

*Working paper Cnr-Ceris, N.16/2014*

I DATI AMMINISTRATIVI PER LA  
VALUTAZIONE DELLE POLITICHE:  
RISCONTRI DALL'ESPERIENZA  
PIEMONTESE SUL FSE

Elena Ragazzi and Lisa Sella

**Working  
Paper**

**WORKING PAPER CNR - CERIS**

RIVISTA SOGGETTA A REFERAGGIO INTERNO ED ESTERNO

ANNO 16, N° 11 – 2014

Autorizzazione del Tribunale di Torino

N. 2681 del 28 marzo 1977

ISSN (print): 1591-0709

ISSN (on line): 2036-8216

**DIRETTORE RESPONSABILE**

Secondo Rolfo

**DIREZIONE E REDAZIONE***Cnr-Ceris*

Via Real Collegio, 30

10024 Moncalieri (Torino), Italy

Tel. +39 011 6824.911

Fax +39 011 6824.966

[segreteria@ceris.cnr.it](mailto:segreteria@ceris.cnr.it)[www.ceris.cnr.it](http://www.ceris.cnr.it)**SEDE DI ROMA**

Via dei Taurini, 19

00185 Roma, Italy

Tel. +39 06 49937810

Fax +39 06 49937884

**SEDE DI MILANO**

Via Bassini, 15

20121 Milano, Italy

tel. +39 02 23699501

Fax +39 02 23699530

**SEGRETERIA DI REDAZIONE**

Enrico Viarisio

[e.viarisio@ceris.cnr.it](mailto:e.viarisio@ceris.cnr.it)**DISTRIBUZIONE**

On line:

[www.ceris.cnr.it/index.php?option=com\\_content&task=section&id=4&Itemid=64](http://www.ceris.cnr.it/index.php?option=com_content&task=section&id=4&Itemid=64)**FOTOCOMPOSIZIONE E IMPAGINAZIONE**

In proprio

Finito di stampare nel mese di Ottobre 2014

**COMITATO SCIENTIFICO**

Secondo Rolfo

Giuseppe Calabrese

Elena Ragazzi

Maurizio Rocchi

Giampaolo Vitali

Roberto Zoboli

**Copyright © 2014 by Cnr-Ceris**

All rights reserved. Parts of this paper may be reproduced with the permission of the author(s) and quoting the source.

Tutti i diritti riservati. Parti di quest'articolo possono essere riprodotte previa autorizzazione citando la fonte.


# I dati amministrativi per la valutazione delle politiche: riscontri dall'esperienza piemontese sul FSE<sup>1</sup>

[Administrative data and policy evaluation: insights from an experience on FSE in Piedmont]

*Elena Ragazzi e Lisa Sella\**

*National Research Council of Italy  
Institute for Economic Research on Firm and Growth  
CNR-Ceris, via Real Collegio 30,  
10024 Moncalieri (To), Italy*

\* Corresponding author: [l.sella@ceris.cnr.it](mailto:l.sella@ceris.cnr.it)

 011-6824.926

**ABSTRACT:** Impact assessment is usually based on direct surveys of treated individuals. The net impact should be estimated observing untreated individuals which are very similar to the treated group. However, the high cost of implementation does not allow to use large samples, and therefore it is difficult to carry out statistically significant comparison at the local level. The open access to databases which were created for different purposes (e.g., tax, administrative, monitoring) could overcome these limitations with many advantages. In this article we present a validation exercise based on data records provided by the Italian law whenever a change in an employment contract occurs, i.e. the so called compulsory communications (COB). This was possible because of an exceptional access to two different data sources, with the aim to assess the impact in terms of employment of vocational training policies in Regione Piemonte (northwest Italy), that is the COB database and a direct survey performed on students one year after the course. We describe the major differences between indicators calculated on survey data and indicators calculated on COBs. Discrepancies are observed for 20% individuals in the sample analyzed. In addition, the change of source distorts the results of the net impact evaluation. In fact, both the determinants of the employment probability and the importance of the net impact do significantly vary across sources. Whenever an administrative database is used as a source for socio-economic analysis, it is essential to be cautious and critic, validating its reliability by comparison with different data sources. In the absence of such a preliminary validation process, which should take place in close collaboration with regional and national authorities managing the informational systems, researchers do not only risk to uncritically accept information that provide systematic distortions, but they prevent corrective procedures which are fundamental for the development and improvement of the whole system.

**Keywords:** net impact evaluation; validation; micro-data sources; survey; training policies

**JEL Codes:** C81; P46; J08

---

<sup>1</sup> Una versione preliminare dell'articolo è stata presentata al XVII Congresso Nazionale AIV, Napoli, 10-11/04/2014.

## INDICE

I dati amministrativi per la valutazione delle politiche: riscontri dall'esperienza piemontese sul FSE.....	3
1. Introduzione .....	5
2. Perché il ruolo dei dati amministrativi diventerà cruciale nella prossima programmazione?.....	7
3. I dati amministrativi e di monitoraggio: quali problemi? .....	8
4. Il caso di studio: la valutazione delle politiche formative in Regione Piemonte .....	10
4.1 Il fenomeno analizzato e i dati disponibili .....	11
4.2 Confronto fra indicatori .....	13
4.2.1 Indicatore dicotomico puntuale .....	13
4.2.2 Indicatore dicotomico continuo .....	14
4.2.3 Numero di contratti.....	14
5. Gli scostamenti nelle variabili.....	15
5.1 Indicatore dicotomico puntuale.....	15
5.2 Altri indicatori.....	18
6. Gli scostamenti negli impatti .....	19
7. Conclusioni e indicazioni di ricerca futura .....	20
Bibliografia.....	23

## 1. INTRODUZIONE

**I**l dibattito europeo, così come gli orientamenti comunitari, già durante la scorsa programmazione 2007-2013 hanno promosso un sempre più ampio e consapevole utilizzo delle valutazioni di impatto netto, ma tale obiettivo è ancora ben lungi dall'essere raggiunto. Le analisi d'impatto, seppur già da tempo considerate uno strumento indispensabile per la programmazione, hanno conosciuto nelle pubbliche amministrazioni europee una fortuna limitata ed applicazioni abbastanza episodiche, soprattutto in merito alla stima degli impatti netti. I metodi controfattuali, infatti, sono stati per lo più impiegati in valutazioni sperimentali di limitati progetti pilota, ma non in contesti quasi sperimentali e di ampie programmazioni. Ciò è in buona parte dovuto alla complessità del loro disegno e della loro implementazione.

Molti spunti, sia dalla letteratura sia dai documenti di programmazione, sembrano ora indicare come percorso da esplorare per raggiungere tale obiettivo l'utilizzo dei dati di forte amministrativa. Tale strada, che è oggi sempre più spianata dal miglioramento dei sistemi informatici di gestione del monitoraggio e che sembra aprire enormi potenzialità soprattutto nell'ottica dell'integrazione di fonti amministrative differenti, garantirebbe infatti la disponibilità di ampie basi dati a basso costo, in grado di osservare l'universo dei trattati per un periodo lungo a piacere dopo la conclusione dell'intervento. Tale prospettiva, davvero attraente, è però ad oggi ben lontana dall'essere facilmente applicabile e, per ora, la semplificazione è solo apparentemente. Due sono gli ordini di problemi che spiegano la

difficoltà di impiegare i dati amministrativi a fini di analisi e valutazione:

