

n. 10/2007

**Seminario sulla qualità: l'esperienza dei referenti
del sistema informativo SIDI - 2^a giornata**

AA. VV.

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei *Contributi ISTAT* e nei *Documenti ISTAT* sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella *Rivista di Statistica Ufficiale* sono subordinati al giudizio di referee esterni.

Direttore responsabile della Rivista di Statistica Ufficiale: Patrizia Cacioli

Comitato di Redazione delle Collane Scientifiche dell'Istituto Nazionale di Statistica

Coordinatore: Giulio Barcaroli

Membri:	Corrado C. Abbate	Rossana Balestrino	Giovanni A. Barbieri
	Giovanna Bellitti	Riccardo Carbini	Giuliana Coccia
	Fabio Crescenzi	Carla De Angelis	Carlo M. De Gregorio
	Gaetano Fazio	Saverio Gazzelloni	Antonio Lollobrigida
	Susanna Mantegazza	Luisa Picozzi	Valerio Terra Abrami
	Roberto Tomei	Leonello Tronti	Nereo Zamaro

Segreteria: Gabriella Centi, Carlo Deli e Antonio Trobia

Responsabili organizzativi per la *Rivista di Statistica Ufficiale*: Giovanni Seri e Carlo Deli

Responsabili organizzativi per i *Contributi ISTAT* e i *Documenti ISTAT*: Giovanni Seri e Antonio Trobia

CONTRIBUTI ISTAT

n. 10/2007

**Seminario sulla qualità: l'esperienza dei referenti
del sistema informativo SIDI - 2^a giornata**

AA. VV.

Contributi e Documenti Istat 2007

Istituto Nazionale di Statistica
Servizio Produzione Editoriale

Produzione libraria e centro stampa:
Carla Pecorario
Via Tuscolana, 1788 - 00173 Roma

Indice

I sessione: La valutazione degli strumenti di indagine attraverso gli indicatori di qualità

Il comportamento degli intervistatori osservato nelle edizioni trimestrali dell'indagine Cati "Viaggi, vacanze e vita quotidiana"

Barbara Dattilo pag. 5

Analisi dell'effetto intervistatore nella fase di controllo e correzione

Alessandro Martini pag. 21

II sessione: Innovazioni di processo 1

Integrazione tra sistemi di gestione di indagine e il Sistema di Documentazione delle Indagini SIDI: l'esperienza del servizio delle statistiche strutturali sulle imprese

G. Brancato, C. Casciano, C. Pellegrini, C. Schiattone pag. 33

Misure della qualità dei dati per una statistica di fonte amministrativa: i condannati con sentenza irrevocabile

Daniela Capelli, Laura Rossetti, Franco Turetta pag. 49

Gli indicatori di copertura e mancata risposta nella nuova indagine campionaria sui servizi idrici

Grazia Di Bella, Silvia Capuano, Marilena Angela Ciarallo, Antonino Laganà, Simona Ramberti pag. 69

III sessione: Innovazioni di processo 2

La nuova Indagine sulle unità locali delle grandi imprese: contributi al miglioramento della qualità dell'informazione statistica sulla struttura delle imprese.

Luigi Costanzo, Barbara Gentili, Serena Migliardo pag. 87

L'innovazione del sistema IVR per il monitoraggio dell'indagine multiscopo "Condizioni di salute e ricorso ai servizi sanitari 2004-2005"

Laura Iannucci, Manuela Michelini, Giuseppina Pica pag. 101

Il sistema SIDI: uno strumento per le analisi di qualità

Giovanna Brancato, Riccardo Carbini, Concetta Pellegrini, Marina Signore, Giorgia Simeoni pag. 115

Il comportamento degli intervistatori osservato nelle edizioni trimestrali dell'indagine Cati "Viaggi, vacanze e vita quotidiana"

Barbara Dattilo

Istat, DPTS/DCCV/SDS/B

1. Introduzione

A partire dal 1997 l'Istat conduce l'indagine "Viaggi, vacanze e vita quotidiana" per soddisfare l'esigenza di raccogliere sistematicamente dati sulla domanda turistica e garantire, insieme ai dati sull'offerta, la disponibilità di un sistema integrato di informazioni statistiche sul turismo. Le informazioni raccolte, inoltre, mirano ad ottemperare la Direttiva europea 95/57/CE, nata con l'obiettivo di raccogliere e diffondere dati statistici armonizzati in ambito europeo sull'offerta e la domanda nel settore del turismo.

L'indagine è condotta trimestralmente con tecnica Cati (Computer Assisted Telephone Interview) su un campione annuo di 14.000 famiglie (3.500 per trimestre) corrispondenti a circa 40.000 individui (circa 10.000 per trimestre). I periodi di riferimento sono i trimestri gennaio-marzo, aprile-giugno, luglio-settembre e ottobre-dicembre. Al fine di quantificare e analizzare i flussi turistici dei residenti in Italia molte sono le informazioni rilevate sul fenomeno, come la destinazione e le modalità di effettuazione del viaggio oppure le caratteristiche socio-demografiche delle persone che hanno viaggiato.

Per svolgere l'indagine, l'Istat si avvale di una società privata che realizza le interviste e mette a disposizione gli intervistatori. Dal 1997 più società si sono susseguite aggiudicandosi per un periodo generalmente triennale l'appalto con l'Istituto. Ne consegue che sia la continuità con la quale è condotta l'indagine sia il vincolo contrattuale con la società esterna, portano a mantenere gli stessi intervistatori per più edizioni di indagine. In particolare, dal 2001 l'Istat collabora con la stessa società che fino ad oggi ha curato due edizioni triennali dell'indagine. Ciò ha permesso di avere una continuità nella modalità di esecuzione del lavoro - che dura ormai da quasi sei anni - nonché una disponibilità degli intervistatori pluriennale che, tuttavia, nel tempo ha risentito di un parziale ricambio degli stessi nel corso delle varie edizioni dovuto a diverse motivazioni (abbandono spontaneo dell'attività da parte dei rilevatori o sostituzione con personale più idoneo).

Questa modalità di organizzazione del lavoro comporta per l'Istat un'intensa attività di monitoraggio della fase di rilevazione, che viene realizzata sia presso la società che cura la rilevazione sia presso il servizio tecnico competente per l'indagine.

L'attività di monitoraggio dei rilevatori ha spinto ad approfondire studi mirati a valutare il loro lavoro. È stato osservato che gli intervistatori sviluppano diverse linee di condotta nel gestire l'intervista, in parte attribuibili alle peculiarità e capacità di ciascun rilevatore, in parte, riconducibili, presumibilmente, al tempo di permanenza di uno stesso rilevatore su più occasioni d'indagine.

2. Gli intervistatori: uno strumento prezioso per la qualità del dato

Come è noto, una delle fasi più delicate per l'indagine è quella legata all'operato dei rilevatori scelti per la raccolta dati. Infatti, un elevato livello di qualità delle interviste è garanzia di risultati coerenti e affidabili. Tale obiettivo viene perseguito inserendo i rilevatori in un processo formativo complesso che mira sia a rendere gli intervistatori edotti sui contenuti, le classificazioni, le definizioni dell'indagine sia a far comprendere loro l'importanza di ridurre il rischio di rifiuto e di interruzione definitiva dell'intervista. A tal fine, mediante la formazione articolata in briefing teorici, tecnici e pratici vengono impartite le istruzioni sull'argomento d'indagine e sulle regole comportamentali idonee a far acquisire agli intervistatori quelle capacità di comunicazione e persuasione che consentiranno loro tanto di catturare l'attenzione del rispondente quanto di contenere gli esiti succitati.

L'operato dei rilevatori è sottoposto ad un accurato e complesso processo di monitoraggio che è articolato in più fasi.

Una fase è l'analisi dei *reports*, cioè degli indicatori, delle distribuzioni di frequenza delle variabili rilevate dei *files* dati sugli esiti dei contatti telefonici che quotidianamente vengono prodotti dall'attività di rilevazione e che la società provvede ad inviare all'Istat.

Un'altra fase consiste nel seguire lo svolgimento delle interviste durante la rilevazione al fine di osservare più elementi: il comportamento dei rilevatori assunto nei confronti delle famiglie intervistate, la capacità di gestire il contatto telefonico e l'abilità pratica di utilizzare il questionario elettronico. Da questi fattori è possibile capire se eventuali segnali di *alert* che provengono da alcuni degli indicatori giornalieri, siano dovuti a incomprensioni sui contenuti o sui comportamenti da tenere in corso d'intervista o siano, invece, il frutto della messa a punto di strategie d'intervista volutamente deviate.

Infine, il piano di monitoraggio prevede l'effettuazione di telefonate di controllo da parte del personale dell'istituto su un sottogruppo di famiglie (circa il 20%) che hanno portato a termine l'intervista. Tra i vari obiettivi di questo strumento di monitoraggio, c'è quello di appurare l'effettiva realizzazione dell'intervista telefonica e di controllare le modalità di conduzione della stessa, verificando sia l'esistenza delle informazioni per tutte le sezioni del questionario e per tutti i componenti della famiglia, sia la bontà di applicazione e il rispetto delle norme comportamentali impartite ai rilevatori durante i *briefing*.

3. Gli indicatori per la misura del rendimento degli intervistatori

Al termine di ogni rilevazione trimestrale con l'obiettivo di monitorare la qualità del dato anche secondo gli standard del sistema SIDI, oltre alla disamina quotidiana dei reports forniti dalla società, sono calcolati gli aggregati che successivamente intervengono nel calcolo degli indicatori di copertura e mancata risposta totale secondo SIDI.

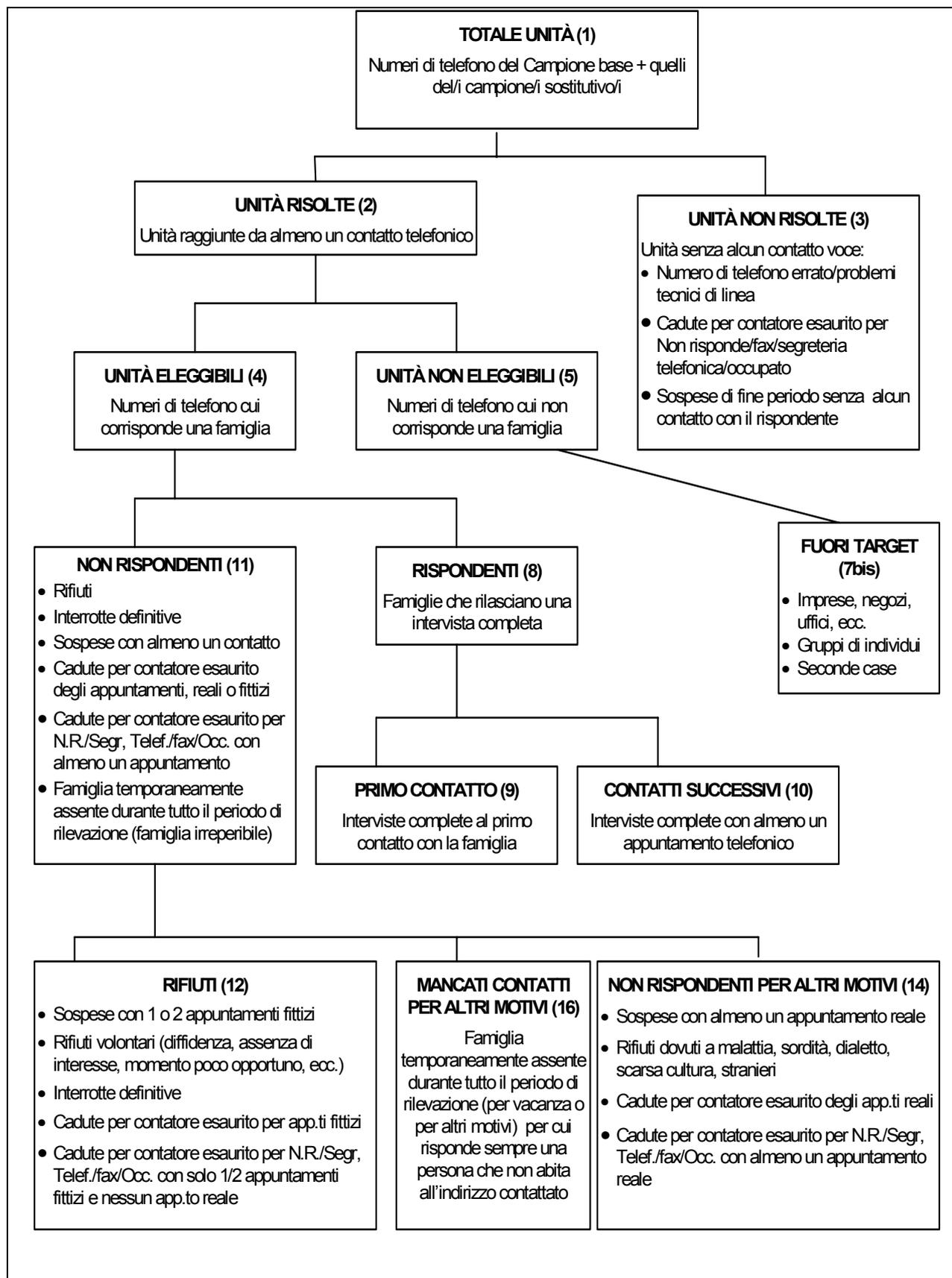
Rispetto alle indicazioni generali fornite da SIDI per il calcolo di tali aggregati, è stato necessario adattare le definizioni previste da SIDI alle specificità dell'indagine in oggetto sia per quanto riguarda la tecnica di indagine utilizzata (telefonica Cati), sia per quanto riguarda alcuni aspetti del disegno campionario e delle modalità di gestione dei contatti telefonici (figure 3.1 e 3.2).

La disponibilità congiunta degli indicatori calcolati per SIDI insieme a quelli prodotti dalla ditta per il monitoraggio quotidiano della rilevazione contribuisce al controllo di qualità di fine rilevazione.

Nel presente lavoro, al fine di poter utilizzare gli indicatori SIDI anche per il controllo dell'operato dei singoli rilevatori, si è proceduto a calcolarli per ciascun intervistatore, andando ad integrare quelli già previsti nella metodologia di monitoraggio dell'indagine.

Gli indicatori utilizzati sono riportati nelle tabelle che seguono insieme alla descrizione di come sono calcolati.

Figura 3.2 – Adattamento della classificazione SIDI delle unità secondo gli aggregati utili per il calcolo degli indicatori di copertura e mancata risposta totale al caso dell’Indagine “Viaggi, vacanze e vita quotidiana”.



In particolare, gli indicatori SIDI analizzati in questo lavoro sono riportati nella tabella 3.1.

