0.38	0.39	0.32
Participation in the community	0.44	0.42	0.44	0.45
Employment & Income	0.51 **	0.47 ***	0.46 **	0.57 ***
Housing	0.52 **	0.49 **	0.50 **	0.55 **
Relationship with Schools and Services	0.50 **	0.45 **	0.53 **	0.41 **
RISK factors				
Child	-0.63 **	-0.67 ***	-0.65 **	-0.69 ***
Family	-0.44 *	-0.47 *	-0.42 *	-0.51 **
Environment	-0.23	-0.22	-0.22	-0.32
PROTECTION factors				
Child	0.52 **	0.59 **	0.50 **	0.52 **
Family	0.48 *	0.56 *	0.46 *	0.53 *
Environment	0.49	0.52	0.44	0.46
Overall risk evaluation	-0.47 **	-0.58 ***	-0.41 ***	-0.55 ***

*** [**] (*) Difference statistically significant at the 99% [95%] (90%) level of confidence.
T(G-1) critical values considered, where G=12 is the number of territories.
Robust standard errors clustered by territorial area (bootstrapped for type A estimates).
Type A estimates: Simple difference of pre-post differences.
Type B Kernel estimates: Ordered Probit model for scores – regression of levels on D, T2*D and X=pre-treatment variables; Probit model for family situation – regression of the situation indicator on D and X=pre-treatment variables; values not comparable with the correspondent estimates in the other columns.
Type B Stratification estimates, 5 blocks: linear regression of the pre-post differences on D*block1, . . . , D*block5, propensity score *block1, . . . , propensity score *block5 (then the conditional ATEs are combined).

programme is confirmed on almost all outcomes. The impact seems to be particularly intense on the child's total risk of out-of-home placement and developmental needs. The effects on parents' responses to children's needs are instead dubious because of being statistically weaker and not confirmed by the application of different estimation methods. Anyway, despite the limited duration of the work with families, their overall situation at T2 improves, as parents' degree of "autonomy" from professionals' support increases. If on the one hand the risk factors decrease, on the other resources and strengths significantly improve thanks to the programme.

This empirical evidence offers indications to professionals and all persons responsible for the programme implementation on how potentially modify practices and organizational aspects to achieve better outcomes. By relating the results of the counterfactual analysis with the process data on the individual micro-plans, questions arise about professionals' attention unbalanced to child's factors and the practical difficulties to discern and leverage on strengths and resources of parents and their relationships when, respectively, evaluat-

ing and working with families [28]. Furthermore, we expect that a longer duration for the intervention, whenever professionals deem it appropriate, could reinforce results, on parents' outcomes.

Doubts however arise in the interpretation of the impact evaluation results as well. Firstly, the data collected by professionals not directly involved in the programme (in the control group) is questionable respect to both quality and meaning. Secondly, the estimated impact, relative to an average effect over the entire treated group, does not consider intersubjective variability in outcomes and in the intensity and type of heterogeneous individual interventions.

In the light of these considerations, the evaluation of P.I.P.P.I. with counterfactual approach is now moving on to the estimation of differential effects inside the programme itself. This strategy does not require to identify a control group of families not involved in the programme nor to gather additional data. Moreover, the rich amount of information already available from the research instruments provided by the evaluation plan allows to better understand what kind of families are compared in the analysis, whether they are families with certain type of micro-plans or families that have taken advantage from specific activities.

In the research described in this paper the idea was to compare the programme with the standard practice of child protection, but the border between the two types of work can be very vague in some cases. It is indeed not possible to exclude a distortive contamination of the programme on the compared standard practices of social work with families and children. Consider, for example, the high percentage of control families whose professionals were previously trained in P.I.P.P.I. (73%). The specific actions characterizing the programme are also frequently observed in the control group, particularly regarding the relationship with schools, which results activated for more than 70% of non-treated families.

