Challenges for the Evaluation of the P.I.P.P.I. - Programme of Intervention for Prevention of Institutionalisation: between Partecipative and Experimental Pathways

Sara Serbati[a], Francesca Santello[a], Sara Colombini[b], Paola Milani[a]

[a]*Department of Philosophy, Sociology, Education and Applied Psychology, University of Padua (Italy)*
[b]*Department of Economy, University of Modena (Italy)*

_____

Abstract. Evaluation is constantly requested by governments and decision-makers, to prove that social policies and actions undertaken are effective in responding to problems. Also, programmes contrasting child neglect are involved in such request to guarantee that children enjoy their childhood and ensure access to quality service. This paper focuses on an Italian evaluation experience of such a programme, thanks to the collaboration between University of Padua and Italian Ministry of Welfare. It is called *P.I.P.P.I. - Programme of Intervention for Prevention of Institutionalisation.* The paper questioned and challenged the experimental designs normally used for these evaluation purposes, highlighting how knowledge of effective treatments is far from the practices delivered. The study purposes an innovative evaluation path, intertwining the participative evaluation where the professionals build their own knowledge through an evaluation *in the field,* with the choice of matching as (quasi) *experimental* evaluation, responding to the Government's request of effective investments.

Correspondence concerning this article should be addressed to Sara Serbati, Department of Philosophy, Sociology, Education and Applied Psychology, University of Padua (Italy), Via Beato Pellegrino 28, 35137 Padova (Italy) email: sara.serbati@unipd.it

Introduction

After the adoption of the Convention on the Rights of the Child (CRC, 1989) an international movement was started, aiming to develop measures and policies reflecting the child's best interests that went hand in hand with another transformation seeking to ensure access to quality services. Also, guided by many pieces of legislation (Council of Europe, 2006; European Commission, 2013), the European Member States have designed and implemented policies addressing child poverty and social exclusion and promoting children's well-being. For governments and decision makers *evaluation* becomes a central feature in order to obtain indications for their decisions.

This paper focuses on an Italian evaluation experience of a programme contrasting child neglect, widespread in the entire State. Run through the collaboration between University of Padua and Italian Ministry of Welfare, it is called *P.I.P.P.I.*, that stands for *Programme of Intervention for Prevention of Institutionalisation* and it is inspired by the fictional character Pippi Longstocking, a creative and amazingly resilient girl known all over the world. Here the meaning of evaluation is assumed in a twofold way: evaluation is the process undertaken by professionals with families for assessing-planning-acting and evaluation is to study the achieved results in order to give indications for decision makers. In the first meaning professionals are the main actors for an evaluation *in the field*, in the second meaning an *experimental* evaluation is undertaken. The paper aims to explore how these two meanings of the evaluation processes live together in the P.I.P.P.I.. With this aim, the study will purpose:

- to question the literature about evaluation of international programmes contrasting child neglect in order to highlight shortcomings and challenges;
- to describe aims and processes of the evaluation *in the field*;
- to describe aims and processes of the *experimental* evaluation;
- to describe processes and results in the selection of treated and non-treated families.

The results will be discussed in order to consider if the evaluation undertaken by the P.I.P.P.I. could be a strategy to reach in a reliable way an answer for responding to the two evaluation meanings considered.

The National P.I.P.P.I. Programme

The P.I.P.P.I. aims to prevent out-of-home placement and to respond to problems connected to child neglect in view of all children's

right to quality care. The first and the second stage of the P.I.P.P.I. programme were carried out on two-year period each (2011-2012; 2012-2013) in 10 Italian cities. Moreover, in 2014-2015 and 2015-2016 the first and the second steps of scaling up have begun, where 82 new territories and approximately 1000 children are involved. Now they are running the 5th (2016-2017) and 6th (2017-2018) steps with 68 new territories. In the P.I.P.P.I., 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 current scientific knowledge (Dubowitz et al., 2005; Lacharité, Ethier, & Nolin, 2006). Considered as a complex social problem, child neglect should not be defined by focusing solely on the description of parental behaviour. In accordance with the bio-ecology of human development (Bronfenbrenner, 1979; 2005), the P.I.P.P.I. aims to respond to children's needs with a collective action: it is not only a matter of helping parents, but of promoting the responsibility of public institutions (i.e. social agencies) to develop relationships for families in their communities. For this reason, the programme provides for each family the simultaneous activation of four "specific activities" involving families' different ecosystem levels. The "specific activities" are integrated in a shared care plan, co-constructed for and with each child and family, following the evaluation *in the field*. They are:

(1) Home-care intervention: this in-home activity aims at addressing relationships problems and modifying behaviours. Practitioners meet with the families approximately twice a week for a minimum of at least four hours a week.

(2) Parent and Child Groups: parents and children are involved in group activities with other families, fostering reflective practice, encouraging exchange and interaction between participants. Meetings are weekly or bi-weekly and usually last approximately three hours.

(3) Natural family helpers: each family is provided with a support family or a family helper whose aim is to offer support in concrete aspects of daily life.

(4) Cooperation between schools/families, social and health services: teachers, with the other professionals and families, outline actions (both individualized and involving the entire class) that will favor a positive school environment where children can learn social and emotional competences.