Tabella 3.1- Indicatori SIDI

Indicatore	Descrizione delle quantità che lo compongono	Elementi del grafo in figura 1	Nome etichetta
Tasso di Risposta	unità rispondenti / (unità eleggibili + unità non risolte)	(8) / [(4)+(3)]	COM
Tasso di Rifiuto	unità che si rifiutano / unità eleggibili	(12) / (4)	RIFV
Tasso di Mancata Risposta per Altri Motivi (esclusi rifiuti e mancati contatti)	unità non rispondenti per motivi diversi dai rifiuti e dai mancati contatti / unità eleggibili	(14) / (4)	RIF_NV

Gli indicatori della metodologia di monitoraggio dell'indagine scelti per effettuare le analisi sulla *performance* degli intervistatori sono riportati nelle tabelle 3.2 e 3.3.

Tabella 3.2 – Indicatori della metodologia di monitoraggio dell'indagine “Viaggi, vacanze e vita quotidiana”

Indicatore	Descrizione delle quantità che lo compongono	Nome etichetta
Tasso di mancato contatto	unità eleggibili con o senza un appuntamento cadute per contatore esaurito di N.R.-Segr. Telef.-Occ. / (unità eleggibili + unità non risolte)	MC
Tasso di mancato contatto senza appuntamento	unità eleggibili cadute per contatore esaurito di N.R.-Segr. Telef.-Occ. senza un appuntamento / (unità eleggibili + unità non risolte)	MCNAPP
Tasso di mancato contatto dopo appuntamento	unità eleggibili cadute per contatore esaurito di N.R.-Segr. Telef.-Occ. con almeno un appuntamento / (unità eleggibili + unità non risolte)	MCAPP
Tasso appuntamento	totale appuntamenti fissati / totale dei tentativi di contatto	APP

Tabella 3.3 – Tassi sulle durate

Indicatore	Descrizione delle quantità che lo compongono	Nome etichetta
Durata media delle interviste complete (in minuti)	Somma delle durate delle interviste complete / numero complessivo di interviste complete	DCOM
Durata media (in minuti) delle interviste rifiutate e delle interviste con mancata risposta per altri motivi (esclusi rifiuti e mancati contatti)	Somma delle durate delle interviste rifiutate e con mancata risposta per altri motivi / numero complessivo di interviste rifiutate e con mancata risposta per altri motivi	DRIF
Durata media delle interviste rifiutate (in minuti)	Somma delle durate delle interviste rifiutate / numero complessivo di interviste rifiutate	DRIFV

Durata media (in minuti) delle interviste con mancata risposta per altri motivi (esclusi rifiuti e mancati contatti)	Somma delle durate di ciascuna intervista con mancata risposta per altri motivi / numero complessivo di interviste con mancata risposta per altri motivi	DRIF_NV
--	--	---------

Per misurare la partecipazione al lavoro, per ciascun intervistatore, sono stati calcolati i seguenti tassi *ad hoc* della tabella 3.4:

Tabella 3.4 – Tassi sulla partecipazione al lavoro

Indicatore	Descrizione delle quantità che lo compongono	Nome etichetta
Tasso di assiduità al lavoro (in percentuale)	Presenze degli intervistatori (in giorni)/ totale dei giorni di lavorazione	ASS
Tasso medio di produttività (in unità)	Interviste complete / Presenze degli intervistatori (in giorni)	PROD

La qualità dei dati raccolti dagli intervistatori è stata verificata scegliendo di costruire altri due indicatori *ad hoc* relativi al contenuto del questionario. In particolare sono state scelte due variabili ritenute in grado di misurare in modo sintetico ma significativo sia il comportamento dell'intervistatore sia la qualità di quanto rilevato. Sulle interviste complete, per ciascun intervistatore, sono stati calcolati i seguenti due indicatori:

- numero medio di componenti della famiglia (NCOMP);
- numero di schede viaggio descritte (NSKVGG).

Per comprendere appieno l'importanza del secondo dei due indicatori succitati è necessario definire il concetto di 'scheda viaggio'. La compilazione delle schede viaggio, infatti, costituisce la parte centrale del questionario, nella quale sono raccolti i dati relativi ad ogni viaggio effettuato da ciascun componente della famiglia. Dal punto di vista dell'intervistatore questa parte del questionario è quella che forse comporta il maggior impegno perchè la scheda prevede la descrizione di tutte le caratteristiche del viaggio effettuato e, di conseguenza, richiede di sottoporre numerosi quesiti al rispondente. Il lavoro del rilevatore tuttavia è facilitato dal meccanismo di gestione delle schede viaggio all'interno del Cati. Infatti se più familiari hanno viaggiato insieme, cioè hanno effettuato un viaggio con le stesse caratteristiche, non verrà descritto lo stesso viaggio tante volte quanti sono i partecipanti al viaggio. Tutte le informazioni contenute nella scheda viaggio già compilata per un membro della famiglia che ha svolto un viaggio insieme ad altri familiari (tale persona viene chiamata 'titolare' del viaggio) vengono automaticamente riportate a posteriori dal sistema Cati anche per gli altri partecipanti al viaggio (chiamati 'aggregati' al viaggio), per ciascuno dei quali verrà creata una scheda viaggio identica a quella appena descritta.

Questo meccanismo implica che una misura idonea del lavoro svolto dall'intervistatore è data dal numero di schede viaggio realmente descritte e non dal numero di viaggi complessivamente risultanti dall'intervista.

4. Primi risultati dell'analisi degli indicatori

Il periodo scelto per effettuare una prima analisi descrittiva sulla *performance* degli intervistatori va dal primo trimestre del 2001 al terzo trimestre del 2006 (ultimo trimestre disponibile) durante il quale i rilevatori sono stati messi a disposizione da una stessa società. Ciò ha consentito di mantenere costanti alcune condizioni che, invece, se fossero cambiate avrebbero potuto incidere sulla produttività dell'indagine e non consentire la piena confrontabilità in un periodo che abbraccia due contratti continui.

Complessivamente, dunque, è stato possibile osservare la *performance* di 272 intervistatori suddivisi in 23 trimestri di rilevazione. È da notare che il numero di rilevatori convocati dalla società e successivamente formati dal personale Istat è stato superiore a quello composto da quanti effettivamente hanno partecipato alla rilevazione. Infatti nel periodo considerato sono stati formati circa 350 intervistatori, dei quali circa il 22% ha abbandonato, durante o a fine formazione, ancor prima di iniziare a lavorare. Tra gli intervistatori che invece hanno partecipato alla rilevazione, ciascuno ha dato la sua disponibilità di trimestre in trimestre per cui il periodo di collaborazione, in termini di presenza per trimestre, è stato variabile da intervistatore ad intervistatore. Alcuni hanno partecipato ad una sola edizione, altri a più edizioni, non necessariamente consecutive, pochi altri ancora sono stati sempre presenti sin dalla prima edizione del 2001.

Una prima analisi della *performance* degli intervistatori è stata fatta calcolando i rapporti di correlazione tra gli indicatori descritti nel paragrafo 3, osservati sul complesso degli intervistatori e sull'intero periodo. Si è poi analizzato l'andamento nel tempo di alcuni indicatori, osservato per le 23 edizioni di indagine.

Sulla base di queste prime analisi, dell'esperienza del monitoraggio in sala di rilevazione presso la società e considerando i risultati di approfondimenti effettuati per alcune edizioni dell'indagine, si è ritenuto opportuno approntare un'analisi che tentasse di evidenziare un'evoluzione temporale del comportamento di alcuni intervistatori seguiti in più edizioni consecutive.

4.1 Analisi dei rapporti di correlazione tra i principali indicatori

Da una prima analisi sulle correlazioni tra gli indicatori sono emerse le seguenti considerazioni.

- Il tasso di appuntamento è positivamente correlato sia con il tasso di risposta sulle interviste complete a seguito di almeno un appuntamento sia con il tasso di mancata risposta per altri motivi. Nel primo caso ciò sta ad indicare, in generale, un buon utilizzo dello strumento appuntamento perché mediante esso gli intervistatori riescono a trasformare molti contatti telefonici in interviste complete. Quindi il tempo speso nella fase iniziale dell'intervista da parte dei rilevatori, anche solo per ottenere un appuntamento, risulta un buon investimento perché in grado di aumentare il numero di complete. Nel secondo caso, il fatto che il tasso di appuntamento sia positivamente correlato anche con il tasso di mancata risposta per altri motivi, sta ad indicare che molti degli appuntamenti presi sono legati a situazioni non ostili nei confronti dell'indagine, dovute a problemi di lingua o dialetto, malattia, lutto, impossibilità a ricontattare la famiglia dopo aver concordato con essa uno o più appuntamenti. Data la rilevanza dello strumento 'appuntamento', durante la formazione molte energie vengono spese per cercare di far capire agli intervistatori l'importanza di usare anche strategicamente tale strumento. Infatti la valutazione congiunta per ogni intervistatore del tasso di appuntamento, di rifiuto e di sospensione¹ durante la rilevazione, consente di misurare la qualità del lavoro svolto

¹ Il tasso di sospensione è un altro degli indicatori previsti dalla metodologia di monitoraggio dell'indagine dato dal rapporto tra le interviste senza un esito definitivo (sospese per appuntamento, per non risponde, occupato, ecc.) e il totale dei contatti telefonici.

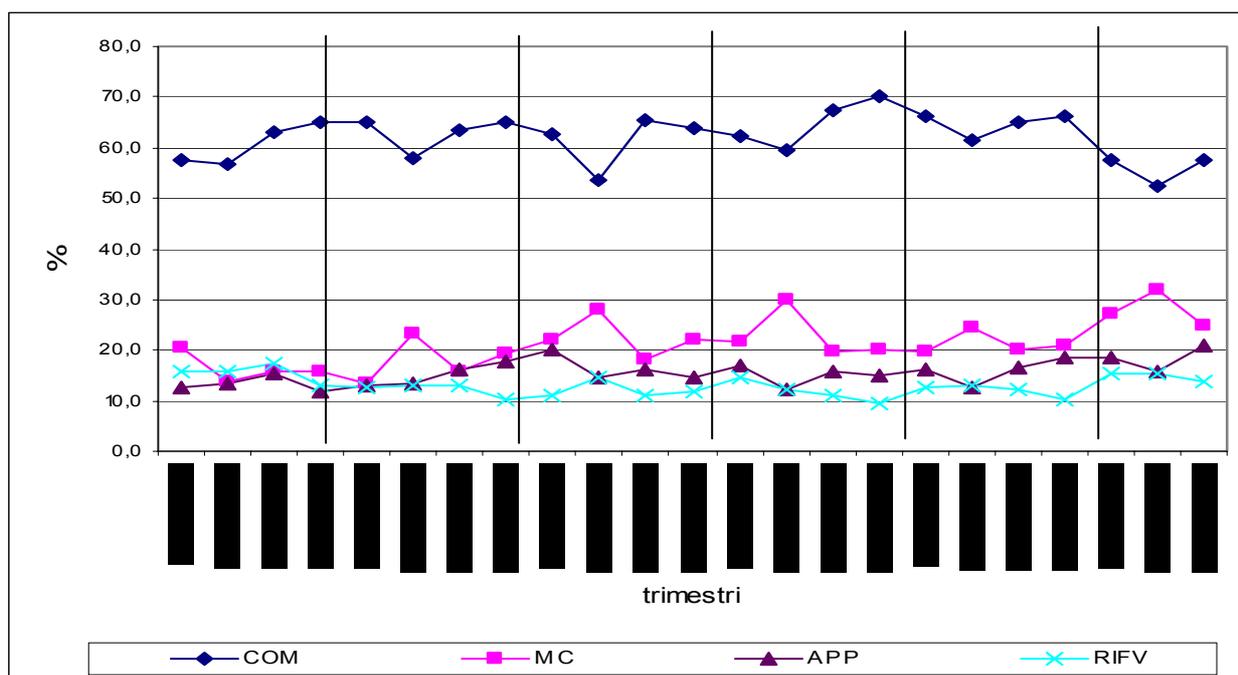
dagli intervistatori: situazioni che presentano un basso livello sia nel tasso di appuntamento che di sospensione accompagnate da un alto tasso di rifiuto denotano in genere uno scarso utilizzo dell'appuntamento con la famiglia.

- Il tasso di mancato contatto dopo appuntamento risulta correlato positivamente con il tasso di mancata risposta per altri motivi. Così come nel caso della correlazione tra il tasso di appuntamento e il tasso di mancata risposta per altri motivi, ciò lascia intendere che dietro questo tipo di esito vi siano delle 'mancate risposte' dovute più alla 'non volontarietà' dei rispondenti, cioè legate ad eventi gravi come la malattia o il lutto familiare oppure alla impossibilità a ricontattare la famiglia, piuttosto che ad una precisa volontà di rifiutare l'intervista. Questa relazione tra tassi probabilmente è da imputare ad un comportamento complessivamente corretto degli intervistatori durante il primo contatto con la famiglia.
- Le durate del rifiuto e della mancata risposta per altri motivi sono negativamente correlate con i tassi di assiduità al lavoro e produttività degli intervistatori seppure in misura piuttosto debole. Ciò lascia intendere che una maggiore presenza dei rilevatori e un maggior numero di interviste complete effettuate probabilmente aiutano l'intervistatore ad acquisire capacità nel condurre l'intervista, consentendogli di gestire al meglio rifiuti e mancate risposte per altri motivi. Tuttavia è proprio dietro tali capacità che possono consolidarsi, in alcuni casi, comportamenti non idonei. La successiva analisi temporale su un sottogruppo scelto di intervistatori mira proprio a verificare eventuali elementi di criticità nella loro *performance* (cfr. par. 5).

4.2 Analisi descrittiva dei principali indicatori nel tempo.

Dopo una prima fase che abbraccia tutto il 2001 e parte del 2002, durante la quale non si osservano andamenti particolarmente caratterizzati dei principali tassi, ad eccezione di una crescita che si potrebbe definire fisiologica del tasso di risposta, a partire dal 2° trimestre 2002 si osservano andamenti abbastanza ciclici dei tassi. In particolare, il tasso di rifiuto e il tasso di appuntamento si compensano ciclicamente ogni tre trimestri assumendo approssimativamente lo stesso valore nel 2° trimestre di ogni anno (Figura 4.1). In questo trimestre il tasso di appuntamento assume il valore più basso in ogni anno, mentre il tasso di rifiuto quello più alto. Di conseguenza, sempre nello stesso trimestre, il tasso di risposta assume quello più basso. Sempre nel secondo trimestre, il tasso di mancato contatto (con o senza appuntamento) assume il valore più elevato. Plausibilmente ciò può essere dovuto al fatto che la rilevazione del secondo trimestre è effettuata nel mese di luglio, mese durante il quale le famiglie sono irreperibili per tutto il periodo di rilevazione. Considerata la stagionalità e la natura stessa del fenomeno e tenendo presente gli studi fatti nel 1996 durante la fase di progettazione dell'indagine per individuare la lunghezza ottimale del periodo di riferimento tenendo conto anche di possibili distorsioni causate dall'effetto memoria, i migliori periodi di riferimento scelti corrispondono ai trimestri solari, mentre il migliore periodo di rilevazione è dato dalle 3 settimane immediatamente successive a ciascun trimestre di riferimento. Di conseguenza qualunque altro periodo di riferimento e/o di rilevazione non apporterebbe un miglioramento né in termini di qualità del dato né in termini di qualità del processo. Negli altri periodi di osservazione invece, il tasso di appuntamento e di rifiuto tendono ad opporsi, con elevati valori del primo e bassi valori del secondo, come di norma ci si aspetterebbe a seguito di una corretta gestione dell'intervista.