1. i sistemi che gestiscono le basi dati sono stati creati per finalità differenti, che rendono l'analisi difficoltosa, se non impossibile. Un esempio concreto è rappresentato dalle banche dati sulle Comunicazioni Obbligatorie (COB), ovvero le comunicazioni che i datori di lavoro sono tenuti a trasmettere in caso di assunzione, proroga, trasformazione, cessazione di un contratto. In quanto tali esse registrano degli eventi e non degli stati, rendendo abbastanza oneroso il loro impiego per valutare lo stato occupazionale degli individui. Un altro caso, ben più problematico, si ha quando le singole unità (individui, imprese) non sono caratterizzate da una notazione univoca nei vari database (o nelle loro sezioni) ed è dunque impossibile ricostruire i micro-dati indispensabili per la valutazione di impatto.
2. Gli amministratori pubblici, sotto la cui autorità ricadono i sistemi informativi, spesso non posseggono un'adeguata cultura del dato e cadono sovente nella trappola per cui un qualsiasi valore registrato sia automaticamente considerato portatore di informazioni, indipendentemente dalla sua qualità. In realtà, come tutti i ricercatori che si trovano a lavorare su dati secondari fanno, la qualità del dato dipende, oltre che da completezza, dettaglio e affidabilità, anche da una serie di informazioni accessorie (per esempio quando è stato raccolto? con che procedura? da chi?) che permettono di qualificarne la portata informativa. Variabili apparentemente molto simili (per esempio lo stato occupazionale dichiarato

dall'individuo e quello rilevato dal centro per l'impiego) possono infatti portare a risultati drasticamente differenti, ed è dunque essenziale disporre, oltre che della banca dati, anche di un kit di informazioni contenutistiche, che risultano invece assenti nei dizionari attualmente disponibili, e che sono oltretutto disperse fra più operatori e a volte anche fra più soggetti diversi, a causa dei frequenti fenomeni del decentramento amministrativo e di esternalizzazione dei servizi.

I valutatori che sperimentano l'utilizzo di questa tipologia di dati devono investire molto tempo a familiarizzarsi e documentarsi sulla fonte (con costi dunque occulti e difficili da preventivarsi), pena il rischio di errori e distorsioni che risultano poi molto difficili da diagnosticare o spiegare con la conseguente analisi quantitativa. D'altronde, anche la strada alternativa di costruire direttamente l'andamento della variabile obiettivo attraverso inchieste presso i beneficiari/destinatari non è esente da problemi. L'alto costo dell'indagine sul campo impone di ricorrere a tecniche campionarie, introducendo margini di errore nelle stime.

Inoltre, l'informazione raccolta si basa sulle dichiarazioni degli interessati e soffre anch'essa di rischi di distorsioni (timore del giudizio altrui, difficoltà ad astrarre un giudizio o a ricordare, differenti criteri di valutazione, influenza di come è realizzata l'intervista, etc.). Dunque, a priori è difficile scegliere quale sia la strada migliore per rispondere ad esigenze valutative e analitiche.

Come valutatori indipendenti del POR-FSE della Regione Piemonte, abbiamo condotto, ormai da tre anni, delle indagini survey

campionarie con disegno controfattuale per valutare l'efficacia delle politiche formative in termini di inserimento lavorativo.

Recentemente, abbiamo anche iniziato a lavorare sulla banca dati SILP (Sistema Informativo Lavoro della regione Piemonte), che contiene i dati delle COB piemontesi, inizialmente solo per integrare alcune informazioni mancanti in fase di campionamento, più recentemente per la validazione dei dati raccolti tramite survey.

In questo contributo vogliamo dunque trarre vantaggio dalla particolare situazione in cui ci troviamo, disponendo contemporaneamente di due fonti di informazioni differenti sulla stessa politica (dati amministrativi e inchieste sul campo), per presentare un'analisi comparativa dei due percorsi di valutazione, individuando le possibili fonti di errori di misura e di stima. A partire dal confronto fra i dati e fra le stime di impatto ottenute con le due fonti, l'articolo vuole descrivere alcuni fra i principali problemi che possono sorgere e fornire una prima road map per l'implementazione di una strategia di valutazione basata su dati amministrativi. Particolare attenzione verrà data all'analisi delle differenze fra i due campioni nella misurazione delle variabili esplicative e di risultato. L'articolo parte da una contestualizzazione della problematica nell'ambito della programmazione europea (§2) e prosegue con una schematizzazione dei termini di confronto fra analisi basate sulle survey e analisi basate sui dati amministrativi (§3). Viene quindi presentato il servizio di valutazione svolto (§4), da cui sono tratti i dati per la successiva validazione dei due percorsi di analisi, in termini di variabili misurate (§5) e di impatti generali e su target specifici (§6).

## 2. PERCHÉ IL RUOLO DEI DATI AMMINISTRATIVI DIVENTERÀ CRUCIALE NELLA PROSSIMA PROGRAMMAZIONE?

La futura programmazione sarà sempre più centrata sull'efficacia e per questo occorre costruire una capacità di leggere i risultati per imparare dalle lezioni passate. Lo riconosceva già il rapporto Barca (2009): *“Capacity to apply lessons must be strengthened; and the failure to develop knowledge on “what works”, i.e. on whether interventions are producing effects, and “for whom” must be remedied”*. Perché questo non resti uno slogan, da un lato è necessario adottare e imporre (attraverso opportuni meccanismi che leghino i finanziamenti al raggiungimento dei risultati) un atteggiamento visionario: capire dove si vuole andare e premiare chi ci arriva. Dall'altro lato è necessario essere umilmente realisti, perché questo richiede capacità tecniche e dotazioni che non si costruiscono dal nulla: *“to promote impact evaluation as one of the methodological backbones of cohesion, policy must at the same time be visionary and humble. Visionary, since it must create the technical, administrative and political momentum for impact evaluation to be actually pursued as an important source of information. Humble, since building this technical capacity requires time, without expectation of a “quick fix” able to provide uncontroversial answers on which parts of cohesion policy are working and where”* (Barca, 2009). La percezione dell'insufficiente preparazione delle Autorità di Gestione a realizzare valutazioni di impatto era già da tempo argomento di discussione.

Già nel 2008 la Commissione si esprimeva esplicitamente sulla necessità di aumentare la

capacità istituzionale di tutti i livelli di autorità riguardo alla valutazione di impatto: *“Looking ahead, it is advisable for the EU to establish the improvement of institutional setting for impact evaluation as a priority within its larger ambition to strengthen the public “institutional capacity” at all levels in Convergence regions and Cohesion countries”* (Commission, 2008). Questo perché le valutazioni di impatto fatte fino a quel momento erano poche, inficciate da dati di monitoraggio di scarsa qualità (Bachtler et al., 2006) e in generale poco credibili a causa della governance multi-livello e dell'eterogeneità dei contesti e degli interventi (Baslé, 2006).

Tornando al rapporto Barca, l'azione per la disponibilità di dati è una delle sei guidelines delle sue conclusioni:

- *Rigorous separation from the use of indicators;*
- *A focus on counterfactual methods;*
- *Prospective impact evaluation;*
- *Experimenting with randomisation;*
- *Commission role: clearing house and think tank;*
- *Data availability.*

In particolare, per annullare i vincoli derivanti dalla mancanza di informazioni adeguate, viene sottolineata l'opportunità che i dati necessari alla valutazione di impatto vengano raccolti congiuntamente al processo di implementazione, allargando dunque il ruolo dell'attività di monitoraggio. *“The availability of data is often a constraint in the use of impact evaluation. Many of the necessary data could (or can only) be collected in the process of implementing the interventions which they should help evaluating, but they fail to be collected because their need is not identified in time”*.

Infine, la possibilità di integrare i dati di monitoraggio con dati derivanti da altre fonti amministrative amplifica enormemente le potenzialità delle valutazioni di impatto, facendo cadere molti vincoli (temporali, di numeri, di budget, di variabili disponibili). Questa strada sembra sempre più percorribile per l'evoluzione dei sistemi informativi delle pubbliche amministrazioni. Ma la difficoltà e i rischi connessi a tali operazioni sono spesso sottovalutati, nella valutazione di opportunità, rispetto all'elenco dei vantaggi. Come evidenzia la Guida per la valutazione dello sviluppo socio-economico (DG for Regional Policy., 2013), i database amministrativi contengono dati secondari e occorre dunque sapere come sono stati prodotti e gestiti prima di poterli utilizzare. *“What samples were used, how were outcomes defined, what is the timescale covered, what is the unit of analysis? It is only by asking these questions that a judgement can be made about their usability in a particular evaluation”*.