3

Challanges and Opportunities for the Evaluation of the P.I.P.P.I. Programme

Over the last two decades, government policies regarding programmes contrasting child maltreatment and child neglect have increased throughout Europe and also at an international level. Well known and outstanding are initiatives such as *Sure Start* in the UK and *Homebuilders* in the USA. But many other initiatives are widespread over the world. Some examples are the *Triple P* (Sanders, Markie-Dadds, & Turner, 2003), *Zeppeling* (Neuhauser et al., 2015), *SafeCare* (Gershater-Molko, Lutzker, & Wesch, 2002), *MST* (Huey, Henggeler, Brondino, & Pickrel, 2000). These programmes share an investment in defining treatment manuals in order to guide practitioners' actions for realizing effective interventions. And in order to assure effective intervention, evaluation is a central question. The gold standard is recognized in RCTs, the Randomized Controlled Trials, which are used to determine the effectiveness of the interventions. Since the Nineties, there has been a growing call to use RCTs, the 'gold standard' able to establish what is called the Evidence Based Practice (EBP). Evidence is 'what works' to obtain best outcomes. In parallel, the social work field increasingly requires the use of Evidence-Based Knowledge, where practitioners use the best available research and knowledge that guide their practice (Mullen, 2014; 2016). Undoubtedly, there is much of an emancipatory spirit here which aims to make the practices implemented more appropriate to respond to the people's needs (Lorenz, 2012).

However, many meta-evaluation and review studies (e.g. Bunting, 2004; Chamberland & Dufour, 2004; Dagenais, Begin, Bouchard, & Forti, 2004; Euser, Alink, Stoltenborgh, & Bakermanskranenburg, 2015; Heneghan, Horwitz, Leventhal, 1996; Leveille & Chamberland, 2010; Turney & Taylor, 2014; Knorth, Harder, Zandberg, & Kendrick, 2008; MacMillan et al., 2009) highlight in their conclusions the lack of scientific rigour: evaluative studies are often based on quasi-experimental designs of a pre/post-intervention type without comparison groups and they use a small sample size not allowing for generalization. Also, they suggest a modest positive change, with a combination of high hopes and much more modest accomplishments and/or disappointments (Wandersman, 2009). The lack of information available on the implementation of interventions constitutes another important limitation.

Moreover, several major reports and many authors highlight that a gap exists between the knowledge produced by evaluation and daily practices currently delivered (Biesta, 2007; Clapton et al., 2008; Fixsen,

4

Naoom, Blasé, Friedman, & Wallace, 2005; Gonzales, Handley, Ackerman, & O'Sullivan, 2012; Hammerseley, 2013; Rubin, 2014; White & Dudley-Brown, 2012). It seems that something exists that hinders the use of effective assumptions inside the daily practice. Effectiveness, even if demonstrated, seems to remain inside the time of implementation. In order to understand this gap between scientific knowledge and the real practice, it is possible to consider three epistemological shortcomings.

A first shortcoming concerns the claim of Evidence-Based tradition to establish scientifically (meaning 'scientific' as a synonym of the experimental model proper of the natural sciences) effective procedures, defined as 'what works'. These procedures, once identified, are not questionable and should be applied as defined, because they passed the test of 'science'. Professionals' and clients' freedom of choice, thought and judgment does not appear a priority. In this regard the question from W. Lorenz (2012) appears correct: "was it possible (and appropriate) to transfer the scientific approach that proved successful in relation to the nature to the social sphere without robbing its members of the very freedom gained by the advance of reason and subjecting their behaviour instead to laws and regularities over which they had little control?"

A second shortcoming, linked to the previous one, concerns the will of Evidence-Based tradition to standardized processes. Indeed, the standardization of processes is not appropriate in a field such as social work. Not all processes can be validly standardized. Standard process can be managed directly through procedural or technological means, but non-routine processes are best managed by indirect means, such as competence, professional values, visions and missions, through culture in other words (Featherstone, White, & Morris, 2014; Lorenz, 2016).

A third shortcoming is about the focus on 'what works' for people to address life challenges and enhance wellbeing. This means that it has a very high confidence in the possibility of intervening with individuals to recover abilities and skills that allow them to exercise their choices. But the nature of social work is relational: child development and families' empowerment do not happen in a vacuum, and child neglect or maltreatment should not be defined by focusing solely on the description of parental behaviour. "Problems are not a matter of the child or of the parent; they are ever shared, thus they show a disease situated not in the child or in the parent, but in the relationship between them. It is the "interactive space" defined by the relationship parent-child in the familiar, social, cultural, historical network" (Dumas, 2007).

Trying to respond to these shortcomings, P.I.P.P.I. confirms the central importance of evaluation for demonstrating effectiveness, but also

tries to find different ways to respond to it. From the Evidence-Based tradition, P.I.P.P.I. maintains two features, translating the first for the evaluation *in the field*, and the second for the *experimental* evaluation.

Evaluation *in the Field*: the Participative and Transformative Evaluation

In P.I.P.P.I. a pre-/ post design is employed to compare Time 0 families' situation at the intake of the intervention and Time 1, at the conclusion. Innovatively and responding to the three shortcomings, P.I.P.P.I. intends this design following the principles of participatory-research. Negotiation is the main feature of participatory-research (Guba & Lincoln, 1989), which allows the questioning of practices, rules, habits, routines, etc. The expected outcome of negotiation is the transformation, in order to achieve better conditions. Thus, research instruments are intended as negotiation instruments, used directly by participants, in order to improve and transform their practices (Serbati, Pivetti, & Gioga, 2015).

Using the research instruments participants create learning and negotiation contexts where families and professionals together could try to experiment a new balance or make new decisions enabling them to improve the children's development. Dewey's thoughts (1933) inspired this work of thinking and reasoning, in order to negotiate between all the actors the meanings of what it is important for the child's wellbeing. And research instruments are viewed as a means to reach this purpose. Data collection procedures of the programme, realized a path called participative and transformative evaluation (P.T.E. – Serbati & Milani, 2013). Data were collected directly by professionals working with families, considering all of them as co-researchers. P.T.E. used measures and data in a double function: (1) of accountability, in order to collect information about the quality of the interventions; (2) of negotiation (Guba & Lincoln, 1989) between all the people who are important to the child's development. People were expected to work together around data and measures. In this way, it is introduced a performance-based culture within the Italian Child Protection sector. P.I.P.P.I. not requires to professionals to realise actions that are proved scientifically and described in a manual. In this case, the theoretical discussion of professional practice would remain far from real practice, which would have no internal legitimacy with the reasons that explain the action (Geertz, 1973, 1997; Soulet, 2014). P.T.E. intends professionals as co-researchers that work in examining and challenging the actions outlined by the programme. The features that characterize the quality of the interventions are built in an internal way, by the professionals themselves

6

through reasoning and dialogue capacities. It means that not only single persons are important, but also groups and relationships between participants.