Figura 4.1 – I principali indicatori sulla rilevazione



Sulle durate medie delle interviste complete influiscono diversi fattori. Primo fra tutti la lunghezza del questionario che dipende sia dalla stagionalità del fenomeno sia dal numero di quesiti in esso inclusi.

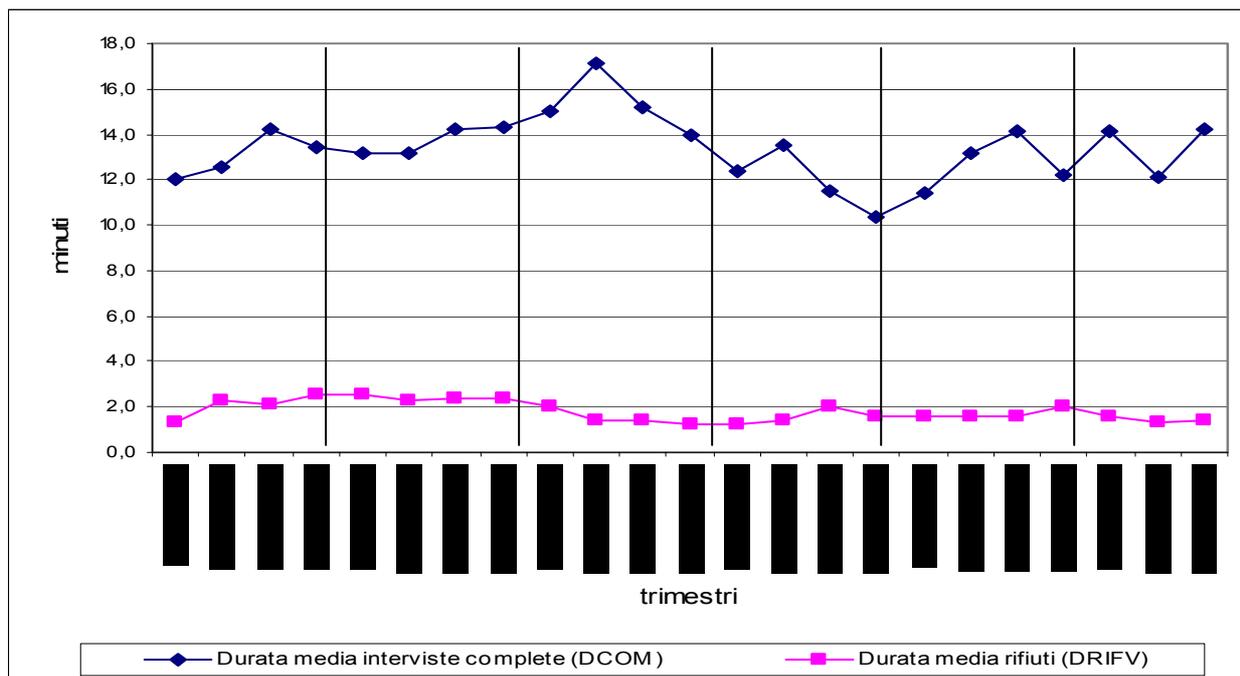
Il turismo, infatti, essendo un fenomeno fortemente stagionale è molto praticato soprattutto nel terzo trimestre dell'anno. Ciò comporta per questo trimestre un maggior impegno sia da parte delle famiglie sia da parte degli intervistatori nella fase di rilevazione.

Riguardo il numero dei quesiti inclusi nel questionario, si evidenzia che oltre alle sezioni dedicate al turismo (che in ogni trimestre comunque subiscono delle variazioni – seppur contenute - per la modifica e/o l'aggiunta/eliminazione di nuovi quesiti), il questionario include una sezione 'open' di carattere sociale dedicata ad altri aspetti della vita quotidiana. Di trimestre in trimestre, in questa parte del questionario, si sono succeduti diversi argomenti, che hanno affrontato, anche contemporaneamente, tematiche differenti: gli incidenti domestici, il ricorso a servizi ospedalieri, le visite mediche e gli accertamenti diagnostici, l'utilizzo di *pc* e internet, i luoghi, le quantità e la qualità degli acquisti di beni non durevoli, gli aiuti ricevuti e gli aiuti dati, il non utilizzo di internet, il fumo, l'acquisto di beni durevoli, le spese per assicurazioni e comunicazioni, la condizione di salute. Conseguentemente l'inclusione o meno di una o più di queste sezioni ha inciso sulla durata dell'intervista.

Un altro fattore che influenza la durata dell'intervista è sicuramente la numerosità della famiglia intervistata. Infatti l'intervista raccoglie informazioni per ciascun componente della famiglia per cui maggiore è la numerosità della famiglia e maggiore è il tempo richiesto per completarla, soprattutto se vi sono componenti che hanno viaggiato e ciascuno di essi ha effettuato numerosi viaggi.

Ultimo, ma non minore intermini di importanza, è il fattore 'intervistatore' che, in base alla propria capacità e personalità, influisce notevolmente sulla durata del contatto telefonico con la famiglia.

Figura 4.2 – Durate medie dei contatti con le famiglie

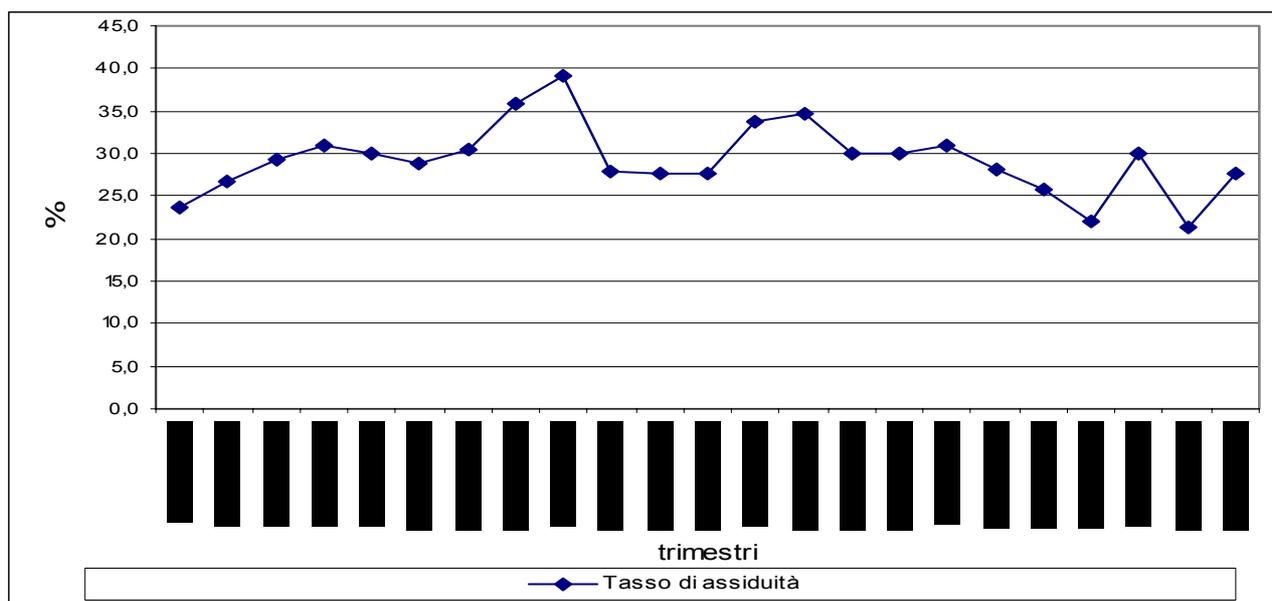


La durata media delle interviste complete è compresa, nel tempo, approssimativamente tra i 10 e i 17 minuti (Figura 4.2).

La stagionalità del fenomeno spiega i picchi osservati nelle durate medie delle interviste complete nel terzo trimestre di quasi ciascun anno. Invece, le maggiori durate osservate per tutto il 2003, in particolare nel 2° trimestre, sono l'effetto dell'introduzione di più sezioni di carattere sociale nel questionario.

La durata media dei rifiuti è approssimativamente costante in tutti i periodi osservati, con valori di poco superiori ai due minuti nel biennio 2001-2002 e, a partire dal 2003, stabili su valori inferiori ai due minuti.

Figura 4.3 – Tasso di assiduità media al lavoro



Analizzando il tasso di assiduità nel tempo, si osserva un andamento irregolare: in crescita fino al 1° trimestre 2003, poi tendenzialmente in diminuzione, con riprese in alcuni trimestri di ciascun anno (Figura 4.3). Ciò potrebbe dipendere dall'introduzione di nuovi intervistatori che integrano i momenti di "stanchezza" di quelli più esperti. Di fatto nel 2001 e nel 2002 c'è stato un continuo rinnovamento del parco intervistatori. In particolare nel 1° 2003 si è registrato il più alto tasso di assiduità al lavoro tra tutti i periodi considerati. Anche i successivi picchi del 1° e 2° 2004 e del 1° e 3° 2006 hanno corrisposto a momenti di rinnovamento degli intervistatori per l'inserimento di nuovi rilevatori. In realtà anche nel 4° 2004, 2° 2005 e 3° 2005 c'è stata l'introduzione di nuovi rilevatori ma, diversamente dal passato, non si è osservata una ripresa dell'assiduità al lavoro. Questa controtendenza potrebbe essere stata in parte il riflesso di alcuni cambiamenti nel vincolo contrattuale che lega gli intervistatori alle società che operano nelle ricerche di mercato e che gestiscono *call-center*. Infatti, a partire dalla fine del 2004 con l'attuazione della legge Biagi su occupazione e mercato del lavoro, i contratti degli intervistatori, precedentemente di collaborazione coordinata e continuativa, sono stati trasformati in contratti a progetto. Contemporaneamente, vi è stata una sostanziale ristrutturazione della società che conduce le interviste che in qualche modo potrebbe aver influito sulla disponibilità a lavorare degli intervistatori.

5. Analisi temporale della *performance* degli intervistatori

Dalle precedenti analisi descrittive emerge che un elemento di criticità dell'indagine sembrerebbe essere legato alla permanenza prolungata degli intervistatori sulla rilevazione che potrebbe portare a momenti di stanchezza o a maturare comportamenti non idonei. Si è cercato quindi di analizzare tale fattore mediante un'opportuna analisi.

La disponibilità di una serie storica così lunga nel tempo potrebbe indurre a pensare di avere una "banca dati" tale da consentire l'osservazione di lunga durata della *performance* degli intervistatori. Tuttavia, per misurare la *performance* degli intervistatori nel tempo è necessario che siano rispettate alcune condizioni, che costituiscono le ipotesi che hanno portato alla scelta del tipo di analisi adottata.

Innanzitutto, fondamento per qualunque tipo di approccio è che gli intervistatori siano presenti contemporaneamente su più trimestri di osservazione.

Per seguire l'evoluzione temporale della loro *performance* è indispensabile, poi, osservare il gruppo di intervistatori su trimestri contigui poiché l'alternanza di presenza/assenza a più edizioni di indagine potrebbe comportare un approccio all'intervista diverso rispetto a quello di chi vi lavora continuativamente.

Inoltre, la quantità delle informazioni raccolte, essendo variabile di trimestre in trimestre in relazione alla stagionalità del fenomeno e alle diverse esigenze informative che si presentano riflettendosi sulla struttura e sulla lunghezza del questionario, costituisce un elemento di cruciale importanza per valutare la *performance* degli intervistatori. Questi elementi infatti si ripercuotono sia in termini di durata dell'intervista sia in termini di qualità misurata mediante i diversi indicatori. È necessario dunque, che gli intervistatori siano confrontati rispetto alle stesse informazioni. Ciò porta necessariamente a scegliere un sottogruppo di intervistatori presente sugli stessi trimestri di rilevazione - contigui - in un arco temporale che abbracci, complessivamente, più edizioni di indagine possibili. Per superare il problema della mobilità degli intervistatori, che li porta a lavorare con discontinuità da un trimestre all'altro e a restringere ad una durata al più annuale un'analisi su trimestri contigui, si potrebbe essere indotti a pensare di poter osservare eventuali specificità del comportamento degli intervistatori nel tempo considerando coloro che hanno lavorato su gruppi di trimestri di diversa numerosità (4, 8, 12, 16, 20 e 23 trimestri), a prescindere da quali essi siano per ciascun gruppo. Tale tipo di analisi, tecnicamente possibile, non è concettualmente accettabile perché non permetterebbe di osservare il comportamento degli intervistatori rispetto alla somministrazione ai rispondenti di eguali questionari. Eventuali variabili di risposta rappresentate dagli indicatori fin qui considerati, come ad esempio il numero di schede viaggio descritte, potrebbero risentire di una scelta non razionale del tipo di trimestri considerati. Occorrerebbe introdurre un vincolo sul tipo di trimestre incluso tra quelli considerati, ma ciò assottiglierebbe il numero di intervistatori osservabili in ciascun gruppo.

Per testare la *performance* nel tempo degli intervistatori è stata eseguita, dunque, un'analisi fattoriale multipla (AFM) su 33 intervistatori presenti contemporaneamente sui quattro trimestri del 2003.

Occorre ricordare che il 2003 è stato un anno che non ha visto nuovi ingressi di intervistatori e i 33 selezionati avevano già lavorato per almeno una edizione dell'indagine e, complessivamente, in media, per quattro trimestri prima del 2003. Quindi la *performance* è stata testata su un gruppo di rilevatori esperti che ha continuato a lavorare per altri quattro trimestri oltre quelli già effettuati.

L'analisi fattoriale multipla (AFM) consente di individuare le relazioni tra variabili osservate su uno stesso gruppo di individui in più occasioni. Tale tecnica fa parte degli metodi *multivary* su matrici tridimensionali del tipo "unità x variabili x occasioni", dove la terza dimensione può essere, ad esempio, sia di tipo geografico sia di tipo temporale, come in questo caso². L'applicazione dell'AFM al caso degli intervistatori contemporaneamente presenti per tutto il 2003, ha portato alla costruzione di 4 matrici del tipo "intervistatori x indicatori" osservate in quattro diverse "occasioni" di indagine, corrispondenti ai quattro trimestri del 2003.