Il prossimo paragrafo intende dunque soffermarsi su una disamina generale delle tipologie di problemi che un ricercatore deve di apprestarsi ad incontrare quando intende utilizzare dati amministrativi a scopi valutativi.

### 3. I DATI AMMINISTRATIVI E DI MONITORAGGIO: QUALI PROBLEMI?

La nostra riflessione si è avviata a partire dalla lettura del rapporto della commissione che, negli USA, ha analizzato l'utilizzo dei dati amministrativi per ricerche di rilevanza politica (Hotz et al. eds., 1999) Il rapporto è stato pubblicato nel 1998; malgrado fossero passati più di 15 anni, in esso pareva di leggere alla perfezione, il nostro attuale

sforzo per integrare i dati amministrativi, in particolare quelli derivanti dalle COB.

Le principali categorie di problemi in cui si incorre nell'utilizzo di banche dati amministrative o di monitoraggio e, in particolare, nella loro integrazione sono:

1. ottenere libero accesso ai dati;
2. riconoscere univocamente l'individuo (tra DB, nel tempo,...);
3. il processo di validazione dei dati;
  - a. capirne la natura;
  - b. esaminare errori e missing;
  - c. valutare le fonti di distorsione;
  - d. Essere informati e gestire i cambiamenti nelle regole del DB.

Il problema dell'accesso ai dati è molto diffuso e non deve essere sottovalutato. Le amministrazioni sono poco inclini a consentire a soggetti sia esterni che interni l'accesso ai dati amministrativi per molti motivi, a partire dallo scarso apprezzamento per le analisi che possono derivarne. Soprattutto, non essendo in grado di codificare ed esplicitare chiaramente tutto quel patrimonio di informazioni accessorie che sono indispensabili per l'interpretabilità e il corretto utilizzo dei dati, esse hanno difficoltà ad allargarne l'utilizzo, sia per il timore di interpretazioni scorrette, sia per la diffusa saturazione dello scarso personale competente a guidare l'accesso alle banche dati. Inoltre, recentemente è divenuto sempre più di ostacolo il problema della privacy, a causa dei maggiori vincoli e controlli che non sono ancora accompagnati da un quadro normativo esaustivo. Infine accade, seppur di rado, di imbattersi in una scarsa disponibilità all'accesso derivante dalla coscienza che, oggi più che mai, l'informazione è una risorsa di grande valore, che non va dispersa ma utilizzata in modo strategico. Nella nostra



esperienza, ancora più che una chiara definizione preliminare dei termini con cui avverrà l'accesso ai dati (es. quali informazioni, quali accessi, strategie per garantire la confidenzialità, proprietà dei risultati e loro disseminazione), è stato premiante un lungo lavoro relazionale per far percepire la valutazione/ricerca come una risorsa importante in fase di programmazione, piuttosto che un obbligo o un controllo ispettivo. Mentre il problema della tracciabilità dell'individuo fra i diversi database è risolto qualora si possa disporre del codice fiscale come identificativo della persona, l'interpretazione e pulizia dei dati sono complicate da una serie di problematiche, che rendono l'impiego del dato amministrativo oneroso e la sua interpretazione spesso poco precisa. Come già accennato sopra, i database non sono corredati di informazioni che qualifichino il dato (quando è stato raccolto, con quale procedura, da chi, cosa significano esattamente le diverse modalità). Accade di trovarsi per esempio di fronte a campi con nome simile o addirittura uguale, ma con contenuto diverso. Se non si riesce a comprendere la genesi di quel campo, si rischiano dunque interpretazioni errate. Tali informazioni sono disperse fra più operatori, spesso anche in differenti enti pubblici (per il decentramento delle competenze) o privati (per l'esternalizzazione dei servizi). La difficoltà nel capire la natura dei dati aumenta quando i database derivano dalla sistematizzazione e stratificazione nel tempo di sistemi informativi inizialmente separati o quando vi siano stati cambiamenti nelle definizioni o nelle procedure. Il ricercatore/valutatore deve dunque attendersi, almeno nelle prime esperienze di impiego delle banche di micro-dati, di dover investire

molto tempo a impraticarsi, indagando e verificando la portata informativa dei vari campi, prima di poter utilizzare i dati disponibili. Indipendentemente da queste problematiche, che sono sicuramente di non semplice soluzione ma non sono comunque strutturali, deve essere chiaro che le banche dati amministrative e di monitoraggio nascono con finalità differenti rispetto a quelle di analisi e valutazione e che quindi la loro portata informativa non è sempre adeguata alla domanda di ricerca. La Tabella 3.1 pone a confronto, su varie dimensioni di interesse per l'analista, le tre possibili fonti di dati, ovvero le indagini ad hoc, i dati di monitoraggio delle politiche o dei progetti e i dati amministrativi indipendenti dalla politica oggetto di valutazione. Le indagini ad hoc consentono ovviamente di focalizzare al meglio i destinatari e le informazioni raccolte, ma soffrono di numerosità limitate a causa dei vincoli di budget e organizzativi. La loro portata informativa è inoltre limitata all'ampiezza e precisione di visione del soggetto intervistato, che può non avere su certi temi sufficienti elementi per rispondere, mentre su altri può essere portato a risposte errate per timore del giudizio altrui e per difficoltà ad astrarre un giudizio o a ricordare o ancora a comprendere bene il quesito.

Il lavoro presentato in questo articolo parte da un'esigenza valutativa, ovvero misurare l'impatto in termini di occupazione delle politiche formative, e mette a confronto i risultati nel caso si impieghino due fonti informative differenti. Esso è dunque in grado di individuare l'esistenza di problemi, quantificando gli scostamenti, mentre resta tutta da percorrere la strada per comprenderne l'origine e dunque valutare la reale fonte di errore.

Tabella 3.1: Confronto fra dati raccolti sul campo, dati amministrativi e dati di monitoraggio

	<b>Survey</b>	<b>Monitoraggio</b>	<b>Amministrativi</b>
<b>Universo</b>	Esattamente quello dell'indagine, purché disponibili le liste dei trattati	Esattamente quello di riferimento della politica, non necessariamente coincidente col la popolazione valutata	Differente. Difficile ottenere l'esatta coincidenza e difficile valutare la comparabilità
<b>Copertura temporale</b>	Limitata. In genere solo <i>ex-post</i> . Problemi di affidabilità della memoria degli intervistati	Limitata all'arco temporale di attività della politica	Ampliabile a piacere, ma problemi di coerenza intertemporale delle serie storiche
<b>Numerosità dei campioni</b>	Limitata dal budget e dai problemi di caduta di intervista	Limitata ai soli trattati – no informazioni su controfattuale (salvo esperimenti)	Ampia
<b>Indicatori</b>	Ad hoc per l'indagine, ma solo per informazioni note all'intervistato	Limitati alle esigenze di gestione della politica e progettati per quello	Ci si deve adattare ai dati disponibili; difficile valutare gli scostamenti
<b>Campioni controfattuali</b>	Limitati dal budget	Limitati ai no-shows e drop-out	Grandi dimensioni ma bassa omogeneità

#### 4. IL CASO DI STUDIO: LA VALUTAZIONE DELLE POLITICHE FORMATIVE IN REGIONE PIEMONTE

Il presente articolo parte da un'analisi di impatto realizzata per la valutazione delle politiche formative finanziate con fondi FSE dalla Regione Piemonte e intende utilizzare i riscontri di tale attività per rispondere alla domanda di ricerca: *“su quale fonte di dati è opportuno disegnare l'esercizio di valutazione di impatto netto?”*.

In passato gli esiti occupazionali in uscita erano investigati dalle stesse agenzie

formative, ma la pratica è stata abbandonata a favore di indagini campionarie dirette, per l'evidente rischio di inquinamento dei dati.