Experimental Evaluation: the Matching Method

The choice of the P.T.E. brings some limitations to the opportunity in using RCTs for evaluating the programme. The randomization required by the RCT designs invests in individuals. To a randomized selection of the children that entered in the programme would follow a randomized selection of professionals that build the teams responsible for the family's care process. It would collide with some organizational and ethic requirements:

- the programme makes an investment in professionals' trainings (three bi-annual trainings according to the professionals' role in the implementation). It requires a controlled number of participants;
- the programme makes an investment in participatory and negotiation processes. It required participants with high internal motivation to invest time and efforts in these processes;
- the programme makes an investment in developing and valorize participants' decision making in team. It requires involving them since the families' selection stage, through team work.

The reasons described above do not permit to use a RCT design. However, the P.I.P.P.I. wants to regard a second feature from the Evidence-Based tradition: the comparison with a group that does not participate to the programme, families and practitioners that follow the Child Protection mainstream activities. Excluding RCTs, the P.I.P.P.I. is now taking the choice of a non experimental method: the *matching*.

In quasi-experimental settings, where there is not a random assignment of individuals to the intervention and non-treated (control) group, to the extent the second group resembles – or is statistically adjusted to resemble – the treated one on relevant characteristics and experiences, what results from their comparison can offer a credible estimate of the programme effect. When a non-random selection process intervenes in the choice of which units to include in the control group, the latter cannot be considered equivalent to the treated one: the two groups might differ in the variables potentially affected by the programme regardless the exposure to it. This a priori difference, known in literature as "selection bias" (Heckman, 1997), prevents us to directly compare treated and non-treated units: the outcomes observed for the non-treated families are not a good approximation of the

7

same outcomes we had observed for the participants in the absence of the programme (counterfactual situation; Holland, 1986).

In non-experimental settings a feasible procedure to statistically adjust for systematic differences between treated and non-treated groups is *Matching* (Rosenbaum & Rubin, 1983). To remove selection bias, matching lines up comparison units according to a sufficient number of observable pre-intervention characteristics: any variable which plays a role in the selection process and also affects the outcome. The researcher therefore assumes to re-establish experimental conditions by comparing units with the same observable characteristics. Clearly the credibility of this selection on observables assumption (Heckman & Robb, 1985) implies a deep understanding of the determinants of the assignment process.

In practice matching involves pairing treatment and control units whose values of their observable characteristics are as similar as possible. While the procedure is straightforward with a small number of covariates, finding matches can become very difficult when there are many variables to control for. To deal with the dimensionality problem, *Propensity Score Matching* can offer a solution. It implements matches by means of the Propensity Score, a single-index variable which summarizes all information contained in the multidimensional vector of pre-treatment characteristics X. The Propensity Score, as defined by Rosenbaum and Rubin 1983, is the conditional probability of receiving the treatment given the X:

$$p(X) \equiv Pr\{T=1|X\},$$

where $T=\{0,1\}$ is the indicator of exposure to treatment. Any standard probability model for binary dependent variable – usually parametric Probit or Logit regression models – can be used to estimate the Propensity Score. Only the need to obtain an estimate of the Propensity Score conditionally on which all covariates have approximately identical value is important (Balancing property, Imbens, 2000).

Propensity Score matching methods build a representative control group by pairing each participant with members of the baseline "biased" non-treated group on the base of the score p(X). The comparison group for each treated unit is chosen with a pre-defined criteria of proximity between the observed propensity scores of treated and non-treated. Having defined a neighborhood around the observed Propensity Score of each treated observation within to select comparison units, the next step is the choice of appropriate weights to associate the members of the comparison set. Statistical literature provides several alternatives (see Leuven & Sianesi, 2003; Becker & Ichino, 2002 for a detailed review). The most commonly

applied techniques are Nearest Neighbor Matching, Radius Matching, Kernel Matching and Stratification Matching. While Nearest Neighbor Matching assigns weight one to the nearest comparison unit and zero to the others, Radius Matching considers all the members within the neighborhood, giving them the same weight. Kernel matching gives a weight inversely proportional to the distance between Propensity Scores. Finally Stratification method, which is a bit different method, divides the support of Propensity Score in blocks and within each block, where the Propensity Score can be assumed approximately constant, compares treated and non treated units.

In the present evaluation study all these methods were tested to find the appropriate comparison group for the families in P.I.P.P.I. As we will see, Kernel and Stratification Matching turned out to be preferable in order to improve on balancing of observables and also obtain a comparison sample representative of the entire treated group of families. Each method was indeed evaluated in respect to its capability to obtain a final "unbiased" control group whose observable characteristics are on average statistically equal to the ones observed for the treated (internal validity); secondly, we tried to employ all information contained in each baseline group.