Gli indicatori calcolati per ciascun intervistatore in ognuno dei trimestri considerati sono il tasso di risposta, il tasso di mancato contatto senza appuntamento, il tasso di mancato contatto con appuntamento, il tasso di appuntamento, il tasso di rifiuto, il tasso di mancata risposta per altri motivi, la durata media delle interviste complete, la durata media dei rifiuti, la durata media delle mancate risposte per altri motivi, il tasso di assiduità al lavoro, il tasso di produttività media giornaliera, il numero medio di componenti per famiglia per intervistatore, il numero medio di schede viaggio compilate per intervistatore.

² Per ulteriori approfondimenti metodologici si veda Bolasco, Sergio. "Analisi multidimensionale dei dati". Roma, Carocci, 1999

Da una prima analisi parziale dei dati, emerge che le correlazioni tra i primi quattro fattori parziali omologhi – espresse nella matrice di correlazione tra fattori parziali – presentano dei valori quasi sempre molto bassi ad indicare che il fenomeno è variabile nel tempo. Infatti per quanto riguarda l'analisi interstrutturale, la matrice dei coefficienti Lg dei legami tra gruppi-trimestre mostra che tra i coefficienti di periodi contigui, quello minore è il coefficiente tra il primo e il secondo periodo (0,564) mentre è stabile nei periodi successivi (0,705 e 0,706), ad indicare un cambiamento abbastanza immediato della struttura, cui segue una staticità nei passaggi tra periodi successivi. Tuttavia il legame maggiore in assoluto si ha tra il primo e terzo periodo (0,918), come ad indicare un ritorno, dopo due trimestri, ad una situazione iniziale.

Tabella 6.1 Coefficienti Lg di legame fra trimestri

Trimestri	Trimestri			
	1	2	3	4
1	1,849			
2	0,564	1,648		
3	0,918	0,705	2,115	
4	0,670	0,511	0,706	1,649

La proiezione su uno stesso piano fattoriale delle variabili osservate nei quattro trimestri porta inevitabilmente ad una perdita di informazione anche se l'inerzia spiegata dai primi 4 fattori risulta essere pari al 50,66%.

Sul primo fattore le variabili che danno il maggior contributo sono il tasso di appuntamento (che contribuisce in modo quasi sempre crescente nel tempo alla spiegazione dell'asse con valori pari a 3,7 nel 1° trimestre; 4,0 nel 2° trimestre; 5,4 nel 3°; 4,7 nel 4°), la durata media delle interviste complete che caratterizza il fattore in modo più o meno decrescente rispetto al tempo (4,0 nel 1° trimestre; 2,9 nel 2°; 3,7 nel 3°, 1,7 nel 4°) e il numero medio di schede viaggio compilate (5,9 nel 1° trimestre; 4,8 nel 3°; 4,6 nel 4°). Il tasso di assiduità dà un contributo importante solo nel 1° trimestre (4,5) mentre è più contenuto nel terzo (2,9). A partire dal 3° trimestre assume importanza anche il tasso di rifiuto (3,5 nel 3° trimestre e 4,4 nel 4°). Ognuno di questi punti è collocato sul semiasse negativo. Questo fattore può essere interpretato come il fattore degli intervistatori che nel corso dei trimestri riescono ad acquisire capacità nel gestire in modo efficiente ma anche 'scaltro' l'intervista, riuscendo ad ottenere le interviste dedicandovi poco tempo ciascuna e riducendo il ricorso agli appuntamenti, a sconto però di un più contenuto numero di viaggi descritti. Tali capacità vengono acquisite nel tempo, con l'abbattimento del tasso di rifiuto. Sono questi gli intervistatori che potrebbero sviluppare nel tempo condotte non adeguate agli standard di qualità in termini di precisione e affidabilità del dato rilevato.

Anche il secondo fattore è fortemente caratterizzato dalla durata media delle interviste complete (che contribuisce in modo sempre crescente alla spiegazione dell'asse con valori pari a 4,6 nel primo trimestre; 5,8 nel 2°; 6,2 nel 3°; 6,4 nel 4°), le cui coordinate si trovano sul semiasse negativo. Invece il tasso di mancato contatto dopo appuntamento presenta contributi –seppur più contenuti- nel primo trimestre (3,3), nel secondo (4,2) e nel quarto (1,5). Entrambi i punti sono situati sul semiasse positivo. Nel primo, terzo e quarto periodo i contributi dei tassi di durata media delle mancate risposte per altri motivi sono pari rispettivamente a 3,9; 2,6 e 2,1 con coordinate negative. Questo fattore è caratterizzato anche dal tasso di risposta (5,5), dal tasso di mancato contatto senza appuntamento (4,5) e dal tasso di mancata risposta per altri motivi (4,5) rilevati nel secondo periodo. Quasi tutti i periodi, dunque, che vedono opporsi le durate medie sia delle interviste complete che delle mancate risposte per altri motivi al tasso di mancato contatto dopo appuntamento, identificano un buon comportamento degli

intervistatori che dedicano tempo alle interviste -a prescindere dall'esito- ottimizzando l'uso dello strumento dell'appuntamento. Questo fattore potrebbe essere definito delle durate medie.

Il terzo asse fattoriale essendo fortemente e quasi esclusivamente caratterizzato dal tasso di produttività media (6,9 nel 1° trimestre; 5,5 nel 2°; 11,7 nel 3°; 5,3 nel 4°) può essere definito come fattore della produttività. La produttività su questo asse è particolarmente esplicativa nel terzo trimestre per poi assumere il valore più basso nel quarto trimestre. Tutte le coordinate sono situate sul semiasse negativo. Il tasso di risposta contribuisce alla spiegazione del fattore nel primo trimestre (4,5) e soprattutto nel terzo (9,8) con coordinate situate sul semiasse negativo. I tassi di mancati contatti -con o senza appuntamento- sono esplicativi più nel terzo periodo (rispettivamente 4,4 e 10,9) che nel quarto nel quale i contributi sono più contenuti (rispettivamente 2,0 e 2,1). Le coordinate per questi due tassi sono positive. Questo terzo asse quindi, discrimina maggiormente gli intervistatori sulla base della produttività, che in certi periodi si accompagna ad un maggior tasso di risposta e ad una riduzione dei mancati contatti.

Il quarto asse fattoriale è caratterizzato prevalentemente dal tasso di rifiuto e dal tasso di appuntamento soprattutto nel 2° trimestre (10,8 e 3,4) e nel 3° (10,1 e 5,9), i cui punti sono situati sul semiasse negativo. Il numero medio di schede viaggio descritte contribuisce in modo crescente alla spiegazione dell'asse (1,1 nel 1° trimestre; 2,9 nel 2°; 5,2 nel 3°; 6,1 nel 4°), con coordinate tutte positive. Tale asse, detto del rendimento, oppone il tasso di rifiuto e dell'appuntamento al numero di schede viaggio descritte, discriminando gli intervistatori più capaci e più ricettivi delle indicazioni impartite circa la gestione dei rifiuti e degli appuntamenti durante la formazione. Tuttavia il buon rendimento degli intervistatori è accentuato soprattutto nei trimestri centrali.

Fig. 6.1 – Anno 2003, i fattori comuni F1(efficienza) e F2 (durate medie)

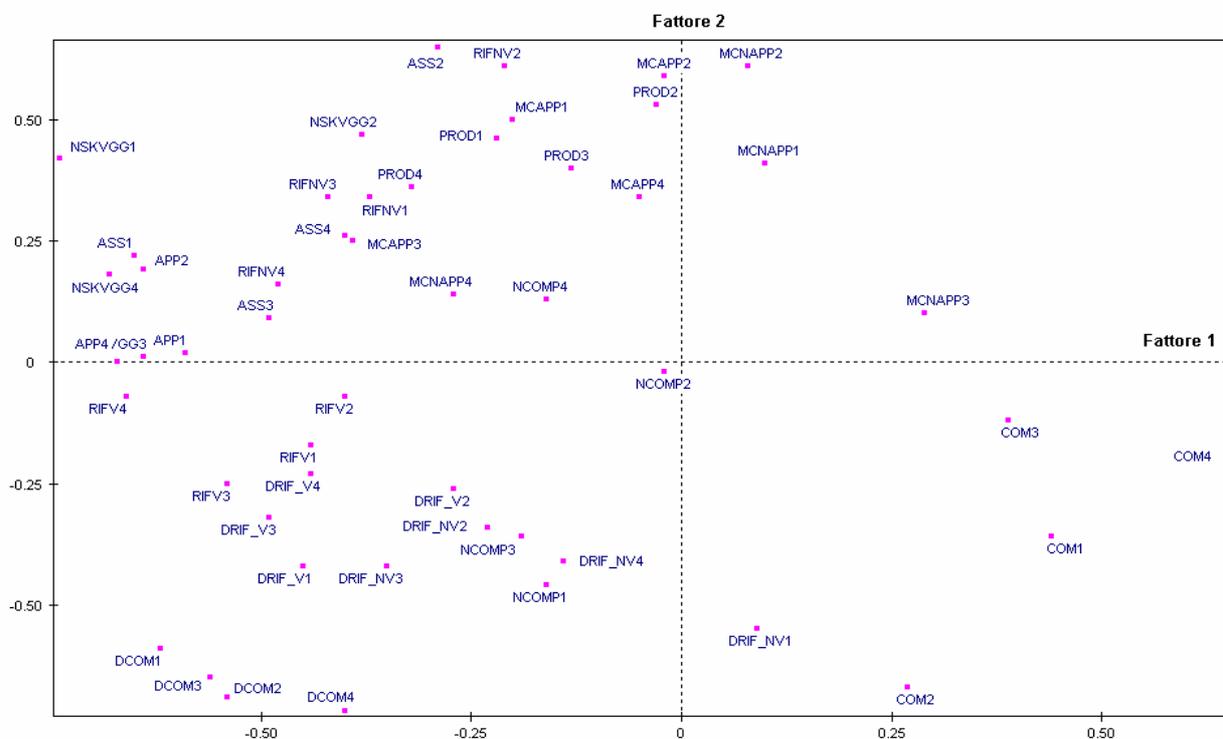
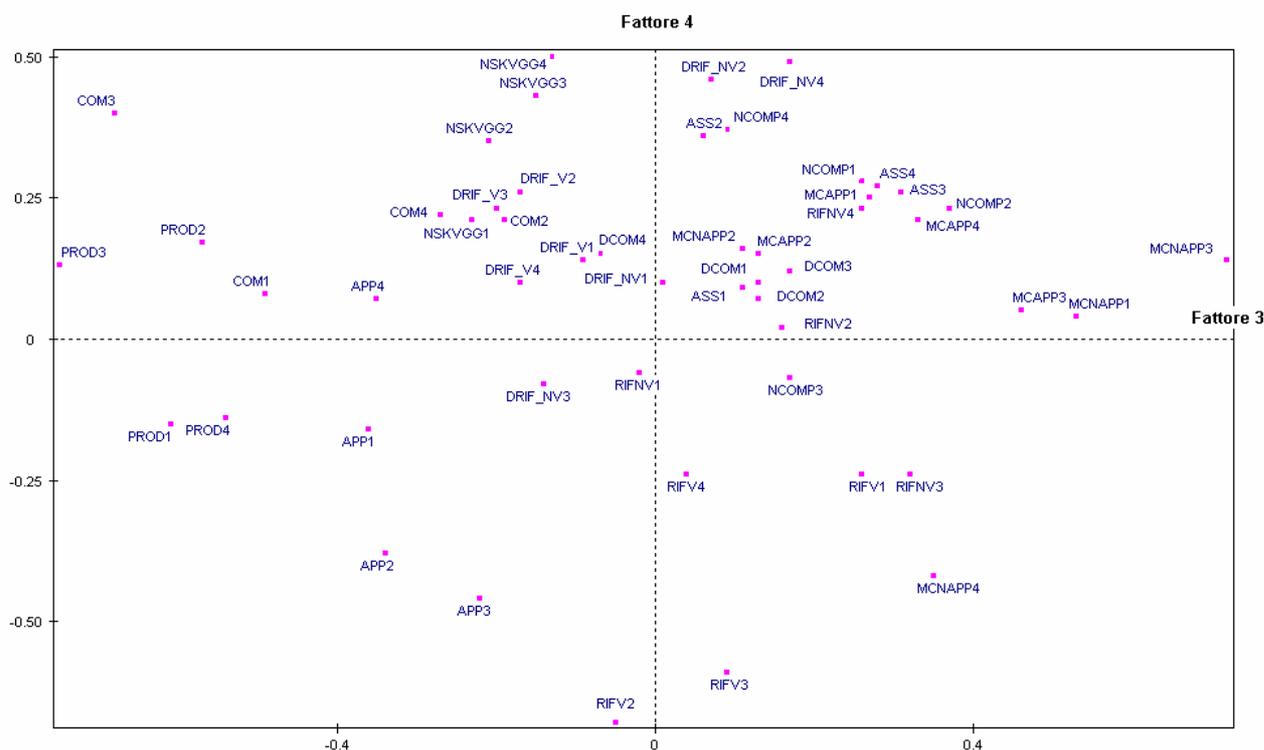


Fig. 6.2– Anno 2003, i fattori comuni F3(produttività) e F4 (viaggi descritti)



6. Conclusioni

Gli intervistatori, senza dubbio, rappresentano uno degli elementi fondamentali per il raggiungimento di un livello di qualità dei dati soddisfacente. La consapevolezza della loro importanza deve indurre tanto l'Istat quanto le società private che lavorano per l'Istat e che congiuntamente si servono del loro lavoro, a renderli coscienti del ruolo che svolgono, il quale rappresenta un anello di strategico rilievo nella catena di produzione del dato.

Dall'analisi dei principali indicatori e delle correlazioni tra gli stessi, complessivamente osservati per i 23 trimestri di rilevazione, emergono comportamenti degli intervistatori coerenti alle indicazioni illustrate durante la formazione, come dimostrano ad esempio il corretto utilizzo dello strumento dell'appuntamento, della gestione dei rifiuti e delle mancate risposte per altri motivi.

Certamente una continuità lavorativa con la stessa società, casualità che si è verificata per l'indagine "Viaggi, vacanze e vita quotidiana", è garanzia di stabilità nello svolgimento del processo produttivo e nella cura del rapporto con i rilevatori, con una ricaduta positiva che inevitabilmente si riflette sull'intero percorso di produzione del dato.