Contemporaneamente l'archivio SILP, che raccoglie i dati dei centri per l'impiego, inclusi quelli sulle COB, si presenta come una nuova opportunità da esplorare, a fronte dei numerosi vantaggi indicati nel paragrafo precedente.

Ci si chiede dunque se sia auspicabile svolgere l'esercizio sui soli dati amministrativi, risparmiando i costi di survey e ampliando enormemente la numerosità dei campioni e l'arco temporale analizzabile.

In questo paragrafo verranno descritte in dettaglio le fonti alternative disponibili e gli indicatori da esse derivabili, mentre nei paragrafi successivi verrà realizzata una cross-validazione dei dati survey e dei dati amministrativi attraverso una quantificazione dello scostamento delle diverse variabili nei due casi.

L'analisi, oltre alla verifica sulla misurazione del livello di occupazione con i due metodi, verificherà anche come varia la stima degli impatti (incremento di occupazione attribuibile alla politica formativa) utilizzando i dati delle due fonti sul medesimo campione di individui.

In altri termini, a parità di campione (ovvero quello intervistato nell'indagine survey), si osserverà se e come variano gli indicatori che descrivono lo stato occupazionale e la misura dell'impatto netto. Resta vero che tale scelta risponde esclusivamente ad esigenze di diagnosi metodologica, mentre una valutazione disegnata su dati amministrativi consentirebbe un'analisi sull'intero universo (vantaggiosa soprattutto per il gruppo controfattuale) e avverrebbe con criteri e tecniche assolutamente differenti.

#### *4.1 Il fenomeno analizzato e i dati disponibili*

A partire dall'anno scolastico-formativo 2009/2010 la Regione Piemonte ha deciso di affidare al Valutatore Indipendente il compito di realizzare un'indagine di Placement sugli esiti occupazionali della frequenza ai corsi di qualifica e di specializzazione. Fino ad oggi sono state realizzate tre indagini con tecnica CATI (Computer Aided Telephone Interview), relative ai formati 2010, 2011 e 2012, intervistati a circa un anno di distanza dalla conclusione dei corsi. Il presente lavoro

si riferisce all'indagine 2012 sui qualificati nell'anno 2011.

La popolazione target è costituita da tutti i qualificati e specializzati che non fossero occupati all'atto dell'iscrizione. Restringendo l'analisi ai soli corsi di qualifica e specializzazione, mediamente più lunghi rispetto ad altri tipi di corso (es. patenti di mestiere, abilitazione professionale, ecc.) e caratterizzati dalla presenza obbligatoria di uno stage, si è specificamente concentrata l'attenzione su corsi che implicassero l'uscita dal mercato del lavoro e una sorta di discontinuità con i percorsi lavorativi precedenti. Per questo non deve essere scordato che la valutazione di placement da cui parte questo lavoro si riferisce a politiche di recupero dello svantaggio, piuttosto che a politiche di accumulo del capitale umano. Ciò consente una valutazione appropriata, ancorché non generalizzabile, dell'impatto delle politiche formative in Regione Piemonte, prevalentemente indirizzate a soggetti deboli, che generalmente manifestano gravi difficoltà nell'inserimento lavorativo. I corsi valutati si riferiscono a un'ampia gamma di professioni, in cui si rileva però una netta prevalenza del comparto dei servizi, in particolare della cura della persona (servizi socio-sanitari, estetica, acconciatura) e nella filiera della ristorazione.

Il campione di 1532 individui è estratto da una popolazione di 9605 individui, con stratificazione in base alle tipologie di azione e alla frequenza di politiche attive post-formazione e con riproporzionamento degli strati in base alle variabili anagrafiche (genere, età, nazionalità).

Per quanto riguarda il campione di confronto, un'attenta analisi del contesto ha suggerito di estrarre il campione di controllo

dai cosiddetti no-shows (Bell et al., 1995) e dai drop-out, ovvero dalla lista degli studenti che non erano occupati all'atto dell'iscrizione e che o non si sono presentati al corso di formazione o non lo hanno terminato con successo.

Tali individui risultano massimamente omogenei rispetto ai formati per situazione occupazionale precorso (dato osservabile) e per atteggiamento nei confronti della formazione (variabile non osservabile), come testimoniato dall'avvenuta iscrizione. Il numero di tali soggetti non è altissimo, per cui il campione di controllo ha necessariamente una dimensione molto ridotta, 491 individui, ed è stato campionato con lo stesso disegno del gruppo principale.

Le interviste hanno indagato, oltre allo stato occupazionale, le condizioni di contesto, la soddisfazione riguardo al corso, le attività svolte per la ricerca del lavoro e le caratteristiche dell'impiego (nel caso degli occupati) o quelle del percorso di studio (per il rientro in istruzione). Per una descrizione più dettagliata del database ottenuto e per la giustificazione delle scelte metodologiche si rimanda a Ragazzi e Sella (2013b).

I dati raccolti permettono dunque un'analisi degli esiti occupazionali che non si limita allo stato dicotomico occupato/disoccupato, ma che riguarda anche gli aspetti qualitativi dell'impiego e del percorso precedente. Al contrario, le possibilità di valutare gli aspetti longitudinali (alternanza e durata degli spell lavorativi e di disoccupazione) è limitata dall'imprecisione nella memoria degli intervistati.

Per integrare i dati dell'indagine, sia in fase di definizione dell'universo, sia a valle per verifiche sui percorsi di inserimento, si è contemporaneamente utilizzata la banca dati

SILP, che raccoglie le registrazioni delle comunicazioni obbligatorie, unitamente ad altre informazioni derivanti dai centri per l'impiego. Le COB riguardano tutti gli eventi di instaurazione, proroga, trasformazione, cessazione di un rapporto di lavoro, e avvengono secondo i modelli unificati definiti dal Ministero del Lavoro, della Salute e delle Politiche Sociali, da parte di tutti i soggetti obbligati e abilitati. Tali comunicazioni assolvono agli obblighi di comunicazione nei confronti degli enti previdenziali. Longitudinalizzando tali eventi, cioè trasformando una serie di eventi riferiti allo stesso individuo in una serie di spell lavorativi, è in via di principio possibile ricostruire lo stato occupazionale dell'individuo in un certo arco di tempo. Sarebbe quindi possibile utilizzare tali dati per verificare il placement dei destinatari di politiche formative e, più in generale, di politiche attive o passive del lavoro (Canu, Conzimu e Garau, 2010). Naturalmente le due strade, quando anche non presentino i problemi che verranno sviscerati nei prossimi paragrafi, non sono perfettamente coincidenti e hanno una serie di vantaggi/svantaggi che li differenziano. Di seguito alcuni elementi che abbiamo preso in considerazione nel bilancio comparativo fra le due fonti (che saranno, per sintesi, indicate come survey e COB):

- *Flessibilità*: la survey permette la raccolta di informazioni ad ampio spettro, perfettamente tarate per le esigenze valutative specifiche e per gli eventuali approfondimenti. Al contrario, le COB sono caratterizzate da notevole rigidità, in quanto ci si deve limitare alle informazioni disponibili negli archivi e la cui accessibilità è limitata dai problemi (di privacy e di validazione) descritti nel §3;

- *Costo*: se per ogni unità aggiuntiva intervistata tramite survey occorre preventivare un costo supplementare, obbligando la maggior parte delle volte a procedere con disegni campionari, nessun costo marginale va previsto nel caso delle COB e l'indagine può essere svolta per l'intero universo prescelto;
- *Tempi di realizzazione*: nel caso delle survey sono lunghi, in quanto la necessità di ridurre le cadute di intervista per non introdurre distorsioni sistematiche non diagnosticabili implica il bisogno di procedere a numerosi richiami prima di sostituire un nominativo. L'allungamento dei tempi di indagine sul campo può essere esso stesso fonte di distorsioni;
- *Possibili fonti errore*: nel caso delle survey vanno citati gli errori di memoria e i margini di soggettività nell'interpretare i quesiti e nel valutare i fenomeni da parte degli individui, la distorsione sistematica da "hard-core" (Cochran, 1977), l'irreperibilità di alcune tipologie di soggetti (per esempio stranieri, giovanissimi), le ambiguità interpretative legate all'impostazione del quesito specifico, errori indotti dagli intervistatori che materialmente effettuano le interviste. Ma anche, nel caso delle COB, ci sono delle possibili fonti di errori (e altre potrebbero emergere da un'esplorazione più approfondita delle procedure di funzionamento): ad esempio, viene segnalata una scarsa affidabilità del dato (non aggiornato, poco curato, caricato con ritardo variabile), vi sono casi di ambigua portata informativa (per esempio nel caso di spell lavorativi sovrapposti). In generale, come già si è detto, molte informazioni accessorie sono imprecise.