The Instruments

P.I.P.P.I. asks professionals to use a wide range of instruments (Serbati, Ius, & Milani, 2016), but in this study they will be considered the *Questionnaire for Socio-Demographic Information* and the *PreAssessment*. Both of them are completed for each child entered in the study (treated and non-treated groups). The Questionnaire is built as a checklist. *Pre Assessment,* created for the P.I.P.P.I. implementation and inspired to the works of Braconnier and Humbeeck (2006), is used both in the P.T.E. path and for the *matching method* evaluation. It consists of 5 sections:
  A. Vulnerability conditions and social interventions accessed by the family and by the children. It is built as a checklist;
  B. Family history. It is built as a free text to be composed starting from 20 guide-questions, which are described in a handbook;
  C. Risk and Protection Factors. It is built as a six-point Lickert-scale (from 1=few factors to 6=many factors) regarding three areas borrowed by the British *Framework for the Assessment of Children in Need and their Families (FACNF*-DoH, 2000): Children's Needs, Parenting Competences, Family and Environmental Factors, which are described in a handbook.

9

D. Quality of the family and social services' relationship. It is built as a six-point Lickert-scale (from 1=difficult to 6=comfortable);

E. Overall evaluation. It is built as a six-point Lickert-scale (from 1=absence of risk to 6=high risk), to be completed considering the previous sections.

PreAssessment is used by professionals before the beginning of the implementation in order to include the treated and non-treated families in the study. The same instrument but in the version of *PostAssessment* will be used by professionals at the end of the study, after 18 months. *PreAssessment* is used as a guide for a shared decision-making following the path of the P.T.E.: professionals' team makes comparison and negotiation between their different points of views in order to find an agreement about the opportunity for a family to access the programme.

After the completion of the professionals' team, sections A, C, D, E are used for the *matching*.

Inclusion of the Participanting Families

The treated and non-treated groups of families are selected by practitioners following some criteria given by the programme: families with children aged 0-11 y.o., families that experienced child neglect, families with previous or current access to the mainstream activities without success, families with good relationship with professionals. To facilitate the process of assessment, professionals use *PreAssessment* in team for a wide range of families. Since *PreAssessment* results, the teams decide which families do not access to the study (those scored 1 or 6 in section E) and which access to the programme (those scored 2 to 5 in section E). During the fourth implementation of the P.I.P.P.I. (2014-2015) 868 *PreAssessment* were completed for 47 territories, but only 446 families accessed to the programme. 10 of these 47 territories were selected for taking part in the *matching method* evaluation. The 10 territories were selected on the basis of the following criteria:
- territories already participating to a previous implementation;
- territories where all the activities of the P.I.P.P.I. are integrated in the mainstream services;
- territories with declared availability to collaborate in scientific actions of the P.I.P.P.I.

The professionals' teams of these territories selected 15 families to compose the non-treated group that will be used for the *matching method* evaluation (Table 1). The selection process is the same of the treated group

10

(using *PreAssessment)*, thus assuring a good baseline for the non-treated group of families to be used for the *matching*.

Table 1

Families and children participating to matching method evaluation

Region	Territory	Treated		Non-treated	
		Families	**Children**	**Families**	**Children**
Emilia Romagna	Correggio-Guastalla	10	10	-	-
	Ferrara	9	9	-	-
	Modena	-	-	15	15
	Reggio Emilia	-	-	15	15
Lombardia	Bergamo	10	12	14	14
	Mantova	9	10	15	15
	Milano	10	12	15	15
Piemonte	Alessandria	9	11	15	15
	Fossano	10	10	15	15
	Torino	19	23	15	18
Veneto	Venezia	9	9	15	15
	Vicenza	12	12	15	15
Total		107	118	149	152

Correggio-Guastalla and Ferrara participated in the fourth implementation but do not give their availability for the selection of non-treated group. For it, they were substitute by similar territories (Modena and Reggio Emilia) of the same Region (Emilia-Romagna) that participated only to the third implementation.

Empirical Results

The rich amount of information at our disposal from the *Questionnaire for Socio-Demographic Information* and the *PreAssessment* makes us confident about the credibility of the matching method in finding an appropriate comparison group for the sample of families in P.I.P.P.I, i.e. the selection on observables hypothesis can be assumed with a certain degree of confidence. Table 2 reports some descriptive statistics for children of the two baseline comparison groups, where the significant statistically differences are highlighted.

Table 2

Pre-matching differences between treated and non-treated children (part1)

	Treated children	Non-treated children	Difference
Child's socio-demographic information			
Age (mean)	8.4	8.0	0.4
Female (%)	35.6	36.8	-1.2
Foreigner (%)	19.1	28.9	-9.8**
Previous out-of-home placements (%)	9.3	16.4	-7.1**
Child's family information			
Number of family components (mean)	3.6	4.0	-0.4
Number of children in the family (mean)	1.9	2.1	-0.1
Children per type of family (%)			
Both biological parents	33.1	44.1	-11.0*
Single parent	36.4	30.9	5.5
Single parent and other adults	6.8	9.2	-2.4
Stepfamily	17.8	13.2	4.6
Other types	5.9	2.6	3.3
Num. of facilities already functioning (mean)	3.2	3.7	-0.5
Years duration of the care process (mean)	4.0	4.4	-0.3
PreAssessment, section C scores (percentage of children)			
High risk (points 5-6)			
Child's developmental needs	27.1	21.1	6.1
Parenting capacity	33.9	36.8	-2.9
Family and environmental factors	11.0	12.5	-1.5
Low protection (points 1-2)			
Child's developmental needs	31.4	24.3	7.0*
Parenting capacity	33.1	40.8	-7.7
Family and environmental factors	29.7	24.3	5.3
Family-social services relationship			
Not good (points 1-3)	32.2	50.7	-18.5**
Overall risk evaluation			
High (points 5-6)	20.3	23.7	-3.3

** (*) Difference statistically significant at the 95 (90)% level of confidence.

(part 2)