Purtuttavia, nell'attività dei rilevatori costantemente dedicati all'indagine, talvolta sono stati riscontrati sintomi di stanchezza -come un calo nell'assiduità sul lavoro- probabilmente legati ad una certa ripetitività del lavoro che per alcuni di essi si è protratto per molti trimestri di rilevazione. Anche l'analisi fattoriale sul 2003 su un sottogruppo di intervistatori già esperti da più trimestri, ha evidenziato come si delineino nel tempo condotte differenti tra gruppi di intervistatori. Alcuni infatti, acquisiscono capacità nel gestire l'intervista in modo efficiente, in quanto a rapidità del contatto e contenimento dei tassi di appuntamento e rifiuto, sebbene ciò sconti un minor numero di viaggi descritti. Altri, invece, mostrano una maggiore dedizione proprio in termini di tempo dedicato all'intervista riuscendo a concluderla senza ricorrere ad un numero eccessivo di appuntamenti. Altri ancora sono discriminati in base ad un'elevata produttività media in termini di numero di interviste complete effettuate, che in alcuni periodi si accompagna ad elevati tassi di risposta e bassi tassi di mancato contatto. Infine, un ultimo gruppo di intervistatori sembra essere quello che meglio riesce a mettere a frutto le indicazioni

ricevute in fase di formazione circa il contenimento dei rifiuti e sulla cura nel rilevare i viaggi descritti, riuscendo anche a limitare il numero di appuntamenti.

Sulla base di alcuni segnali di *alert* sarebbe auspicabile trovare delle strategie finalizzate a motivare nuovamente gli intervistatori che lavorano da più tempo, mettendo in risalto la professionalità acquisita grazie alla assiduità al lavoro dimostrata di trimestre in trimestre. Ciò potrebbe essere raggiunto sia mediante momenti formativi mirati, sia mediante forme varie di incentivazione al lavoro che dovrebbero essere inevitabilmente concordate con la società che cura la rilevazione.

Inoltre, probabilmente, la disponibilità di una rete interna di rilevazione per le indagini Cati campionarie - sulla scia dell'esperienza maturata con la rete di rilevazione Capi per l'indagine sulle Forze di lavoro - consentirebbe di creare un gruppo di intervistatori più stabile, sebbene il *turnover* sia un elemento non del tutto comprimibile. Ciò consentirebbe di ottimizzare gli interventi sulla rete di rilevazione grazie ad una gestione diretta dei rapporti con la rete stessa, in modo da soddisfare più facilmente adeguati livelli di qualità.

Bibliografia

AA. VV. (2005) "Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche", Istat, Metodi e Norme, no.25, Roma.

AA. VV. (2003), "Metodologia e organizzazione dell'indagine multiscopo sulla domanda turistica 'Viaggi e vacanze' ", Istat, Metodi e Norme, no.17, Roma.

Bolasco, Sergio (1999). Analisi multidimensionale dei dati. Roma, Carocci.

Brancato G, Pellegrini C, Simeoni G. e Signore M. (2004) "Standardising, Evaluating and Documenting quality: the implementation of Istat information system for survey documentation – SIDP". Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004). Mainz, 24-26 May 2004. CD ROM

Escofier, B. e Pagès J. "L'analyse factorielle multiple". Paris, Dunod, 1988.

Analisi dell'effetto intervistatore nella fase di controllo e correzione

Alessandro Martini ISTAT - Servizio SDS

1. Introduzione

Un approccio integrato alla qualità, nel contesto di un'indagine statistica in cui sia presente la figura dell'intervistatore, richiede necessariamente che tutte le strategie mirate al miglioramento continuo vengano definite intorno a questa figura, per le dirette implicazioni che il suo operato ha sui risultati ottenuti.

Le indagini periodiche si prestano perfettamente alla realizzazione di un ciclo continuo, volto a pianificare e mettere in atto interventi migliorativi, verificandone ogni volta l'efficacia ed agendo coerentemente alle indicazioni emerse dalla fase di verifica, affinando ad ogni edizione successiva tutti gli strumenti metodologici per raggiungere sempre nuovi obiettivi in termini di qualità.

In questo contesto il Sistema Informativo di Documentazione delle Indagini dell'Istituto (SIDI) rappresenta, oltre che un importante diario di bordo, un punto di partenza per approfondire aspetti ancora poco esplorati del processo d'indagine.

I dati cui si è fatto riferimento sono i dati individuali dell'indagine Multiscopo Sicurezza dei Cittadini 2002. L'analisi illustrata è stata realizzata a partire dagli indicatori di controllo e correzione gestiti da SIDI, attraverso l'utilizzo di modelli multilivello, tenendo conto della diversa tipologia di interventi in cui si articola la fase di controllo e correzione dei dati dell'indagine e valutando per ciascuna di essa il contributo apportato dal comportamento dell'intervistatore. Verranno inoltre analizzate le caratteristiche socio-demografiche, professionali ed attitudinali dei rilevatori, valutandone ex post l'importanza rispetto alle *performance* ottenute, al fine di ottenere indicazioni utili per le strategie di reclutamento e la loro formazione.

2. L'indagine Sicurezza dei cittadini

A partire dagli anni '90 le tecniche di rilevazione assistite da computer, CAI (Computer assisted Interview) hanno avuto un crescente successo, in particolare nell'ambito delle indagini sociali.

Tali tecniche presentano indubbi vantaggi, basti citare la tempestività della raccolta dei dati, la possibilità di introdurre controlli nella fase di immissione delle informazioni, il contenimento dei costi e la semplificazione della gestione della rete di rilevazione.

Le opportunità che possono essere colte con un approccio del genere possono essere, però, anche più strettamente collegate con alcune specificità dell'indagine, ad esempio la natura particolarmente sensibile degli argomenti trattati e la complessità del questionario, come nel caso dell'indagine ISTAT Sicurezza dei cittadini.

Tale indagine è stata svolta sin dalla prima edizione del 1997 con tecnica CATI ed è la prima indagine su ampia scala svolta dall'ISTAT con tale tecnica.

La seconda edizione dell'indagine è stata condotta da marzo ad ottobre 2002, su un campione costituito da 60.000 famiglie, estratte dall'elenco ufficiale degli abbonati. Tra i componenti della famiglia con almeno 14 anni, viene estratta casualmente la persona da intervistare. Le informazioni raccolte hanno il fine di far emergere aspetti sia oggettivi che soggettivi della domanda di sicurezza dei cittadini.

Il questionario è costituito da 17 sezioni ed è articolato in una struttura gerarchica in modo da non essere somministrato interamente a tutti i rispondenti. Vengono somministrate a tutti i rispondenti la scheda generale, contenente le informazioni strutturali sulla famiglia, le sezioni riguardanti la percezione della sicurezza, il rischio di criminalità in strada, nella zona in cui si vive, e le strategie che le famiglie adottano per proteggersi da tali rischi.

Due sezioni, dette di screening, presentano batterie di domande volte a sondare l'esistenza di vittimizzazione nel periodo di riferimento considerato, rispetto ad alcuni reati contro la persona ed il

patrimonio, sia individuale che familiare. Solo nel caso in cui l'intervistato risponda in modo affermativo a tali sezioni, gli vengono somministrate le sezioni di approfondimento che riguardano analiticamente i reati subiti, dove e come sono avvenuti, le caratteristiche degli autori, il danno subito e l'esito avuto.

Il questionario contiene, inoltre, una sezione, somministrata esclusivamente alle donne di età compresa tra i 14 ed i 59 anni, che tratta il tema delle molestie e violenze sessuali. Vista la natura particolarmente sensibile degli argomenti trattati, sono state selezionate esclusivamente intervistatrici.

La piena coscienza del ruolo cruciale che esse sono chiamate a ricoprire ha suggerito la predisposizione di un sistema integrato di formazione e valutazione del loro operato.

Alla selezione delle intervistatrici ha contribuito il gruppo di lavoro dell'ISTAT, anche avvalendosi di psicologhe esperte di comunicazione. Prima dell'indagine le intervistatrici sono state divise in gruppi per partecipare ai briefing teorici e tecnici della durata di cinque giorni. Durante il periodo di rilevazione sono stati continuamente predisposti sia debriefing di carattere generale che interventi più mirati, apparsi opportuni alla luce delle risultanze del monitoraggio giornaliero del lavoro sul campo, articolato utilizzando sia strumenti tipici di un approccio di tipo "qualitativo" che analisi e tecniche quantitative.

Rispetto al monitoraggio qualitativo sono state predisposte delle schede per il monitoraggio in sala, effettuato da personale ISTAT, per raccogliere in modo omogeneo le impressioni di chi svolge questa attività, ponendo particolare attenzione alle modalità con cui le intervistatrici gestiscono tutte le fasi della comunicazione con le famiglie: dal primo contatto alla gestione delle situazioni più difficili e delicate. Inoltre è stato predisposto un sistema di ricontatto con le famiglie, che ha permesso di controllare il lavoro svolto con i rispondenti stessi, verificando la correttezza delle principali informazioni raccolte e l'aderenza del comportamento dell'intervistatrice alle specifiche metodologiche illustrate durante la formazione.

Riguardo gli strumenti più "quantitativi", l'edizione del 2002 ha costituito l'occasione per integrare maggiormente nel sistema informativo di indagine il monitoraggio giornaliero di indicatori e tassi definiti rispetto agli esiti di contatto con le famiglie, ormai di uso consolidato per le indagini CATI. Sono stati acquisiti gli standard internazionali AAPOR, rivedendo in alcune parti lo schema progettuale del sistema di monitoraggio dell'indagine del 1997, in modo da ottenere un quadro di valutazione armonizzato e coerente con altre fonti. Nel controllo di processo sono state, inoltre, impiegate carte di controllo di tipo SHEWART, che vengono costruite ed analizzate quotidianamente per giorno e per intervistatore, su un set di tassi ed indicatori particolarmente significativi. Tali tecniche di controllo statistico della qualità sono di uso ormai consolidato nei processi di produzione industriale, e consentono di cogliere l'insorgere di anomalie nel processo di rilevazione sotto forma di segnali di fuori controllo.

Appare chiaro come siano state colte pressoché tutte le possibilità offerte dalla tecnica di rilevazione adottata di enfatizzare i controlli in corso d'opera, in linea con la convinzione che il controllo del processo in tutte le sue fasi possa garantire risultati finali con più elevati standard di qualità e che i controlli in corso d'opera siano i più idonei a prevenire effetti distorsivi legati ad errori non campionari.

I controlli ex-post possono, infatti, al massimo limitare i danni di errori non campionari in vari modi insorti. L'utilizzo di una tecnica assistita da computer rende numericamente meno rilevanti alcuni interventi di trattamento dei dati d'indagine, come il piano di check e correzione, ma ciò non ne ridimensiona l'importanza nei confronti di specifiche dimensioni della qualità del dato statistico, in primo luogo sulla tempestività e l'accuratezza.

Nell'indagine Sicurezza dei Cittadini tale fase è risultata comunque molto onerosa, ha richiesto misure e valutazioni indispensabili alla fase di validazione dei dati e per alcuni aspetti si è rivelata un elemento critico nella definizione dei tempi necessari per il rilascio dei dati definitivi e la loro diffusione.

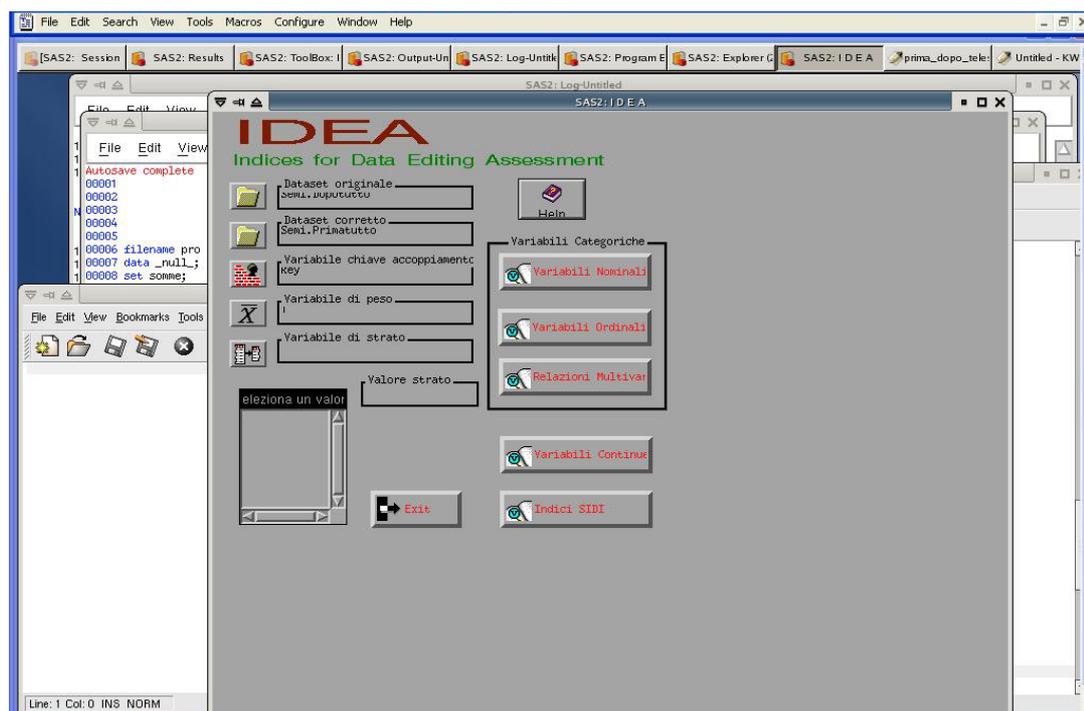
L'utilizzo di un questionario elettronico ha chiaramente limitato molto le possibilità di avere dati affetti da incoerenze, grazie alla gestione delle compatibilità dei dati, dei filtri di sezione e di domanda con i relativi percorsi, come risulta dalla documentazione di questa fase dell'indagine nel sistema informativo SIDI.