Infine, con l'analisi delle COB si perdono alcuni rapporti di lavoro: a parte il lavoro nero, che risulta difficile da indagare anche con una survey demoscopica e che richiede tecniche di intervista diretta, delle sulle COB non vengono registrate tutte le forme di lavoro autonomo (incluso il coadiuvante familiare); inoltre se il riferimento è a una banca dati regionale, non sono visibili i rapporti di lavoro contrattualizzati fuori regione.

#### 4.2 Confronto fra indicatori

Per analizzare le prestazioni delle due fonti informative al fine di valutare gli esiti occupazionali delle politiche formative, sono stati identificati tre indicatori differenti, calcolabili con entrambe le fonti, in modo da verificare la robustezza dei risultati trovati a diverse specificazioni della variabile obiettivo.

##### 4.2.1 Indicatore dicotomico puntuale

Questo indicatore è detto dicotomico perché confronta esclusivamente lo stato occupato/non occupato, senza ulteriori specificazioni (es. studente, inattivo) e senza considerazione degli aspetti qualitativi dell'occupazione. Inoltre esso è detto puntuale, in quanto si riferisce allo stato occupazionale in un preciso istante di tempo, nella fattispecie a ottobre 2012, ovvero a circa un anno dalla fine del percorso formativo. Per quanto riguarda la survey, l'indicatore è stato misurato attraverso un quesito specifico, che indaga la condizione professionale prevalente nel mese suddetto. Per quanto riguarda le COB, invece, si sono rilevate le date di inizio e fine dei rapporti di lavoro degli individui, restringendo l'analisi ai soli contratti attivi nell'ottobre 2012.

Tale indicatore è di semplice definizione e misurazione, ma troppo naive per rappresentare adeguatamente le caratteristiche degli spell occupazionali (es. durata, tipologia, grado di precarietà, ecc.). Inoltre, vi sono alcuni problemi aggiuntivi, legati alle caratteristiche specifiche delle fonti informative. Per quanto riguarda le survey, l'indicatore deriva da una autodichiarazione sulla condizione professionale prevalente che, come tale, è soggettiva e a rischio di errori di memoria e di omissioni volontarie (per esempio per il timore di cosa potrebbe succedere in seguito alla dichiarazione di lavorare in nero). Nel caso delle COB, invece, il criterio di definizione dello stato è oggettivo e unico, ma resta discrezionale la definizione di occupazione: si può definire occupato un individuo se in quel mese ha lavorato almeno un giorno, oppure ci si deve limitare a spell occupazionali (anche multipli) che coprano tutti i giorni del mese? Oppure è preferibile qualche altra definizione intermedia?

#### 4.2.2 *Indicatore dicotomico continuo*

L'accento dell'indicatore, in questo caso, è posto sulla durata degli spell occupazionali, prendendo in considerazione tutto il tempo lavorato in un certo intervallo di tempo (gennaio-ottobre 2012). Pur considerando solo due possibili stati alternativi (occupato/non occupato), esso misura un'intensità e non solo lo stato in un certo periodo. Dunque, rappresenta il placement meglio dell'indicatore puntuale, ma non cattura comunque la multidimensionalità del fenomeno. Nel caso della survey, un apposito quesito chiedeva il numero di mesi lavorati fra gennaio e ottobre 2012. Tale scelta soffre di imprecisioni, a causa degli errori di memoria e

dell'approssimazione legata alla formulazione del quesito.

Nel caso delle COB, si è calcolata la durata complessiva dei rapporti attivi fra gennaio e ottobre 2012. Questo ha comportato, dal punto di vista del data-management, la necessità di gestire le sovrapposizioni.

Inoltre, è un metodo che può essere applicato solo per misurare eventi sufficientemente lontani nel tempo, in quanto attualmente si osservano frequenti ritardi nella registrazione delle comunicazioni. In compenso, questo metodo permette un calcolo preciso del numero di giorni lavorati, fermo restando che non si possono osservare gli episodi di auto-impiego, di lavoro non contrattualizzato o al di fuori del territorio regionale.

#### 4.2.3 *Numero di contratti*

Questo indicatore fornisce informazioni complementari rispetto ai precedenti, in quanto esso non misura tanto l'intensità lavorativa quanto piuttosto, considerato congiuntamente al numero di mesi lavorati, la frammentarietà della carriera. Comunque, anche tale indicazione assume una rilevanza molto diversa a seconda dei settori e delle professioni.

Anche in questo caso, nella survey è stato inserito un quesito specifico sul numero di contratti avuti nel solito intervallo di tempo gennaio-ottobre 2012, per cui valgono gli stessi caveat dell'indicatore dicotomico continuo in merito a difetti di memoria e tendenza all'approssimazione.

Nel caso delle COB, si è calcolato il numero di rapporti di lavoro per individuo, facendo riferimento al codice identificativo dei singoli rapporti.

## 5. GLI SCOSTAMENTI NELLE VARIABILI

In questo paragrafo si mostrano gli esiti del calcolo degli indicatori presentati nel §4.2, computandoli per lo stesso campione di individui oggetto dell'indagine di placement 2012, ma utilizzando entrambe le fonti informative descritte. I risultati mostrano non solo la misura delle discordanze, ma anche le loro caratteristiche.

### 5.1 *Indicatore dicotomico puntuale*

La Tabella 5.1 mostra lo stato occupazionale ad ottobre 2012 di ogni individuo inserito nel campione principale o di controllo, così come auto-dichiarato nella survey, e così come risulta dalla longitudinalizzazione delle COB.

I dati sulla diagonale principale rappresentano gli individui per cui si hanno informazioni coincidenti, ovvero per cui le due fonti informative risultano in accordo. Al

contrario gli individui inclusi nelle celle poste sulla diagonale secondaria (A e B) presentano discordanze fra valore dichiarato e valore registrato. Essi ammontano a 408 record, pari al 20,2% del campione complessivo.

Nella casella A si trovano coloro che hanno dichiarato di lavorare in occasione dell'intervista ma che invece non risultano occupati in base all'archivio SILP. Molte spiegazioni di tipo diverso sono possibili, che vanno dalle categorie escluse dall'archivio (come il lavoro autonomo o fuori regione) agli errori di memoria. La Tabella 5.2 permette quindi di approfondire la tipologia di lavoro svolto, quantificando le forme che non sono presenti nell'archivio delle COB (lavoro nero, autonomo, stage).

Dei 201 che lavorano ma non compaiono nelle COB, 125 dichiarano di essere dipendenti o parasubordinati; quindi, più del 60% di questi casi anomali dovrebbe comparire nell'archivio, a meno che non si tratti di lavoro fuori regione o all'estero.