	Treated children	Non-treated children	Difference
Vulnerabilities (percentage of children)			
Economic deprivation			
Economic/working condition	61.9	66.4	-4.6
Low level of parents' education	48.3	44.1	4.2
House	33.1	43.4	-10.4*
Poverty	22.0	25.0	-3.0
Traumatic and/or stressing event	52.5	33.6	19.0**
Perturbations of family equilibrium			
Parents' conflict	42.4	49.3	-7.0
Absence of one/both parents	39.8	32.2	7.6
Stepfamily	19.5	20.4	-0.9
Adoption	0.8	0.0	0.8
Child's problem			
Psychological disease	41.5	32.2	9.3
Disability	16.9	17.1	-0.2
Psychiatric disease	2.5	7.9	-5.4*
Parents' problem			
Psychological disease	39.0	30.9	8.1
Disability	7.6	16.4	-8.8**
Psychiatric disease	8.5	15.8	-7.3**
Other risk factors			
Child neglect	33.9	28.9	5.0
Social deprivation	29.7	30.9	-1.3
Immigration	26.3	25.0	1.3
Deviant/at risk behavior	22.0	29.6	-7.6
Family under care for generations	22.0	13.8	8.2**
Child witness of violence	16.9	15.8	1.2
Substance or alcohol dependence	16.1	13.2	2.9
Bad neighborhoods	10.2	5.9	4.2
Detention	5.9	8.6	-2.6
Physical/emotional abuse	5.1	6.6	-1.5
Early school leaving	4.2	9.9	-5.6*
Sexual abuse	3.4	5.9	-2.5

** (*) Difference statistically significant at the 95 (90)% level of confidence.
Robust standard errors clustered by territorial area, computed using a bootstrap with 500 replications.

     The relevant differences observed for some variables confirm the presence of selection bias, which prevents a direct comparison between the baseline groups. Among non-treated families there are more foreigners and

children who live with both natural parents. The percentage of children with previous out-of-home placements is higher in the non-treated group and also the presence of psychiatric diseases for parents. In the meantime, the treated group presents a high frequency of traumatic or stressing events.

Child Protection previous intervention is here more recurrent. Among treated families, professionals recognize lower levels of protection factors for the Child's Needs and higher levels of risk. They also signal better relationships with social services.

The significant differences between the two baseline comparison groups make necessary to implement matching, as previously described. Table 3 reports the results obtained from the estimation of the Logit model. For each variable is reported the marginal effect on the probability to participate in the programme.

Results partly confirm the empirical evidence from the simple comparison of the baseline groups and, in general, refer to aspects which reduce the propensity score. For example, children who experienced previous out-of-home placements have a probability of about 20 percentage points lower to be included in the programme than children without this experience. We estimate marginal effects of approximately the same magnitude and sign when parents' problems of health become a vulnerability condition or professionals recognize parents' conflict and in general alterations in family composition as risk factors. Also the presence of both natural parents seems to reduce the propensity score by 20 percent.

Slightly lower, but statistically significant yet, are the coefficients associated to the number of facilities and services already functioning: much higher is the number (5 or more), lower is the participation probability (-12,5%). Other variables affecting negatively the access to the programme are: being foreigner (-12%) and poor quality relationships between family and social services (-14%). The only variable with a positive effect seems to be the presence of traumatic or stressing situations, which increase the participation probability by about 16 percentage points.

14

Table 3

Logit estimates of the Propensity score: marginal effects of the explicative variables

Age		Economic deprivation	
0-5 years	-5.1%	Yes	-5.3%
6-10 years	-4.6%	No (baseline)	
11 years or more (baseline)		Perturbations of family equilibrium (a)	
Gender		Yes	-20.1%**
Female	-0.5%	No (baseline)	
Male (baseline)		Traumatic and/or stressing event	
Nationality		Yes	16.2%**
Foreigner	-11.6%*	No (baseline)	
Italian (baseline)		Parents' or child's psychological diseases	
Previous out-of-home placements		Yes	2.8%
Yes	-20.6%**	No (baseline)	
No (baseline)		Child's disability or psychiatric diseases	
Years duration of the care process		Yes	-5.8%
0-1 years	-2.4%	No (baseline)	
2-5 years	0.1%	Parents' disability or psychiatric diseases	
6 years or more (baseline)		Yes	-19.7%*
Number of children in the family		No (baseline)	
1	1.0%	Child abuse or witness of violence	
2	4.6%	Yes	-6.7%
3 or more (baseline)		No (baseline)	
Type of family		Child neglect	
Both natural parents	-21.7%**	Yes	8.0%
Single parent	-6.6%	No (baseline)	
Stepfamily	-3.3%	Social deprivation or immigration	
Other types (baseline)		Yes	7.3%
Number of facilities already functioning		No (baseline)	
5 or more	-12.5%*	Other factors of risk	
3 or 4 (baseline)		Yes	5.4%
1 or 2	-0.9%	No (baseline)	
Family-social services relationship			
Bad	-14.2%**		
Good (baseline)			

** (*) Difference statistically significant at the 95 (90)% level of confidence.
Average marginal effects. Robust standard errors clustered by territorial area, computed using a bootstrap with 500 replications.
(a) Parents' conflict or absence, problematic adoption or family re-composition.

15

Figure 1 represents the probability distribution of the estimated Propensity Score separately for treated and controls, i.e. the fraction of children observed for each value of the propensity score, in the common support. Of course, children in the treated group have on average a higher probability to be involved in the programme than the others: while for high values of the Propensity Score there are very few non-treated units comparable to the treated ones, when the Propensity Score is low there are much more control than treated units.