Figura 1 - Gli indicatori di controllo e correzione del sistema SIDI

Indicatore	Nord Ovest	Nord Est	Centro	Sud	Isole	Italia
Totale Record						60000
Totale Variabili						597
Totale Variabili Soggette a Imputazione						595
Tasso di Imputazione	7.11	6.97	7.11	6.86	6.82	6.99
Tasso di Modificazione	0.01	0.01	0.01	0.01	0.01	0.01
Tasso di Imputazione Netta	0.01	0.01	0.01	0.01	0.01	0.01
Tasso di Cancellazione	7.09	6.96	7.09	6.84	6.8	6.97
Tasso di Non Imputazione	92.89	93.03	92.89	93.14	93.18	93.01
Tasso di Valori Non Blank Immutati	3.74	3.75	3.8	3.69	3.65	3.73
Tasso di Valori Blank Immutati	89.15	89.27	89.08	89.44	89.54	89.28
Primo Quartile del Tasso di Imputazione per Variabile						0
Terzo Quartile del Tasso di Imputazione per Variabile						0.02
Numero di Variabili con Tasso di Imputazione Superiore al 5%						59
Numero di Variabili con Tasso di Imputazione Superiore al 2%						108
Primo Quartile del Tasso di Imputazione per Record						0.06
Terzo Quartile del Tasso di Imputazione per Record						0.08
Numero di Record con Tasso di Imputazione Superiore al 5%						52185
Numero di Record con Tasso di Imputazione Superiore al 2%						60000
Percentuale di Modificazione	0.14	0.14	0.14	0.14	0.14	0.14
Percentuale di Imputazione Netta	0.12	0.12	0.12	0.14	0.12	0.13
Percentuale di Cancellazione	99.74	99.74	99.73	99.73	99.74	99.74
Percentuale di Valori Non Blank Immutati	4.03	4.03	4.1	3.96	3.91	4.01
Percentuale di Valori Blank Immutati	95.97	95.97	95.9	96.04	96.09	95.99

Gli indicatori sono stati calcolati utilizzando il software IDEA, messo a punto dal servizio studi dell'istituto. Vista la mole di variabili coinvolte nel piano di check e correzione e il numero di osservazioni del campione si è preferito effettuare l'elaborazione su un server Linux, migrando l'applicazione, sviluppata in SAS/AF, con risultati decisamente soddisfacenti in termini di prestazioni.

Figura 2 - Il software IDEA in ambiente Linux



Il piano di controllo e correzione è stato progettato per ristabilire la coerenza logica, statistica e matematica dei dati raccolti, rispetto a specifiche tipologie di errore, determinate per lo più dalla complessa tematica dell'indagine e dalle modalità di comunicazione intervistatrice-rispondente.

Il piano di check e correzione è articolato essenzialmente in due fasi: la prima prevede l'individuazione deterministica delle mancate risposte parziali, dovute ad una interruzione definitiva dell'intervista, avvenuta, in ogni caso, dopo aver completato le sezioni ritenute fondamentali del questionario, ovvero non prima della somministrazione della sezione 17. In questi casi, infatti, le informazioni mancanti possono riguardare solo l'abitazione in cui la famiglia vive, il reddito, i sistemi di sicurezza presenti nell'abitazione e alcune caratteristiche strutturali relative agli altri componenti della famiglia, come il titolo di studio conseguito, la condizione professionale, etc.

La metodologia adottata per il trattamento di tali mancate risposte parziali si basa sul metodo del donatore, che utilizza tutte le informazioni di contesto disponibili per identificare il donatore in maniera corretta. Il software utilizzato è RIDA, che implementa la metodologia basata su una distanza mista minima, adatta anche per variabili qualitative utilizzando una metrica opportuna, con il vantaggio del mantenimento ottimale delle distribuzioni multivariate originali dei fenomeni.

Alcune tipologie di errore tipiche della tematica d'indagine e della tecnica di rilevazione erano già state individuate nella precedente edizione dell'indagine, pertanto, nella fase di progettazione del questionario e nella fase di implementazione informatica, sono stati predisposti tutti gli strumenti necessari a contenerle il più possibile.

Gli interventi più rilevanti sono strettamente correlati a problemi di comunicazione che possono insorgere tra rispondente e intervistatrice, e ciò accade tanto più facilmente quanto più i temi toccati scatenano reazioni emotive particolarmente intense in entrambi gli attori in campo.

Difetti di comprensione delle domande o difficoltà a collocare le proprie esperienze possono essere comprensibili, se si considera che in un'indagine di vittimizzazione viene richiesto al rispondente di evocare episodi vissuti particolarmente spiacevoli, a volte i più drammatici della propria vita.

Aspetti relativi alla definizione dei reati, la distinzione di questi tra tentati e consumati, possono trarre in inganno il rispondente ma devono essere assolutamente chiari all'intervistatrice.

Il piano di controllo e correzione ha previsto, quindi, una attenta revisione delle classificazioni dei reati. Ad esempio la sezione in cui si rilevano gli scippi, la prima che viene somministrata al rispondente, ha un effetto di attrazione che porta molte persone a rispondere di avere subito tale reato, tuttavia le domande di approfondimento sulle modalità di accadimento del reato fanno spesso emergere incoerenze che portano a ricodificare quanto descritto tra i borseggi o i furti di oggetti personali. Analogamente un furto in casa, in cui è stato fatto anche uso della forza, viene ricollocato nelle rapine, mentre nei casi di furto in cui nell'approfondimento si verifica che non è stato sottratto nulla, diventano dei tentati furti.

Un'altra problematica strettamente connessa al forte impatto emotivo dei contenuti dell'indagine riguarda l'effetto *telescoping forward*, un errore sistematico del ricordo che porta a collocare alcuni eventi più vicini nel tempo, determinando una sovrastima del fenomeno rispetto al periodo di riferimento. Per la rilevazione dei reati si è adottata quindi una tecnica ad imbuto, con lo scopo di sollecitare il rispondente ad un approfondimento che consenta di contenere questa distorsione.

Nella progettazione del questionario si è tenuto conto di queste problematiche, sono state inseriti controlli on-line che possono correggere alcune incompatibilità nel corso stesso dell'intervista o, nel caso in cui incoerenze più complesse coinvolgano sezioni diverse del questionario, sono state previste nuove modalità di risposta che segnalano l'errore, per potere poi effettuare le correzioni del caso.

Il ricorso a tali strumenti denota, da parte dell'intervistatrice, la giusta attenzione agli aspetti qualitativi dei dati raccolti, una corretta gestione delle modalità di comunicazione con il rispondente, nonché la piena padronanza tecnica del questionario elettronico.

Il controllo a posteriori delle informazioni raccolte dalle intervistatrici nelle note all'intervista e nei campi aperti degli items "Altro specificare" costituisce anch'esso un patrimonio prezioso per il controllo e la correzione dei dati.

L'analisi di tali informazioni spesso contribuisce in modo determinante al ripristino della coerenza e correttezza dell'informazione rilevata, al contrario a volte può accadere che il ricorso a tale strumento non sia completamente giustificato: può capitare che quanto riportato dall'intervistato sia riconducibile

direttamente a modalità di risposta già previste. Per quanto numericamente poco rilevante, questo fenomeno denota un atteggiamento delle intervistatrici superficiale e sbrigativo nella conduzione dell'intervista, unito ad una scarsa conoscenza del questionario. Molto spesso le domande contenenti campi aperti fanno riferimento ad approfondimenti riguardanti le modalità di accadimento di un reato subito, quindi può accadere che siano percepite come meno importanti da parte delle intervistatrici e pertanto somministrate con meno rigore e puntualità.

La fase di controllo e correzione è tipicamente un momento importante della valutazione della performance dello strumento di rilevazione predisposto, ma per quanto esposto finora può esserlo anche per la performance dell'intervistatrice.

3. Il modello gerarchico a due livelli

Obiettivo del presente contributo è pervenire all'individuazione dei fattori esplicativi che maggiormente concorrono a determinare la necessità di applicare metodologie di controllo e correzione ai dati provenienti dalle interviste effettuate nel corso dell'indagine.

Per procedere ad una soddisfacente individuazione delle possibili determinanti si è fatto ricorso a modelli di regressione multilivello, in quanto le analisi descrittive svolte non hanno consentito di raggiungere tale risultato.

Il ricorso a tali modelli è consigliabile nel caso in cui le unità su cui si rileva il fenomeno oggetto di studio (unità di primo livello) risultino naturalmente aggregate in gruppi o classi differenti (unità di secondo livello), eventualmente aggregabili a loro volta (unità di terzo livello) e così via.

In tali casi, infatti, è ragionevole ritenere che la variabilità del fenomeno dipenda, oltre che da variabili esplicative individuali (di primo livello), anche dal fatto che un individuo appartenga ad un determinato gruppo avente caratteristiche peculiari, che lo distinguono dagli altri gruppi. Nel caso in esame, i rispondenti (unità di primo livello) risultano naturalmente associati all'intervistatrice che ha condotto le interviste.

L'applicazione di tali modelli è possibile in quanto la tecnica CATI garantisce due ipotesi di base essenziali:

- l'assegnazione casuale delle famiglie da intervistare alle intervistatrici;
- poter considerare le intervistatrici come un campione casuale di una popolazione non osservabile di potenziali intervistatrici telefoniche.

Le variabili risposta considerate nell'analisi sono variabili dicotomiche del tipo:

$$y_{ij} = \begin{cases} 1 & \text{se i dati individuali hanno richiesto un determinato intervento di check e correzione} \\ 0 & \text{altrimenti} \end{cases}$$

Trattandosi di una variabile binaria il modello prescelto è stato il modello logistico a due livelli, preferito ad un modello di tipo probit per la maggiore facilità di interpretare i risultati attraverso il calcolo degli *odds ratio*.

P_j indica invece la probabilità che i dati di un'intervista effettuata dalla j-esima intervistatrice abbiano richiesto interventi di correzione; di conseguenza il modello nullo è il seguente:

$$y_{ij} = P_j + e_{ij}$$

Il valore assunto dalla variabile risposta è dato quindi dalla probabilità media della j-esima intervistatrice più una componente casuale avente media 0 e varianza strettamente dipendente da P_j .

Dovendo utilizzare una trasformazione che renda normalmente distribuite le probabilità P_j il ricorso ad un modello di tipo logistico fornisce la seguente specificazione:

$$\log it(P_j) = \beta_j$$

$$\beta_j = \gamma + U_j$$

Di conseguenza si ottiene il modello nullo:

$$\log it(P_j) = \gamma + U_j$$

Dove γ indica il valore medio della probabilità per l'intera popolazione, mentre U_j costituisce lo scarto da tale valore per la j-esima intervistatrice.

Gli U_j costituiscono l'elemento distintivo del modello gerarchico, rappresentano gli effetti casuali relativi alle unità di secondo livello. Esprimono quindi l'effetto residuo esercitato da ciascuna intervistatrice nei riguardi della variabile risposta. L'ipotesi di base prevede per tali componenti casuali una distribuzione normale, con media nulla e varianza τ^2 .

Le unità di primo livello sono costituite dalle 60.000 interviste complete, effettuate da 220 diverse intervistatrici. Nell'analisi sono state considerate solo le intervistatrici che hanno effettuato non meno di 30 interviste, risultando così infine 59.822 interviste complete, effettuate da 207 intervistatrici, le unità di secondo livello.

Coerente con quanto suggerito dalla teoria, i passi della procedura di stima sono stati:

- stima del modello nullo, al fine di verificare, attraverso la stima della varianza di secondo livello, la ragionevolezza del ricorso ad un'analisi multilivello;
- stima del modello a intercetta casuale comprendente tutte le variabili di primo livello risultate significative;
- stima del modello a intercetta casuale (due livelli) completo, costituito dalle variabili di primo e secondo livello.

Di seguito verranno riportati e commentati in forma estesa i risultati relativi alla stima del modello a intercetta casuale finale, cioè del modello di regressione logistica a due livelli ad intercetta casuale:

$$\log it(P_{ij}) = \gamma + \sum_{h=1}^r \gamma_h X_{hij} + U_j \quad \text{con}$$

$$U_j \sim N(0, \tau^2)$$

Come detto, la variabile risposta utilizzata è misurata al primo livello, ed è dicotomica, mentre le variabili esplicative sono sia di primo che di secondo livello, ovvero relative sia ai rispondenti che alle intervistatrici, e sono sia categoriche che continue.

P_{ij} indica la probabilità di dover intervenire sui dati della i-esima intervista effettuata dalla j-esima intervistatrice, e quindi si ha:

$$y_{ij} = \begin{cases} 1 & \text{se i dati individuali hanno richiesto un determinato intervento di check e correzione} \\ 0 & \text{altrimenti} \end{cases}$$

Gli effetti casuali di secondo livello U_j , già presenti nel modello nullo, rappresentano ora l'effetto residuo esercitato da ciascuna intervistatrice nei riguardi della variabile risposta, una volta controllato

l'effetto delle variabili esplicative X_h . L'ipotesi di base prevede per tali componenti casuali una distribuzione normale, con media nulla e varianza τ^2 .

Il coefficiente di correlazione intra-classe, per le ipotesi fatte è dato da $\rho = \frac{\tau^2}{\pi^2/3 + \tau^2}$ ed indica quanta parte della variabilità totale è attribuibile a livello delle intervistatrici.

Le elaborazioni sono state condotte utilizzando la procedura NLMIXED del SAS. Tale procedura prevede la specificazione del predittore lineare, come funzione delle variabili esplicative e della funzione link utilizzata.

Quando la convergenza viene raggiunta, l'output di questa procedura fornisce la stima dei parametri, del loro errore standard ed include relativamente a ciascun parametro la verifica della sua significatività attraverso il test t di Wald. Per quanto riguarda la procedura di stima, la proc NLMIXED massimizza un'approssimazione numerica dell'esatta verosimiglianza marginale del modello non lineare, attraverso il metodo di quadratura adattivo di Gauss-Hermite.

4. I modelli stimati ed i risultati ottenuti

Il primo passo della procedura di stima dei modelli è consistito nella stima del modello a sola intercetta casuale, ossia senza introdurre nessuna variabile esplicativa. L'importanza di questa prima analisi risiede nella possibilità di calcolare il coefficiente di correlazione intra-classe, ossia stimare la quota di variabilità attribuibile alle intervistatrici.

Le variabili risposta considerate sono relative agli interventi di controllo e correzione ritenuti maggiormente significativi dell'intero piano di check dell'indagine così per ciascuno di essi è stata definita una variabile dicotomica codificata con 1 nel caso sia stato necessario effettuarlo sui dati individuali, 0 altrimenti.

Gli interventi considerati fanno riferimento alle tipologie di errori di cui si è detto, la mancata risposta parziale alla sezione 17, il ricorso errato ai campi aperti, le correzioni riguardanti le classificazioni dei reati e quelle relative al periodo di riferimento, riconducibili all'effetto telescoping.

Riguardo l'utilizzo dei campi aperti la variabile risposta è stata definita 1 nel caso siano stati commessi almeno due errori nel corso dell'intervista, l'averne commesso uno solo non è stato ritenuto particolarmente grave ed associabile ad un comportamento scorretto ma lascia supporre una occasionale svista. Le variabili coinvolte per ciascuna operazione e lo stadio di avanzamento della fase di controllo e revisione hanno consentito di identificare con buona precisione le diverse tipologie oggetto di analisi.

Nella tabella 1 sono riportati i risultati della stima dei modelli nulli relativi alle variabili risposta considerate.

Si è proceduto inoltre alla stima di un modello utilizzando come variabile di risposta la probabilità di intervistare un rispondente di 65 anni e più, variabile che non dovrebbe essere influenzata da un effetto dovuto all'intervistatore, per verificare l'esistenza di un valore di fondo del coefficiente correlazione intra-classe, risultato pari a 0,002, in linea con quanto noto in letteratura.