*Tabella 5.1: Stato occupazionale ad ottobre 2012, a seconda del tipo di misura*

		COB occupato almeno un giorno		
		No	Si	Tot.
CATI condizione professionale prevalente	No	976	207 (B)	1183
	Si	201 (A)	639	840
Tot.		1177	846	2023

*Tabella 5.2: Tipo di lavoro dichiarato durante l'indagine di placement*

<b>Tipo di lavoro</b>	<b>V.A.</b>	<b>%</b>
Lavoro dipendente	113	56,2
Lavoro senza contratto (nero)	31	15,4
Autonomo	25	12,4
Coadiuvante di impresa familiare	6	3,0
Parasubordinato	12	6,0
Tirocinio, stage, praticantato	14	7,0
<b>Totale</b>	<b>201</b>	<b>100,0</b>

Per i 207 individui inclusi nella casella B, che risultano aver lavorato almeno un giorno ma si sono dichiarati non occupati, risulta più difficile trovare una spiegazione logica se non ipotizzando che, in caso di contratti molto brevi (per esempio un giorno di lavoro o poco più), la percezione complessiva della condizione professionale per il mese di riferimento non sia di occupazione. Ma anche questa congettura viene a cadere, in quanto il 73% di tali soggetti aveva lavorato per tutto il mese, secondo gli archivi COB.

Per verificare ulteriormente la robustezza dei risultati e vedere se essi dipendono in qualche modo da come è stato definito l'indicatore dicotomico puntuale di occupazione, sono state calcolate le correlazioni, riportate nella Tabella 5.3, fra l'indicatore calcolato sui dati survey e diverse formulazioni dell'indicatore calcolato sui dati COB (rispettivamente, l'individuo è classificato come occupato se lavora almeno un giorno, se lavora almeno quindici giorni o se lavora tutto il mese).

*Tabella 5.3 Correlazioni fra diversi indicatori di stato occupazionale*

	<b>Survey</b>	<b>COB Almeno 1 g.</b>	<b>COB Almeno 15 gg.</b>	<b>COB Tutto il mese</b>
<b>Survey</b>	1,0000			
<b>COB Almeno 1 g</b>	<b>0,5851</b>	1,0000		
<b>COB Almeno 15 gg.</b>	<b>0,5919</b>	0,9609	1,0000	
<b>COB Tutto il mese</b>	<b>0,5921</b>	0,9025	0,9392	1,0000



Considerando la prima colonna, che riporta le correlazioni delle varie definizioni di indicatore puntuale COB con i dati survey, si osserva una variazione minima fra le varie tipologie, confermando il permanere dell'incoerenza fra le diverse fonti nelle varie configurazioni.

Infine, occorre portare l'attenzione su un fatto molto importante. Le celle A e B sono caratterizzate quasi dalla stessa numerosità. Questo significa che i tassi di occupazione calcolati con i due procedimenti risultano uguali praticamente identici: rispettivamente 41,8% (survey) e 41,5% (COB). Tale equivalenza non segnala, però, un'elevata

omogeneità fra database, bensì deriva da una compensazione di errori di diversa natura che, come si vedrà nel §6, non è esente da conseguenze allarmanti soprattutto nella valutazione degli impatti, poiché le caratteristiche degli individui presenti nelle due celle non sono uguali. Per analizzare tali differenze nella Tabella 5.4 sono riportate alcune variabili misurate rispettivamente per i casi in cui le due fonti sono concordi a misurare lo stato occupazionale, per il gruppo A e per il gruppo B. Infine l'ultima colonna riporta la probabilità che le differenze registrate fra gruppo A e gruppo B siano non significative.

*Tabella 5.4: Composizione dei sotto-gruppi per alcune variabili rilevanti e test*

<b>Variabile</b>	<b>Diagonale Principale</b>	<b>Gruppo A</b>	<b>Gruppo B</b>	<b>Test differenze</b>
Femmine (%)	49,7	47,8	48,3	0,912
Età (anni)	27,3	26,8	26,0	0,381
Istruzione (anni)	11,3	11,9	10,9	0,001
Stranieri (%)	24,2	19,9	33,8	0,001***
Obbligo Istruzione (%)	26,8	19,9	30,9	0,010*
Specializzazione (%)	38,6	50,7	36,2	0,003**
Disoccupazione precedente (mesi) – per MdL	2,4	2,1	2,5	0,005**
Controllo (%)	23,8	34,8	17,4	0,000***
N	1615	201	207	-
Mesi lavorati gennaio – ottobre 2012 - survey	4,1	6,9	3,5	0,000***
Mesi lavorati gennaio – ottobre 2012 - COB	3,8	1,6	6,3	0,000***

È interessante notare che i gruppi A e B non si differenziano significativamente per caratteristiche quali genere o età, praticamente equi-distribuite (). Invece, sono significative le differenze per nazionalità, livello di istruzione in ingresso, esperienza pregressa; gli individui del gruppo B, ovvero coloro che si dichiarano non occupati pur comparando nelle COB, sembrano essere soggetti più “deboli” per caratteristiche osservabili rispetto al gruppo A, ovvero coloro che si dichiarano occupati ma non compaiono negli archivi amministrativi. Questi ultimi sono soggetti prevalentemente autoctoni (80,1% vs. 66,2%), significativamente più istruiti (12 vs. 11 anni), con spell di disoccupazione antecedenti il corso mediamente più brevi (2,1 vs. 2,5 mesi). Inoltre, la quota di individui del gruppo A appartenenti al campione di controllo (34,8%) è doppia rispetto al gruppo B (17,4%), mentre nel gruppo B è maggiore la quota di individui in formazione iniziale (30,9% vs. 19,9%).

Chiaramente, queste differenze di composizione si riflettono sui risultati delle analisi di impatto, derivando da caratteristiche risultate significative nello spiegare l’occupabilità individuale secondo il modello multivariato al §6.

### 5.2 Altri indicatori

Estendendo il confronto agli altri indicatori (Tabella 5.5), i risultati non cambiano molto: le correlazioni fra gli indicatori calcolati a partire dalle due fonti evidenziano sempre una notevole discrepanza. In particolare, essa risulta leggermente meno accentuata nel caso del numero di mesi lavorati rispetto all’indicatore dicotomico puntuale. Al contrario, la correlazione è molto più bassa nel caso del numero di contratti, variabile che probabilmente soffre più di altre di difetti di memoria e approssimazione e che non è quindi adeguatamente rilevabile attraverso survey.

Tabella 5.5: Correlazioni fra diversi indicatori

<b>Numero di mesi fra gennaio e ottobre</b>		
	<b>Mesi survey</b>	<b>Mesi COB</b>
<b>Mesi survey</b>	1,0000	
<b>Mesi COB</b>	0,6333	1,0000
<b>Numero di contratti</b>		
	<b>Contratti survey</b>	<b>Contratti COB</b>
<b>Contratti survey</b>	1,0000	
<b>Contratti COB</b>	0,3602	1,0000

A questo proposito, le ultime due righe della Tabella 5.4 riportata più sopra evidenziano una sostanziale coerenza fra l'indicatore istantaneo e quello di "durata" dello stato occupazionale. Nel caso di fonti concordi, lo scarto fra la lunghezza media dello spell occupazionale dell'individuo nel periodo gennaio – ottobre rilevato con survey e con COB ammonta ad appena a una decina di giorni (4,1 vs. 3,8 mesi). Nel caso di fonti incoerenti, invece, la discordanza nella rilevazione della durata va nella stessa direzione indicata dall'indicatore puntuale: gli individui che dichiarano di lavorare a ottobre ma non compaiono nei corrispondenti archivi COB (gruppo A), lavorano mediamente 7 mesi secondo survey e appena 2 secondo COB e viceversa per il gruppo B (3,5 secondo survey e 6,3 secondo COB). Pertanto, lo scostamento evidenziato dall'indicatore dicotomico non è un evento casuale, bensì l'evidenza di un errore sistematico di cui è bene rintracciare le cause per non commettere grossolane imprecisioni nella valutazione degli impatti delle politiche.