Figure 1

Estimated probability distribution of propensity score in baseline treated and non-treated group
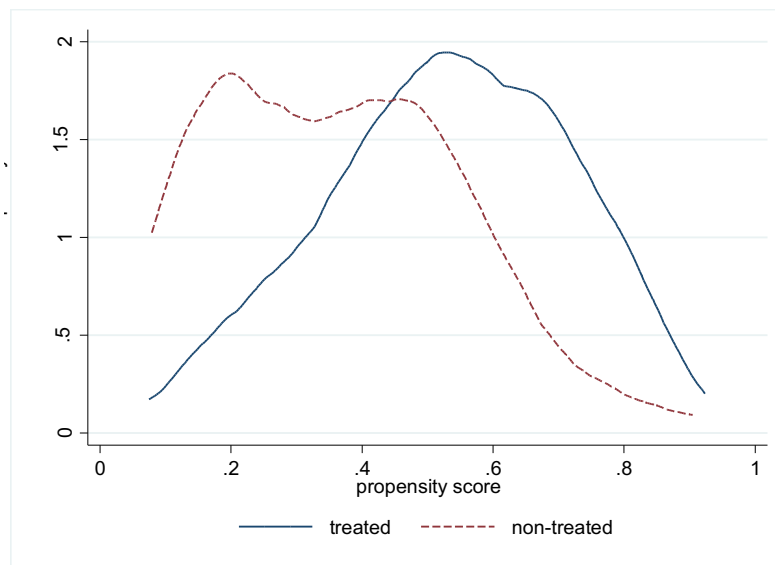
Kernel density estimation. Since the common support option has been selected, the analysis discards 8 observations.

The choice of the most appropriate matching method becomes crucial when the overlap between the treatment and the comparison groups in term of the Propensity Score is not good and the number of non-treated is not much higher than the number of treated. In these situations Nearest and Radius Matching are not recommended because of the poor quality of matches where the common support condition is badly satisfied; a lot of

treated units do not find a match choosing a small size for the neighbor as well.

In order to exploit the information of all non-treated families and also obtain a comparison group for the whole treated sample, we focused on Kernel and Stratification Matching. Both the procedures succeed in balancing all the observed covariates. For brevity, we show results only for Kernel Matching. Table 4 reports the differences between treated and non-treated children after matching. The mean value of each variable is compared between the matched samples.

Table 4

Post-matching differences between treated and non-treated children

	Treated children	Non-treated children	Difference
Age (mean)	8.3	8.1	0.2
Females (ratio)	0.35	0.34	0.01
Foreigners (ratio)	0.19	0.21	-0.02
Previous out-of-home placements (ratio)	0.09	0.09	0.00
Years duration of the care process (mean)	3.8	3.7	0.1
Number of children in the family (mean)	1.9	1.9	0.0
Children per type of family (ratio)			
Both biological parents	0.35	0.36	-0.02
Single parent	0.34	0.33	0.01
Stepfamily	0.19	0.16	0.02
Other type of family	0.13	0.15	-0.01
Number of facilities already functioning (mean)	3.2	3.3	-0.1
Poor family-services relationship (ratio)	0.33	0.33	-0.01
Family vulnerabilities (ratio)			
Economic deprivation	0.79	0.76	0.02
Perturbations of family equilibrium	0.70	0.69	0.01
Parents' or child's psychological disease	0.59	0.60	-0.01
Traumatic and/or stressing event	0.51	0.51	0.00
Social deprivation or immigration	0.42	0.43	-0.01
Child neglect	0.32	0.27	0.05
Child abuse or witness of violence	0.22	0.20	0.02
Child's disability or psychiatric disease	0.20	0.18	0.02
Parents' disability or psychiatric disease	0.13	0.14	-0.01
Other factors of risk	0.50	0.46	0.04

No difference statistically significant. Robust standard errors clustered by territorial area, computed using a bootstrap with 500 replications.

After matching, the non-treated and the treated group are much closer: none of the differences is statistically significant and the problem of selection bias seems to be solved. Reasonably, it is now possible to make a statistical comparison between the two groups using data already collected in the P.T.E. for the pre-post design. This evaluation is now in course.

Discussion

P.I.P.P.I. programme have defined the RCTs as not reliable for its aims. Random selection of the RCTs hinders with the P.I.P.P.I. purpose to engage professionals in building a performance-based culture where they will be able by themselves:
- to focus on results in order to determine the quality of social interventions;
- to focus on possibilities of change of parental practices, family and environment;
- to bear in mind the points of view of all people who are important to the child's development.

This path asks to intend evaluation *in the field*, with professionals engaged as the main actors in the process of evaluation, following the statements of the P.T.E.

However, the logic of the *experimental* evaluation, based on the difference between what was achieved with the intervention and what would happen without it, is assumed by P.I.P.P.I.. Governments and decision makers, such as the Italian Ministry of Welfare, need indications to help them with their decisions and make a constant request for *experimental* evaluation, to prove that social policies and actions undertaken are effective in responding to problems. The request for evaluation is well-motivated by the need to justify economic and social investments. The P.I.P.P.I. choice of quasi-experimental design with the *matching method* evaluation has revealed successful in responding to this request. It consents to select treated and non-treated groups without compromising the professionals' evaluation practices requested by the P.T.E. in the field. It consents to respect the need of the P.T.E. to work not only with single professionals but with professionals' relationships working in teams, well motivated to experiment the programme proposals and available to participate in a two-year training activity.

Moreover, even abandoning the aims of generalization of the RCT, the matching consents to statistically prove the effectiveness of the P.I.P.P.I. programme, thus realizing a (quasi) *experimental* evaluation, besides the participative path realized by evaluation *in the field*.
More, the results of Propensity Score open to an initial comprehension of professionals' choices in families' selection processes. There are variables considered less by professionals for accessing families to the P.I.P.P.I..