Tabella 1 - Effetto intervistatore sul piano di check e correzione

Modello	Standard error	Coefficiente correlazione intra-classe
Mancate risposte parziali	0,006	0,038
Campi aperti	0,016	0,137
Codifica reati screening	0,005	0,039
Periodo riferimento	0,009	0,068

I risultati ottenuti indicano in generale una chiara presenza dell'effetto intervistatore rispetto alle variabili risposta considerate, essendo in ogni caso considerato significativamente superiori ai valori di fondo, riscontrati sia nella letteratura che nelle analisi di controllo.

I valori più bassi si sono ottenuti in corrispondenza degli interventi effettuati per trattare le mancate risposte parziali della sezione 17 e quelli relativi alle classificazioni dei dati.

Nel primo caso va considerato che la discrezionalità da parte delle intervistatrici è limitata dal fatto che non erano a conoscenza che le intervista interrotte in quella fase fossero comunque considerate valide. Le cause della eterogeneità in questo ambito sono perciò da attribuire alla loro diversa capacità di continuare a gestire la comunicazione nell'ultima fase dell'intervista, stimolando il rispondente ad affrontare tematiche sensibili, come il reddito disponibile e le strategie adottate dalla famiglia per difendersi dalla criminalità.

Il coefficiente di correlazione intraclassa si attesta su valori modesti (0,039) anche nella segnalazione di incoerenze nella classificazione dei reati, comportamento che viene valutato positivamente in quanto denota la partecipazione attiva dell'intervistatrice al ristabilimento della coerenza delle informazioni rilevate, grazie alla capacità di sfruttare efficacemente tutti gli strumenti predisposti all'interno del questionario per rilevare correttamente il fenomeno.

Il fatto che non si registri una elevata eterogeneità nel comportamento delle intervistatrici rispetto questo aspetto può essere interpretato come un segnale di buon funzionamento del piano di formazione predisposto: le intervistatrici hanno raggiunto in maniera sufficientemente uniforme uno standard qualitativo accettabile nella gestione del questionario e degli strumenti predisposti per una corretta gestione dell'intervista, comprendendo in pieno l'importanza di una corretta rilevazione dei fenomeni considerati. Superiore il valore registrato per gli interventi effettuati sul periodo di riferimento, sintomo di un atteggiamento più discrezionale delle intervistatrici che fa riflettere sull'opportunità di enfatizzare questo aspetto nella fase di formazione.

Il valore più alto è stato ottenuto nel caso del reiterato errore nel ricorso ai campi aperti, pari a 0,137, ed evidenzia come in questo caso l'effetto rilevatore sia abbastanza elevato, chiaro segnale di come venga commesso con un grado minore di uniformità tra le intervistatrici.

Tali comportamenti errati possono essere percepiti dalle intervistatrici come poco gravi ma denotano, al contrario, una scarsa conoscenza del questionario o un atteggiamento nell'intervista superficiale.

Il fatto che spesso tali campi siano relativi a domande contenute nelle sezioni di approfondimento, relative al dove, come, quando un evento sia successo, e che quindi vengano somministrate ad intervista ormai "conquistata" può far supporre che alcune intervistatrici perseguano una strategia opportunistica, mirata ad una "accelerazione" dell'intervista al fine di ridurre il tempo totale ed effettuare il numero maggiore possibile.

I risultati di questa prima analisi confermano la presenza di un effetto della figura del rilevatore rispetto a questi aspetti, e suggerisce l'opportunità di approfondire l'analisi con lo scopo di ricercarne alcune determinanti, anche se la dimensione modesta dei valori riscontrati rende difficile evidenziare legami che possano spiegare quote consistenti della variabilità riscontrata.

Una volta verificata l'effettiva esistenza di un effetto dovuto alle intervistatrici, lo studio è proseguito introducendo nei modelli delle variabili esplicative a livello del rispondente e dell'intervistatrice.

Le variabili esplicative prese in considerazione per la stima del modello completo sono state scelte per alcuni versi sulla base delle analisi descrittive svolte. Per quanto riguarda le variabili a livello di intervistatrice, inizialmente si è cercato di verificare l'influenza di alcune variabili "strutturali" rilevate presso le intervistatrici durante la loro selezione. Le informazioni socio-demografiche disponibili riguardano il titolo di studio, l'età, gli anni di esperienza lavorativa come intervistatrice telefonica. Tali informazioni nell'edizione del 1997 non erano disponibili pertanto non era stato possibile analizzarne il contributo nella valutazione dei risultati ottenuti.

Tabella 2 - Caratteristiche delle intervistatrici

CARATTERISTICHE	%
Classe di età	
Fino a 30 anni	59,1
Da 31 a 40 anni	23,1
Più di 41 anni	17,8
Titolo di studio	
Dottorato di ricerca	1,4
Laurea	8,6
Diploma	90,0
Esperienza lavorativa	
Un anno	65,0
Due anni	28,6
Tre o più anni	6,4

Relativamente alle variabili categoriche introdotte nel modello, si è reso necessario individuare per ciascuna di esse una modalità base o di riferimento nei confronti della quale valutare l'effetto di tutti gli altri livelli esistenti; si è scelto di considerare come modalità di base la caratteristica più diffusa nel campione in analisi, l'individuo che possiede tutte queste caratteristiche verrà denominato individuo base.

In particolare le caratteristiche individuali introdotte nel modello sono state:

- **il sesso** codificato come 0 se il rispondente è femmina, 1 se maschio
- **classi di età** tale variabile è stata introdotta nel modello attraverso tre variabili di tipo dummy, avendo scelto come base la classe 14-39 le variabili presenti nel modello risultano essere *etaris1*, relativa agli individui 40-64 e *etaris2* per gli individui di 65 anni e più.

Le variabili riferite alle intervistatrici inserite nel modello sono state:

- **classe di età:** la variabile dicotomica *ril_claeta* assume valore 1 se l'intervistatrice ha più di di 30 anni, 0 altrimenti;
- **titolo di studio:** la variabile dicotomica *ril_istr* assume valore 1 se l'intervistatrice ha conseguito la laurea, o un titolo superiore, 0 altrimenti;
- **anni di esperienza lavorativa:** la variabile dicotomica *ril_esp* assume valore 1 se l'intervistatrice ha dichiarato una esperienza lavorativa come intervistatrice di almeno due anni, 0 altrimenti.

Una volta introdotte nel modello le variabile esplicative, sia di primo che di secondo livello, la

formula $\frac{\tau^2}{\pi^2 / 3 + \tau^2}$ non stima più la quota di variabilità attribuibile alle unità di secondo livello ma conserva una certa utilità in quanto comunque ha un valore indicativo di come quella determinata specificazione del modello spieghi la variabilità intra-classe.

Nella tabella seguente sono riportati i risultati ottenuti dalla stima dei modelli considerati.

Tabella 3 - Parametri stimati con il modello ad intercetta casuale completo - variabili esplicative II livello “strutturali”

Variabile esplicativa	Variabile risposta				
	Mancate risposte parziali	Campi aperti	Codifica reati screening	Periodo riferimento	
Rispondente					
Sesso	0.10	-0.07		0.11	0.10
Età 41-64	-0.08	-0.09		-0.31	-0.08
Età 65 +	-0.21	-0.47		-0.59	-0.21
Intervistatrice					
Età 31+	ns	ns		ns	ns
Laurea	ns	0.18		ns	ns
Anni di esperienza lavorativa >2	ns	ns		ns	ns

ns = non significativa a 0,05

L’analisi effettuata introducendo le variabili “strutturali” delle intervistatrici nel modello non ha evidenziato un loro significativo contributo a spiegare la varianza di secondo livello, ossia le differenze riscontrate tra le intervistatrici.

Questo risultato da un lato attesta che la formazione continua delle intervistatrici ha funzionato, è stata cioè in grado di fornire gli strumenti necessari a condurre le interviste e gestire la comunicazione con i rispondenti in modo da assorbire possibili effetti legati alle precedenti esperienze lavorative, alla loro età ed alla loro formazione scolastica o universitaria.

Solo nel caso degli errori nel ricorso ai campi aperti il contributo della variabile associata al titolo di studio appare significativo ($\beta=0,18$ $p=0,02$). Sfruttando la funzione logistica per valutare questo risultato in termini di variazione percentuale della probabilità di dover intervenire sui dati con procedure di controllo e correzione si ha che nel caso di interviste condotte da rilevatrici laureate tale probabilità aumenta del 16,3% rispetto all’individuo base, intervistato da una non laureata.

Per quanto tale errore sia tra i meno gravi, non inficiando la struttura gerarchica del questionario, essendo per lo più domande di approfondimento delle modalità di accadimento dei reati, che non modificano il percorso dell’intervista, denota come già detto un atteggiamento superficiale nella condotta dell’intervistatrice, ed una conoscenza lacunosa del questionario.

Ritenere il ruolo ricoperto come non adatto alle proprie capacità e poco rispondente alle proprie aspirazioni professionali da parte delle intervistatrici costituisce un fattore che incide negativamente sulla qualità dei dati rilevati.

Per altri versi la non significatività delle variabili socio-demografiche disponibili fa emergere il convincimento che variabili tipicamente “di processo” possano contribuire a spiegare con maggiore efficacia la variabilità esistente a livello delle intervistatrici piuttosto delle caratteristiche socio-demografiche osservate in questo caso.

L’analisi è quindi proseguita introducendo, nei modelli relativi alle variabili per quali non si sono evidenziate relazioni significative per le variabili di secondo livello, alcuni indicatori ritenuti tra i più significativi del processo di rilevazione, calcolati per intervistatrice ed introdotti nel modello completo.

Tali indicatori sono:

- la durata media delle interviste complete;
- il tasso di mancate risposte alle domande della sezione 16, relativa a molestie e violenze sessuali;
- il tasso di interruzione definitivo calcolato come rapporto tra il numero di interviste interrotte definitivamente e il totale degli esiti definitivi;
- il tasso di rifiuto calcolato come rapporto tra il numero di rifiuti e il totale degli esiti definitivi.

Gli ultimi due tassi sono stati calcolati sul totale degli esiti definitivi, ossia su tutti i 106.989 nominativi contattati nel corso della rilevazione.

In due casi, gli interventi relativi alla riclassificazione dei reati e alle modifiche ai periodi di riferimento, una variabile che ha contribuito in maniera significativa a spiegare parte della variabilità

dovuta alle intervistatrici è stata la durata delle interviste ($\beta=0,193$ $p=0,001$ per il tentato-consumato), ($\beta=0,191$ $p=0,03$ per il telescoping).

Rispetto alle mancate risposte parziali alla sezione 17, dovute ad una interruzione definitiva dell'intervista non sono emerse relazioni significative con le variabili di secondo livello considerate.

Il valore ottenuto del coefficiente di correlazione intra-classe attesta in ogni caso una certa eterogeneità tra le intervistatrici, che le variabili considerate non riescono a spiegare, un ulteriore sviluppo dell'analisi potrebbe essere condotto utilizzando modelli gerarchici a coefficienti casuali, in modo da poter analizzare più a fondo le possibili interazioni tra caratteristiche del rispondente e dell'intervistatrice.

Tabella 4 - Parametri stimati con il modello ad intercetta casuale completo - variabili esplicative II livello "di processo"

Variabile esplicativa	Variabile risposta			
	Mancate risposte parziali	Campi aperti	Codifica reati screening	Periodo riferimento
Rispondente				
Sesso	0.10	-0.06	0.11	0.10
Età 41-64	-0.08	-0.07	-0.31	-0.08
Età 65 +	-0.21	-0.40	-0.59	-0.21
Intervistatrice				
Durata media interviste	ns	ns	0,193	0,191
Tasso di MPR sezione 16	ns	ns	ns	ns
Tasso di interruzione definitivo	ns	ns	ns	ns
Tasso di rifiuto	ns	ns	ns	ns

ns = non significativa a 0,05

Trasformando le stime ottenute in probabilità attraverso la funzione logistica si ottiene che un aumento pari alla deviazione standard della durata media delle interviste, pari a poco più di 9 minuti, determina un aumento percentuale della probabilità di utilizzo corretto degli strumenti di segnalazione delle incoerenze nei dati pari al 17,4% nel caso delle riclassificazioni dei reati e del 19,3% per i periodi di riferimento dei reati.

Questo risultato mette in luce come un approccio dell'intervistatrice orientato alla consapevolezza del proprio ruolo, svolga un ruolo di primaria importanza anche rispetto alla qualità finale dei dati rilevati.

5. Conclusioni

Il ruolo dell'intervistatore, cruciale in tutte le indagini statistiche in cui sia impiegato, assume ancora maggiori elementi di specificità nelle indagini su vasta scala che trattano tematiche sensibili, come nel caso dell'Indagine Multiscopo ISTAT Sicurezza dei Cittadini 2002.

Le difficoltà nella selezione di un numero cospicuo di operatrici richiede un quadro informativo più esauriente possibile, le analisi fatte hanno cercato di fornire un contributo in tal senso evidenziando tutte le difficoltà a cogliere gli aspetti fondanti della professionalità delle rilevatrici nelle caratteristiche socio-demografiche. Un risultato interessante fa emergere alcuni aspetti problematici nell'impiego di intervistatrici laureate: il senso di frustrazione nel ricoprire un ruolo non ritenuto adatto al loro grado di preparazione ha ricadute negative su alcuni aspetti della qualità dei dati rilevati con possibili rischi di burn-out.

Al contrario ci sono comportamenti delle intervistatrici che hanno un chiaro effetto positivo sulla qualità dei dati, le analisi condotte hanno evidenziato la loro stretta relazione con aspetti peculiari del processo, la durata delle interviste in primo luogo, di cui era stata già colta per altri versi l'importanza e per i quali erano già stati predisposti tutti gli strumenti disponibili per il monitoraggio in corso d'opera.

Il diverso ricorso delle intervistatrici a tali comportamenti appare sicuramente influenzato da strategie utilitaristiche evidentemente in contrasto con una politica orientata alla qualità: il fatto che siano retribuite in base alle interviste effettuate può determinare una scarsa propensione a porre una particolare

attenzione ad alcuni aspetti qualitativi dei dati raccolti o ad introdurre informazioni di contesto, che rappresentano invece un importante valore aggiunto.

I risultati ottenuti da un lato forniscono indicazioni positive riguardo il piano di formazione continua predisposto per supportare il lavoro sul campo delle intervistatrici, per altri versi attestano la validità metodologica di specifiche richieste inserite dal gruppo di lavoro nella definizione dei capitolati di gara delle indagini CATI di propria competenza.