References

1. Pearl J. Causality. Models, reasoning and inference. – Cambridge, UK: Cambridge University Press, 2000.
2. Winship C., Morgan S. Counterfactuals and causal inference. – Cambridge, UK: Cambridge University Press, 2007.
3. Lacharité C., Ethier L., Nolin P. Vers une théorie éco-systémique de la négligence envers les enfants. – *Bulletin de psychologie* 59(4) (2006) 381-394.
4. Dubowitz H., Pitts S.C., Litrownik A.J., Cox C.E.,Runyand D., Black M.M. Defining child neglect based on child protective services data. – *Child Abuse & Neglect* 29 (2005) 493-511.
5. Bronfenbrenner U. - The ecology of human development. Experiments by nature and design. – Cambridge: Harvard University Press, 1979.
6. Bronfenbrenner U. - Making Humans being Human. Bioecological perspectives on Human development. – London: Sage Publications, 2005.
7. Serbati S., Milani P. La tutela dei bambini. Teorie e strumenti di intervento con le famiglie vulnerabili. – Roma: Carocci, 2013.
8. Vaquero Tió E., Ius M., Paola M., Balsells Bailón M.A. Una revisión de la literatura sobre el uso de las TIC en elámbito de la intervención sociofamiliar. – In: R. Roig-Vila (ed.), Tecnología, innovación e investigación en losprocesos de enseñanza-aprendizaje. – Barcelona: Octaedro, 1920-1928, 2016.
9. Parker R., Ward H., Jackson S., Aldgate J., Wedge P. Looking after children: Assessing Outcomes in Child care. – London: HMSO, 1991.
10. Horwath J. - The Child's World: the Comprehensive Guide to Assessing Children in Need. – London: Jessica Kingsley Publisher, 2010.
11. Serbati S., Ius M., Milani P. P.I.P.P.I. Programmeme of Intervention for Prevention of Institutionalization. Capturing the evidence of an innovative programmeme of family support. – *Revista de cercetare si interventiesociala* 52 (2016) 26-50.
12. Braconnier V., Humbeeck B. L'évaluation de l'état de danger dans la famille. – In: Rapport intermédiaire de recherche. – Mons: Université de Mons-Hainaut/CERIS, 2006.
13. Milani P., Colombini S., Di Masi D., Ius M., Santello F., Serbati S., Sità C., Tuggia M., Zanon O. P.I.P.P.I. Programmema di Intervento Per la Prevenzione dell'Istituzionalizzazione. Rapporto di valutazione. Sintesi 2015-16. – *Quaderni della ricerca sociale* 39 (2017).
14. Holland P.W. - Statistics and causal inference. – *Journal of American Statistical Association* 81(396) (1986) 945-960.
15. Heckman J. Instrumental variables: A study of implicit behavioral assumptions used in making programme evaluations. – *Journal of human resources* 32(3) (1997) 441-462.
16. Rosenbaum P.R., Rubin D.B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. – *The American Statistician* 39(1) (1985) 33-38.
17. Heckman J., Ichimura H., Todd P.E. - Matching as an econometric evaluation estimator. – *Review of economic studies* 64(4) (1997) 605-654.
18. Heckman J., Robb R. Alternative Methods for Evaluating the Impact of Interventions. – In: Heckman J. and Singer B. (eds.), Longitudinal Analysis of Labor Market Data. – Cambridge, U.K.: Cambridge University Press, 1985.
19. Abadie A. - Semiparametric difference-in-differences estimators. – *Review of Economic Studies* 72 (1) (2005) 1-19.
20. Heckman, J., Robb R. - Alternative Methods for Evaluating the Impact of Interventions. – In: J. Heckman and Singer B. (eds.), Longitudinal Analysis of Labor Market Data. – Cambridge, U.K.: Cambridge University Press, 1984.

21. Rosenbaum P.R., Rubin D.B. The central role of the propensity score in observational studies for causal effects. – *Biometrika* 70(1) (1983) 41-55.
22. Imbens G.W. - The role of the propensity score in estimating dose-response functions. – *Biometrika* 87(3) (2000) 706-710.
23. Becker S., Ichino A. Estimation of average treatment effects based on propensity scores. – *The Stata Journal* 2(4) (2002) 358-377.
24. Imbens G.W., Wooldridge J. M. - Recent developments in the econometrics of programme evaluation. – *Journal of economic literature* 47(1) (2009) 5-86.
25. Leuven E. and Sianesi B. Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, andcovariate imbalance testing. – [http://ideas.repec.org/c/boc/bocode/s432001.html], 2003.
26. Serbati S., Santello F., Colombini S., Milani P. Challenges for the evaluation of the P.I.P.P.I. Programmeme of Intervention for Prevention of Institutionalisation: between participative and experimental pathways. – *Interdisciplinary Journal of Family Studies* 21(2) (2016) 1-24.
27. Santello F., Colombini S., Ius M., Milani P. P.I.P.P.I.: What has changed? How and why? The empirical evidence. – *Rivista Italiana di Educazione Familiare* 2 (2017) 111-136.
28. Serbati S., Ius M., Milani P. P.I.P.P.I. Programmeme of Intervention for Prevention of Institutionalization. Capturing the evidence of an innovative programmeme of family support. – *Revista de cercetare si interventie sociala* 52 (2016) 26-50.