They are: the presence of both parents, previous children's experience of out-of-home placement, parents' psychiatric diseases or disability, bad relationship with services, the use of a high number of social facilities and activities and immigration. These features seem to respect the programme criteria to include family with good relationship with services or families with previous or current access to the other interventions with no success (i.e. foster or residential care, psychiatric treatment, services for foreigners). On the other hand it seems that professionals prefer to use the P.I.P.P.I. when the family's story presents traumatic or stressing events, or when there are situations of child neglect and isolation. Propensity Score indicates that treated and non-treated groups are homogeneous for the child neglect variable. It respects the request to include in the two groups similar families' situations of child neglect. However, following Barudy's (2004) statements, child neglect could be generated by stressing situations that hinder the possibility of good parenting practices. And it seems that these situations have more probability to be included in the P.I.P.P.I. Even if the data request an in-depth study, these preliminary results of Propensity Score confirm the professionals' competence in making a good selection of families for the P.I.P.P.I., respectful of the requested criteria.

Conclusion

In the social field *experimental* evaluation designs demonstrate some contradictions with the need to change practices: experimental designs happen far from real practice and by themselves are unable to modify the reasons that build the professional actions.
At the same time participative evaluation *in the field*, where professionals are the evaluators of their work, hinders the possibilities to realize an experimental evaluation.
In an innovative way, the P.I.P.P.I. tries to assure the coexistence of these two features exploring a (quasi) *experimental* evaluation path, the matching, which does not hinder with the participative one. By using the same instruments of the Participative and Transformative Evaluation, the *matching* assures the possibility to fund the evaluation of the programme

on experimental assumptions, as requested by the decision makers, and meantime it respects the central importance given to professionals' participation in the construction of the evaluation results. The relevance given by the P.I.P.P.I. to competence, professional values, and visions of participants is respected.

Notes

Paper presented at the CIRF Conference "Dalla famiglia alle famiglie. Compiti di sviluppo e specificità relazionali. Nuovi approcci di intervento e di ricerca." [From the Family to the Families. Development tasks and relational specificities. New interventions and research approaches.], Padua, November 26th 2016.

References

Barudy, J., & Marquebreucq, A. P. (2005). *Les enfants des mères résilientes*. Paris: Solal Editeurs.

Becker, S., & Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal, 2*(4), 358-377.

Biesta, G. (2007). Why 'what works' won't work: evidence-based practice and the democratic deficit in educational research. *Educational Theory, 57*(1), 1-22. doi: 10.1111/j.1741-5446.2006.00241.x

Braconnier, V., & Humbeeck, B. (2006). *L'évaluation de l'état de danger dans la famille*. Rapport intermédiaire de recherche. Mons: Université de Mons-Hainaut/ CERIS.

Bronfenbrenner, U. (1979). *The ecology of human development. Experiments by nature and design*. Cambridge: Harvard University Press.

Bronfenbrenner, U. (2005). *Making Humans being Human. Bioecological perspectives on Human Development*. London: Sage Publications.

Bunting, L. (2004). Parenting Programmes: The Best Available Evidence. *Child Care in Practice, 10*(4), 327-343. http://dx.doi.org/10.1080/1357527042000285510

Chamberland, C., & Dufour, S. (2004). The effectiveness of selected interventions for previous maltreatment: enhancing the well-being of children who live at home. *Child and Family Social Work, 9*, 39-56. doi: 10.1111/j.1365-2206.2004.00302.x

20

Clapton, G., Cree, V. E., Allan, M., Edwards, R., Forbes, R., Irwin, M., MacGregor, C., Paterson, W., & Perry, R. (2008). Ideas in Action Thinking ''Outside the Box'': A New Approach to Integration of Learning for Practice. *Social Work Education*, 27(3), 334-340. doi: 10.1080/02615470701379826

Council of Europe, (2006). *Policy to support positive parenting*, Commission Recommendation of 13.12.2006, Brussels: European Commission. Available at: search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680 5d6dda

Dagenais, C., Begin, J., Bouchard, C., & Forti, D. (2004). Impact of intensive family support programs: a synthesis of evaluation studies. *Children and Youth Services Review, 26*, 249–263. https://doi.org/10.1016/j.childyouth.2004.01.015

Dewey, J. (1933). *How We Think*, New York: D. C. Heath.

DoH, (2000). Framework for the assessment of children in need and their families. The family pack questionnaires and scales. London: The Stationery Office.

Dubowitz, H., Pitts, S. C., Litrownik, A. J., Cox, C. E., Runyand, D., & Black, M. M. (2005). Defining Child neglect based on child protective services data. *Child Abuse & Neglect, 29*, 493-511. https://doi.org/10.1016/j.chiabu.2003.09.024

Dumas, J. E. (2007). *Psychopathologue de l'enfant et de l'adolescent*. Bruxelles: Editions de Boek.

European Commission, (2013). *Investing in Children: Breaking the Cycle of Disadvantage*, Commission Recommendation of 20.2.2013, Brussels: European Commission. Available at: http://ec.europa.eu/justice/fundamentalrights/files/c_2013_778_en. pdf

Euser, S., Alink, L. R. A., Stoltenborgh, M., & Bakermans-kranenburg, M. J. (2015). A gloomy picture: a meta-analysis of randomized controlled trials reveals disappointing effectiveness of programs aiming at preventing child maltreatment. *BMC Public Health, 1068,* 1–14. doi: 10.1186/s12889-015-2387-9

Featherstone, B., White, S., & Morris, K. (2014). Re-*imagining child protection. Towards humane social work with families*. Bristol: Policy Press.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation Research: A Synthesis of the Literature. *Components, 311712,* 1–119. Available at: http://ctndisseminationlibrary.org/PDF/nirnmonograph.pdf

21

Geertz, C. (1973). *The Interpretation of Cultures*. New York: basic Books.

Geertz, C. (1997). *Local Knowledge*. New York: Basic Books.