## 6. GLI SCOSTAMENTI NEGLI IMPATTI

In questa sezione si valuta lo scostamento nella stima dell'impatto netto delle politiche formative utilizzando le due fonti informative a disposizione. L'analisi è effettuata con un modello probit multivariato, che stima in termini percentuali l'impatto netto della formazione professionale sulla probabilità di essere occupati a circa un anno di distanza, controllando per gli effetti che le caratteristiche individuali osservabili hanno su tale probabilità. Il modello di regressione in Tabella 6.1 stima l'impatto sull'occupabilità individuale di variabili quali il genere, l'età,

l'istruzione, la nazionalità, la durata del periodo di disoccupazione precedente il corso e la formazione stessa, di cui si valuta l'impatto anche in interazione con altre caratteristiche individuali. Al di là dei risultati specifici del modello e dell'interpretazione dei singoli coefficienti, trattati in dettaglio in Ragazzi e Sella (2013b), il confronto fra la regressione su dati survey e la regressione su dati COB evidenzia coefficienti differenti per le singole variabili esplicative e, soprattutto, un impatto netto della formazione molto diverso. In particolare, sui dati COB non si riscontrano più effetti significativi dell'età, del genere e della cittadinanza sull'occupabilità. L'istruzione conserva un effetto positivo ma ridotto (0,049 vs. 0,074), al pari dell'impatto negativo della disoccupazione precedente il corso (-0,111 vs. -0,179). La formazione, invece, conserva e amplifica il suo impatto positivo (0,989 vs. 0,740), mentre si riscontra in ambo i casi la maggior efficacia dell'azione formativa sui soggetti meno istruiti (variabile "formazione\*istruzione"). In particolare, è interessante quantificare lo scostamento nella stima dell'impatto fra le due fonti calcolando l'effetto marginale medio (Average Marginal Effect, AME) della formazione. In pratica, invece del più tradizionale effetto marginale sull'individuo medio, calcolato su un unico individuo che presenti caratteristiche medie per tutte le altre variabili, l'AME calcola la media dell'effetto marginale della formazione su tutti gli individui nel campione, stimando la differenza nella loro occupabilità se formati o non formati, a parità di tutte le altre variabili osservate. Se con dati survey la formazione aumenta mediamente l'occupabilità individuale del 14,5%, nel caso dei dati COB la stima si alza al 23,7%, con un considerevole aumento di impatto di circa 10 punti percentuale.

Tabella 6.1: Analisi multivariata su indicatore dicotomico puntuale

Variabile	Modello su dati survey	Modello su dati COB
Età	0,090***	0,038
Età <sup>2</sup>	-0,001***	-0,001
Femmina	-0,384***	-0,128
Extra-comunitario	-0,291**	0,000
Istruzione (anni)	0,074***	0,049**
Disoccupazione pre (mesi)	-0,179***	-0,111***
Ritirato per lavoro	0,852***	0,796***
Formazione	0,740*	0,989***
Formazione * Istruzione	-0,057**	-0,054**
Formazione * Femmina	0,006	-0,040
Formazione * Extra-comunitario	-0,044	0,064
<b>AME Formazione</b>	<b>0,143***</b>	<b>0,237***</b>

\*  $p < 0,1$  \*\*  $p < 0,05$  \*\*\*  $p < 0,01$  N = 1.485

LR  $\chi^2 = 239,9$

## 7. CONCLUSIONI E INDICAZIONI DI RICERCA FUTURA

Tirando le fila di questa prospezione sull'utilizzo dei microdati estratti dagli archivi dei Centri per l'Impiego, per validarne l'utilizzo a fini di studi sul placement, va innanzitutto richiamata la grossa discrepanza che si registra fra gli indicatori calcolati su dati raccolti attraverso survey e quelli calcolati a partire dalle Comunicazioni Obbligatorie. Nel caso del tasso di occupazione puntuale, si osservano discordanze per il 20% degli individui del campione analizzato.

Questa discrepanza appare evidente solamente se si osservano i microdati individuali. Infatti, calcolando semplicemente il tasso di occupazione aggregato non si riscontrano grandi differenze. Ma l'apparente

uguaglianza nasconde "errori" di segno e peso opposto: coloro che hanno dichiarato di lavorare durante l'intervista ma non compaiono sull'archivio delle COB corrispondono grossomodo numericamente a quanti hanno dichiarato di non essere occupati, pur disponendo di un regolare contratto.

Tale fenomeno è particolarmente insidioso, in quanto a fronte di un'apparente regolarità nella misurazione degli stati occupazionali, il cambiamento di dataset distorce i risultati delle analisi di impatto netto. Infatti, le variabili determinanti dell'occupazione e l'entità dell'impatto netto variano notevolmente. Questo accade perché gli individui nei due sottogruppi con indicatori discordanti si compensano numericamente, ma hanno caratteristiche (osservabili o no) differenti, che incidono sulla probabilità di

trovare lavoro. In particolare, è possibile ipotizzare che nel passaggio da survey a COB degli individui più forti abbiano sostituito gli individui inizialmente (pre-formazione) più deboli per caratteristiche anagrafiche o per altri motivi. Confrontando lo stesso modello stimato sulle COB si osserva infatti un aumento dell'impatto netto registrato, a fronte di una scomparsa dello svantaggio iniziale registrato per giovani, donne e stranieri. Inoltre non si registra più il fenomeno di recupero dello svantaggio, ottenuto attraverso una maggiore efficacia della politica formativa per donne e stranieri.

Questo lavoro iniziale apre quindi a molti interrogativi sull'opportunità di utilizzare gli archivi informatici delle pubbliche amministrazioni, nati con finalità differenti (anagrafiche, fiscali, previdenziali). Almeno per quanto si osserva in Italia, nella fase presente esse hanno raggiunto un livello di apparente maturità. Non si riscontrano infatti più grosse lacune o disomogeneità di copertura, e l'alimentazione delle banche dati pare fluire in modo continuo e con ritardi sempre più contenuti. Inoltre man mano che passa il tempo, vengono implementate nuove funzionalità del sistema integrato, per cui dati che inizialmente venivano caricati manualmente con notevoli possibilità di errore umano, verranno sempre più ricavati automaticamente da altre partizioni del sistema.

In realtà la maturità dei sistemi riguarda prevalentemente la loro funzionalità principale (nel caso delle COB l'interruzione dello stato di disoccupazione registrato presso il centro per l'impiego), ma non altrettanto si può dire per quanto riguarda il loro utilizzo a scopi di analisi, che è finora avvenuto per esperienze sporadiche, non coordinate e,

soprattutto, da parte di valutatori o ricercatori che usufruiscono di "scarichi" ad hoc di dati, ma non si interfacciano con i soggetti che gestiscono e alimentano le banche dati. Accade così che pratiche errate (per esempio errori nell'interpretazione delle codifiche) si perpetuino senza essere diagnosticate o corrette, generando distorsioni non valutabili dall'esterno.

Un altro problema è il differente livello di maturità delle differenti parti che compongono il sistema. In Italia il sistema delle COB è nazionale, sotto l'egida del Ministero del Lavoro, ma la sua creazione è competenza delle Regioni. Inoltre l'alimentazione del sistema è prevalentemente a carico dei Centri per l'Impiego ma alcuni tipi di dati o alcuni record provengono anche da fonti diverse (Province, Prefetture, Agenzie di Lavoro Interinale e datori di lavoro o i loro Consulenti del Lavoro). In sintesi è un sistema la cui potenzialità risiede proprio dall'aver posto in rete molti snodi in cui vengono create o gestite informazioni relative ai rapporti di lavoro. Per il momento però non tutti gli attori coinvolti sono dotati delle stesse competenze per interpretare e veicolare correttamente le informazioni, dello stesso livello di motivazione a svolgere tale compito in modo preciso e, per quanto riguarda i soggetti pubblici, del personale sufficiente e competente per far fronte alle nuove mansioni.