Gershater-Molko R. M, Lutzker J. R., & Wesch, D. (2002). Using recidivism data to evaluate project safecare: Teaching bonding, safety, and health care skills to parents. *Child Maltreatment, 7*(3), 277-285. doi: 10.1177/1077559502007003009

Gonzales, R., Handley, M. A., Ackerman, S., & O'Sullivan, P. (2012). Increasing the Translation of Evidence into Practice, Policy, and Public Health Improvements: A Framework for Training Health Professionals in Implementation and Dissemination Science. *Acad Med, 87*(3), 271–278. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3307591/

Guba, E., & Lincoln, Y. (1989). *Fourth Generation Evaluation*. Newbury Park: Sage

Hammerseley, M. (2013). *The Myth of Research based Policy and Practice*. London: Sage

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources, 32*, 441-462. doi: 10.2307/146178

Heckman, J., & Robb R. (1985). Alternative Methods for Evaluating the Impact of Interventions. In Heckman, J. & Singer, B. (eds.), *Longitudinal Analysis of Labor Market Data.* New York: Wiley.

Heneghan, A. M., Horwitz, S. M., & Leventhal, J. M. (1996). Evaluating intensive family preservation programs: a methodological review. *Pediatrics, 97*, 535-542. Available at: pediatrics.aappublications.org/content/pediatrics/97/4/535.full.pdf

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of American Statistical Association, 81*, 945-970. doi: 10.2307/2289064

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451–467. doi: 10.1037/0022-006X.68.3.451

Imbens, G.W. (2000). The Role of Propensity Score in Estimating Dose-Response Functions. *Biometrika, 87*(3), 706-710. doi: 10.3386/t0237

Knorth, E. J., Harder, A. T., Zandberg, T., & Kendrick, A. J. (2008). Under one roof: A review and selective meta-analysis on the outcomes of residential child and youth care. *Children and Youth Services Review, 30*, 123-140. doi: 10.1016/j.childyouth.2007.09.001

Lacharité, C., Ethier, L., & Nolin, P. (2006). Vers une théorie éco-systémique de la négligence envers les enfants. *Bulletin de psychologie*, *59*(4), 381-394. doi: 10.3917/bupsy.484.0381

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components* S432001. Boston College, Department of Economics.

Léveillé, S., & Chamberland, C. (2010). Toward a general model for child welfare and protection services: A meta-evaluation of international experiences regarding the adoption of the Framework for the Assessment of Children in Need and Their Families (FACNF). *Children and Youth Services Review*, *32*, 929–944. https://doi.org/10.1016/j.childyouth.2010.03.009

Lorenz, W. (2012). Response: Hermeneutics and Accountable Practice: Lessons From the History of Social Work. *Research on Social Work Practice*, *22*(5), 492–498. doi: 10.1177/1049731512444167

Lorenz, W. (2016). Reaching the person—social work research as professional responsibility. *European Journal of Social Work*, 1–13. http://dx.doi.org/10.1080/13691457.2015.1137870

Macmillan, H. L., Wathen, N., Barlow, J., Fergusson, D. M., Leventhal, J. M., & Taussig, H. N. (2009). Interventions to prevent child maltreatment and associated impairment. *The Lancet*, *373*(17), 250-266. http://dx.doi.org/10.1016/S0140-6736(08)61708-0

Mullen, E. J. (2014). Evidence-based knowledge in the context of social practice. *Scandinavian Journal of Public Health, 47*, 59–73. doi: 10.1177/1403494813516714

Mullen, E. J. (2016). Reconsidering the "idea" of evidence in evidence-based policy and practice. European *Journal of Social Work*, *19*(3-4), 310-335. http://dx.doi.org/10.1080/13691457.2015.1022716

Neuhauser, A., Ramseier, E., Schaub, S., Burkhardt, S. C. A., Templer, F. & Lanfranchi, A. (2015). Hard to reach families. A methodological approach to early recognition, recruitment, randomization in an intervention study. *Mental Health & Prevention, 3*, 79–88.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika, 70*(1), 41-55. doi: 10.2307/2335942

Rubin, A. (2014). Bridging the Gap between Research-Supported Interventions and Everyday Social Work Practice: A New Approach. *Social Work, 59*, 223-230. doi: 10.1093/sw/swu023

Sanders, M. R., Markie-Dadds, C. & Turner, K. (2003). *Theoretical, Scientific and Clinical Foundations of the Triple P-Positive*

23

Parenting Program: A Population Approach to the Promotion of Parenting Competence. Brisbane, Qld, Australia: The Parenting and Family Support Centre.

Serbati, S., Ius, M., & Milani, P. (2016). P.I.P.P.I. Programme of Intervention for Prevention of Institutionalization. Capturing the Evidence of an Innovative Programme of Family Support. *Revista de Cercetare si Interventie Sociala, 52*, 26-50. The online version available at: www.rcis.ro, www.doaj.org and www.scopus.com

Serbati, S., & Milani, P. (2013). *La tutela dei bambini. Teorie e strumenti di intervento con le famiglie vulnerabili.* Roma: Carocci.

Serbati, S., Pivetti, M., & Gioga, G. (2015). Child Well-Being Scales (CWBS): in the assessment of families and children in home-care intervention: an empirical study. *Child & Family Social Work, 20*, 446-458. doi: 10.1111/cfs.12094

Soulet, P. M. (2014). Analyse qualité oupratique qualité? Quand la qualité devient un enjeu de définitiondu travail social? *Actualités sociales, 51*, 1-4.

Turney, D., & Taylor, J. (2014). Interventions in chronic and severe neglect: What Works? *Child Abuse Review, 23*(4), 231-234. doi: 10.1002/car.2351

Wandersman, A. (2009). Four keys to success (theory, implementation, evaluation, and resource/system support): High hopes and challenges in participation. *American Journal of Community Psychology, 43*, 3-21. doi: 10.1007/s10464-008-9212-x

White, K. M., & Dudley Brown, S. (2012). *Translation of Evidence into Nursing and Health Care Practice*. New York: Springer.