In conclusione si sottolinea dunque che quando si proceda ad utilizzare come fonte per l'analisi socio-economica una banca dati di tipo amministrativo sia indispensabile procedere con cautela e spirito critico, verificando l'attendibilità dei valori che emergono dal suo utilizzo attraverso l'incrocio con altre fonti. In assenza di tale

attività preliminare di validazione, che nel nostro caso deve avvenire con una collaborazione stretta con le autorità regionali e provinciali, non solo si rischia di accettare acriticamente informazioni che, come dimostrato nei paragrafi sopra, portano a distorsioni sistematiche nei risultati, ma si impedisce un lavoro di correzione delle procedure e delle pratiche che è indispensabile per la messa a punto del sistema. Un'analisi come quella qui presentata può fornire i campanelli di allarme necessari per attirare l'attenzione sui problemi. Il passo successivo è quello di identificare le cause che originano gli scostamenti, soprattutto se di tipo sistematico, in modo da poter agire su prassi errate o da poter progettare controlli incrociati che evidenzino le anomalie già in fase di caricamento.

Nei sistemi complessi, quale è un sistema informativo che integri molte fonti informative differenti, i singoli operatori hanno una visibilità limitata alla parte con cui si interfacciano, e si muovono con logiche, conoscenze, competenze e motivazioni differenti. E' dunque da un lato normale che nessuno, o quasi, abbia la percezione di eventuali distorsioni né il potere o la capacità di correggerli quando anche se ne renda conto. Ma proprio per questo è indispensabile che, conclusa la fase di avvio e di messa a regime, venga realizzata una diagnosi approfondita della portata informativa complessiva, e poi venga individuato un regista in grado di riformare le anomalie che si sono create.

Il lavoro di ricerca sta attualmente proseguendo con ulteriori approfondimenti sul caso presentato, volti sia a comprendere in modo più approfondito le caratteristiche della discrepanza (continuando quindi il confronto fra le basi dati con altri indicatori e modelli),

ma soprattutto a comprenderne le cause, in modo da conoscere i limiti di ciascuna delle fonti di informazione e destinarne l'utilizzo ai fenomeni che meglio è in grado di descrivere.

## BIBIOGRAFIA

- Bachtler J., Wren C. (2006) "Evaluation of European Union Cohesion policy: Research questions and policy challenges", *Regional Studies*, 40:02, 143-153.
- Barca F. (2009), *An agenda for a reformed cohesion policy. A place-based approach to meeting European Union challenges and expectations*. Independent Report prepared at the request of Danuta Hübner, Commissioner for Regional Policy [http://ec.europa.eu/regional\\_policy/archive/policy/future/pdf/report\\_barca\\_v0306.pdf](http://ec.europa.eu/regional_policy/archive/policy/future/pdf/report_barca_v0306.pdf)
- Baslé, M. (2006). Strengths and weaknesses of European Union policy evaluation methods: Ex-post evaluation of objective 2, 1994-99. *Regional Studies*, 40(2), 225.
- Bell S., Orr L., Blomquist J., Cain G. (1995) *Program Applicants as a Comparison Group in Evaluating Training Programs: Theory and a Test*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Benati I., Ragazzi E., Sella L. (2013), *Valutare l'impatto della formazione professionale sull'inserimento lavorativo: lezioni da una ricerca in Regione Piemonte*. Rivista Italiana di Valutazione, in press.
- Canu E., Conzimu L., Garau G. (2010) Le politiche per il lavoro della Regione Sardegna. *Formazione Orientamento Professionale*, 10 (5-6): 123-146.
- Cochran W. (1977) *Sampling Techniques, III ed.* New York: John Wiley & Sons.
- Commission of the European Communities. (2008). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the results of the negotiations concerning cohesion policy strategies and programmes for the programming period 2007-2013*. COM(2008) 301/4, Brussels.
- European Commission - DG for Regional Policy (2013). Evalsed – the guide. The resource for the Evaluation of Socio-Economic Development. [http://ec.europa.eu/regional\\_policy/information/evaluations/guidance\\_en.cfm](http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm)
- European Commission - DG for Regional Policy (2013). Evalsed sourcebook: methods and techniques. [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/evaluation\\_sourcebook.pdf](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf)
- Falavigna G., Ragazzi E., Sella L., (2013). *Parcours formatifs et entrée dans le marché du travail pour les femmes. Analyse des différences homme-femme dans le cas des politiques de formation en Piémont*, CNR-Ceris Working Paper, WP 13/13.
- Hotz V., Goerge R., Balzekas J., Margolin F. (1999). *Administrative data for policy-relevant research: assessment of current utility and recommendations for development*. Report of the Advisory Panel on Research Uses of Administrative Data. Joint Center for Poverty Research of the Northwestern University/University of Chicago. [http://public.econ.duke.edu/~vjh3/working\\_papers/adm\\_data.pdf](http://public.econ.duke.edu/~vjh3/working_papers/adm_data.pdf)
- Morton M.H (2009) *Applicability of Impact Evaluation to Cohesion Policy*, working paper connesso al report "An Agenda for a reformed Cohesion" [http://ec.europa.eu/regional\\_policy/archive/policy/future/pdf/4\\_morton\\_final-formatted.pdf](http://ec.europa.eu/regional_policy/archive/policy/future/pdf/4_morton_final-formatted.pdf)

- Ragazzi E. (a cura di), Benati I., Santanera E., Sella L. (2013) Gli esiti occupazionali delle politiche formative in Piemonte. 2° Rapporto Annuale di Placement 2012. Indagine su qualificati e specializzati nell'anno 2011. Moncalieri: CNR-Ceris e Regione Piemonte, in press.
- Ragazzi E. (a cura di) (2008). *Perché nessuno si perda. La Piazza dei Mestieri, un modello per contrastare la dispersione scolastica*. Milano, Guerini e Associati.
- Ragazzi E., (2013). *Effectiveness evaluation of training programmes for disadvantaged targets*, Procedia - Social and Behavioral Sciences, in press.
- Ragazzi E. (a cura di), Nosvelli M., Sella L., (2012). *Gli esiti occupazionali delle politiche formative in Piemonte – 1° rapporto annuale di placement 2011 – Indagine su qualificati e specializzati nell'anno 2010*, CNR-Ceris e Regione Piemonte, Torino, ISBN 978-88-98193-01-1
- Ragazzi E., Sella L. (2011), *L'Assetto Istituzionale dell'Istruzione Professionale in Italia*, in Lauro e Ragazzi 2011.
- Ragazzi E., Sella L.,(2013a) *L'efficacité des politiques de formation: méthodes pour l'évaluation de l'impact*, CNR-Ceris Working Paper, WP 14/13.
- Ragazzi E., Sella L.,(2013b) *Una valutazione di impatto delle politiche formative regionali: il caso piemontese*, CNR-Ceris Working Paper, WP 15/13. Paper presentato alla XXXIV Conferenza Italiana Scienze Regionali, Palermo, 2-3 settembre 2013. Sessione SP35 – Valutazione di politiche, piani e programmi.
- Ragazzi E., Sella L.,(2013c). *Migration and work: the cohesive role of vocational training policies*, CNR-Ceris Working Paper, WP 16/13.
- Ragazzi E., Vitali G. (2005), *I fabbisogni formativi dei soggetti deboli*, Franco Angeli, Milano.
- Sella L., (2013). *Enhancing Vocational Training Effectiveness Through Active Labour Market Policies*, Procedia - Social and Behavioral Sciences, in press.



*Download*

[www.ceris.cnr.it/index.php?option=com\\_content&task=section&id=4&Itemid=64](http://www.ceris.cnr.it/index.php?option=com_content&task=section&id=4&Itemid=64)

Hard copies are available on request,  
**please, write to:**

Cnr-Ceris  
Via Real Collegio, n. 30  
10024 Moncalieri (Torino), Italy  
Tel. +39 011 6824.911 Fax +39 011 6824.966  
[segreteria@ceris.cnr.it](mailto:segreteria@ceris.cnr.it) [www.ceris.cnr.it](http://www.ceris.cnr.it)

**Copyright © 2014 by Cnr-Ceris**

All rights reserved. Parts of this paper may be reproduced with the permission  
of the author(s) and quoting the source.