La valorizzazione del ruolo dell'intervistatore è quindi un obiettivo primario per affrontare indagini con tematiche particolarmente sensibili, da perseguire curando in modo particolare la loro selezione e formazione, ma altresì attraverso la definizione di quegli aspetti, prevalentemente correlati alle forme contrattuali e alle modalità della loro retribuzione, che hanno ricadute fondamentali sulla qualità del processo e quindi, inevitabilmente, sui risultati dell'indagine.

Bibliografia

- Giusti C., Vignoli D. "Ricorso alla contraccizione in Egitto: tra scelte individuali ed eterogeneità regionali" Working Paper 2005/14 Università degli studi di Firenze – Dipartimento di Statistica "G. Parenti".
- Hox, J.J., E.D. De Leeuw e I.G.G. Kreft. "The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model". In *Measurement Errors in Surveys*, a cura di Beær, Paul P. et al. 339-461. New York: J. Wiley & Sons, 1991.
- Luzi O et al., "IDEA Indicatori per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI" Manuale per L'utente del software IDEA.
- SAS Institute Inc. (1999). SAS/STAT® User's Guide, Version 8, SAS Institute Inc., Cary NC.
- J. D. Singer, (1998) "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models" in *Journal of Educational and Behavioral Statistics*, Vol.24, n.4, pp.323-355.
- Istat. Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche. A cura di Maria Giuseppina Muratore, Marina Signore. Roma, Istat, 2005 (Metodi e Norme n.25)

Integrazione tra sistemi di gestione di indagine e il Sistema di Documentazione delle Indagini SIDI: l'esperienza del servizio delle statistiche strutturali sulle imprese

G. Brancato^{*}, C. Casciano[§], C. Pellegrini^{*}, C. Schiattone[§]

Istat

**Struttura di progetto Audit per la qualità*

§Servizio Statistiche strutturali sulle imprese dell'industria e dei servizi

1. Introduzione

Il presente lavoro illustra l'esperienza relativa all'integrazione tra il sistema di "monitoraggio del pervenuto" in uso presso il servizio delle Statistiche strutturali sulle imprese dell'industria e dei servizi (SSI) dell'Istat e il sistema di documentazione delle indagini SIDI diffuso presso l'Istituto, avente come obiettivo quello di automatizzare il calcolo degli indicatori di qualità di Copertura e Mancata Risposta Totale.

Molte indagini del servizio SSI utilizzano un software che permette di registrare le risposte derivanti dalla fase di raccolta dei dati (questionario utilizzabile, impresa non eleggibile con il motivo della non eleggibilità, impresa non rispondente con il motivo della non risposta). Dall'altra parte, gli indicatori standard di SIDI su Copertura e Mancata Risposta Totale si basano proprio sul risultato della fase di contatto delle unità e sono calcolati trasponendo gli eventi derivanti dalla fase di contatto dell'unità (codificati in modo specifico dalle singole indagini), nelle categorie utili per i numeratori e denominatori coinvolti nel calcolo degli indicatori (classi standard comuni a tutte le indagini).

Entrambi i sistemi sono stati sviluppati nello stesso ambiente, Oracle, e pertanto il processo di integrazione è risultato facilitato.

Per realizzare il progetto, è stato creato un sistema di transcodifica tra le modalità utilizzate nel sistema di monitoraggio del pervenuto e quelle utilizzate in SIDI. Quindi all'interno del sistema di monitoraggio del pervenuto è stata aggiunta una funzione, disponibile per una tipologia di profilo di utente, che consente la creazione di una tabella con i dati per gli indicatori da importare poi nel sistema SIDI.

Il progetto non rappresenta solo una attività di integrazione tra sistemi informativi diversi disponibili all'Istituto ma ha anche numerose ricadute. In primo luogo aumenta la validità del sistema SIDI in quanto consente di calcolare gli indicatori in modo coerente e standard per un elevato numero di rilevazioni. Infatti, la confrontabilità di uno stesso indicatore tra indagini diverse potrebbe essere compromessa o falsata dalla eterogeneità nell'utilizzo della classificazione alla base degli indicatori. In secondo luogo, ma non meno rilevante, consente di ridurre il carico sui referenti del sistema SIDI - ossia quelle figure dell'Istituto incaricate dell'aggiornamento dei metadati e degli indicatori per tutti i processi documentati nel sistema - derivante dal calcolo di questi due gruppi di indicatori di qualità: non è più necessario sviluppare programmi ad hoc con le necessarie ricodifiche per calcolare questi indicatori, potendosi avvalere di una funzione predisposta. Infine, offre ai responsabili di indagine un ulteriore strumento di monitoraggio della fase di rilevazione. Tutto ciò contribuisce a migliorare la qualità e l'efficienza all'interno dell'Istituto.

Questa esperienza quindi rappresenta un importante punto di riferimento da generalizzare ed estendere, dove possibile, ad altri contesti dell'Istituto.

2. Il sistema di monitoraggio del pervenuto per le indagini strutturali sulle imprese

Il Sistema Informativo per il monitoraggio del pervenuto viene utilizzato dalla gran parte delle indagini del Servizio delle Statistiche Strutturali sulle Imprese (SSI) per gestire le liste di spedizione dei modelli di rilevazione, per registrare le informazioni relative alla modalità di risposta delle imprese alle varie indagini, per gestire le liste di sollecito e le variazioni anagrafiche delle imprese osservate.

Di seguito l'elenco delle indagini SSI che utilizzano correntemente tale applicazione:

INDAGINE	SIGLA	PERIODICITA'
Rilevazione sul Sistema dei Conti di impresa	SCI	Annuale
Stima Provvisoria Valore Aggiunto delle imprese	SPVA	Annuale
Rilevazione sulle tecnologie dell'informazione, della comunicazione e competitività delle imprese (con supplemento di osservazione per le imprese di intermediazione finanziaria)	TEC	Annuale
Rilevazione della Produzione industriale	PRODCOM	Annuale
Rilevazione sulle Piccole e medie imprese	PMI	Annuale
Rilevazione sulla formazione del personale nelle imprese	CVTS3	Quinquennale
Rilevazione sul costo del lavoro	RCL	Quadriennale
Rilevazione sulle imprese italiane a controllo estero	FATS	Biennale
Rilevazione sull'innovazione nelle imprese	CIS4	Quadriennale
Ricerca scientifica e sviluppo sperimentale nelle imprese	RS1	Annuale
Rilevazione della struttura delle retribuzioni	ESES	Quadriennale

I profili di utenza definiti per questa applicazione sono **UTENTE**, **COORDINATORE** e **AMMINISTRATORE**. L'utenza dell'amministratore del sistema è dedicata alle attività di gestione e manutenzione del sistema, mentre le prime due sono relative agli utilizzatori del sistema, all'interno dei singoli processi di indagine. Di seguito si descrivono le principali funzionalità previste per ogni profilo di utenza.

Il profilo **UTENTE** è utilizzato dai revisori sostanzialmente per:

- la **registrazione dei codici di risposta** che può essere effettuata tramite tastiera, penna ottica o file, in quest'ultimo caso è possibile assegnare il codice risposta a un insieme di imprese;
- la **creazione dei file** delle imprese rispondenti e non rispondenti;
- la realizzazione di **statistiche** immediate: totale delle imprese rispondenti e non; totale dei diversi codici di risposta delle imprese rispondenti e non rispondenti raggruppate per attività economica, addetti, regione, provincia o comune;
- la **visualizzazione delle informazioni anagrafiche** e quelle relative ai solleciti degli ultimi tre anni di una determinata impresa.

Il profilo **COORDINATORE** consente, oltre all'utilizzo delle medesime maschere della versione **UTENTE**, altre funzioni quali:

- la possibilità di effettuare **variazioni anagrafiche** che vengono registrate su una tabella per poi essere eventualmente utilizzate al momento dell'invio successivo del file di spedizione;
- l'**inserimento di nuove ditte** che non erano presenti nel file di spedizione;

- l'esecuzione di diversi tipi di **interrogazioni avanzate** che riguardano **tutte le indagini** e non solo quella su cui si sta lavorando. Ad esempio è possibile visualizzare informazioni riguardanti i contatti (referenti, mail, telefono ecc.) non solo dell'indagine che si sta lavorando ma anche delle altre indagini;
- la visualizzazione, per una singola impresa, del **quadro storico** relativo alle diverse indagini: se ha risposto e a quale indagine, come e in quali anni;
- per alcune indagini del Servizio (SPVA, PRODCOM, RS1), l'aggiornamento automatico delle informazioni anagrafiche registrate nei sistemi informativi di produzione delle indagini stesse.

Tra le principali funzionalità del profilo AMMINISTRATORE, si evidenziano la predisposizione del sistema per ogni singola edizione di indagine, il caricamento del file di spedizione alle imprese e la realizzazione del file definitivo di sollecito.

Il sistema è stato sviluppato in ambiente Oracle. Il database risiede sul server linux PAKISTAN, mentre l'applicazione client è stata sviluppata in Oracle Forms 6.i e risiede sulle postazioni Windows degli utenti.

L'informazione di maggior rilievo trattata in questo sistema informativo è la modalità di risposta delle imprese, che viene associata a un CODICE RISPOSTA. Nel corso degli anni, si è cercato il più possibile di rendere omogenei i codici di risposta tra le diverse indagini. Tuttavia, essendo il sistema di monitoraggio del pervenuto funzionale alla gestione specifica delle informazioni che ritornano dalla fase di contatto, per alcune indagini è stato necessario mantenere dei codici "non condivisi". Ad esempio, per l'Indagine statistica sulla ricerca scientifica (RS1) è necessario mantenere un codice corrispondente a 'Rispondente in cui i dati RS sono nel consolidato di altra impresa'; una modalità di risposta che non ha alcun interesse per le altre rilevazioni che utilizzano il sistema di monitoraggio del pervenuto, ma che per l'indagine RS1 è utile per la definizione delle liste di osservazione degli anni successivi.

Nella tabella seguente vengono riportati, a livello di singola indagine, i codici in comune e quelli specifici necessari all'integrazione con il sistema di transcodifica di SIDI.

DESCRIZIONE	CODICI	RILEVAZIONI										
L'impresa fornisce le sole previsioni di spesa	02	-	-	RS1	-	-	-	-	-	-	-	-
Pervenuta e utilizzabile per il solo anno in corso	03	-	-	RS1	-	-	-	-	-	-	-	-
Rispondente in cui i dati RS sono nel consolidato di altra impresa	04	CVTS3	-	RS1	-	-	-	-	-	-	-	-
Questionario rispedito	10	CVTS3	-	-	-	-	-	-	-	-	SPVA	
Pervenuta utilizzabile	11	CVTS3	FATS	-	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Esigua (fino a 94 addetti)	12	-	-	-	-	SCI	TEC	-	RCL	-	SPVA	
Esigua (da 95 a 99 addetti)	13	-	-	-	-	-	-	-	RCL	-	SPVA	-
Pervenuto e utilizzabile solo dati definitivi (anno precedente)	14	-	-	-	-	-	-	-	-	-	SPVA	-
Pervenuto e utilizzabile solo dati provvisori (anno in corso)	15	-	-	-	-	-	-	-	-	-	SPVA	-
Esigua (fino a 9 addetti) TELECOM	16	-	-	-	-	-	TEC	-	RCL	-	-	
Questionario utilizzabile: ditta modificata successivamente	18	-	-	-	-	-	-	-	-	-	-	PRODCOM
PMI con 100 addetti e oltre	19	-	-	-	PMI	-	-	-	-	-	-	-
Rifiuto esplicito	20	CVTS3	-	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Attività RS extramuros oppure extra e previsioni	21	CVTS3	-	RS1	-	-	-	-	-	-	-	-

Respinta, trasferita, sconosciuta	31	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Cessata, fallita (anno in corso)	40	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Cessata, fallita (anno precedente)	41	-	-	-	-	-	TEC	-	-	-	SPVA	-
Inattiva (anno in corso)	50	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Inattiva (anno precedente)	51	-	-	-	-	-	TEC	-	-	ESES	SPVA	-
In liquidazione (anno in corso)	70	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
In liquidazione (anno precedente)	71	-	-	-	-	-	TEC	-	-	-	SPVA	-
Amministrazione controllata, concordato preventivo, concordato fallimentare (anno in corso)	80	CVTS3	-	RS1	PMI	SCI	TEC	CIS4	RCL	-	SPVA	PRODCOM
Amministrazione controllata, concordato preventivo, concordato fallimentare (anno precedente)	81	-	-	-	-	-	TEC	-	-	-	SPVA	-
Fusa o scorporata (anno in corso)	90	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Fusa o scorporata (anno precedente)	91	-	-	-	-	-	TEC	-	-	ESES	SPVA	-
Pervenuto questionario elettronico	95	-	-	RS1	-	SCI	TEC	-	RCL	-	SPVA	PRODCOM
Pervenuta non utilizzabile (dati incompleti, in accertamento, richiesta chiarimenti, ecc.)	AA	-	FATS	-	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Fuori campo di osservazione (anno in corso)	BB	CVTS3	-	RS1	PMI	-	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Ha risposto in modo negativo ma farà ricerca successivamente	B1	-	-	RS1	-	-	-	-	-	-	-	-
Fuori campo di osservazione (permanente)	CC	CVTS3	-	RS1	-	SCI	TEC	-	RCL	-	-	-
Concessa proroga per 1 mese	DD	-	-	-	-	-	-	CIS4	RCL	-	-	-
Concessa proroga per 2 mesi	EE	-	-	-	-	-	-	-	-	-	SPVA	-
Concessa proroga per 3 mesi	FF	-	-	-	-	-	-	-	-	-	SPVA	-
Modelli in sospenso >249 addetti	LL	-	-	-	-	-	-	-	RCL	ESES	-	-
Modelli in sospenso <250 addetti	MM	-	-	-	-	-	-	-	-	ESES	-	-
Questionario restituito in bianco dietro accordi con l'impresa	PP	-	-	-	-	-	-	-	-	-	-	PRODCOM
Sede legale non produttiva (impresa pluri-localizzata)	QQ	-	-	-	-	-	-	-	-	-	-	PRODCOM
Questionario e Teleform	TT	-	-	-	-	-	-	-	-	-	SPVA	-
Sollecito affidato a ditta esterna	ZZ	-	-	-	-	-	TEC	CIS4	-	-	SPVA	-
Risposta non pervenuta		CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM

Grazie alla collaborazione tra rappresentanti dell'SSI e del gruppo SIDI, è stato possibile associare ogni singolo codice di risposta del sistema di monitoraggio del pervenuto, con i codici SIDI relativi alla classificazione delle mancate risposte e degli errori di copertura.

Questa attività è stata anche sfruttata per il popolamento del data base del sistema informativo AsiaIndagini, che ha come scopo il controllo sistematico e generalizzato del carico statistico sulle imprese soggette a rilevazioni censuarie e campionarie e per il quale non era necessario il dettaglio della codifica del sistema di monitoraggio ma era più appropriato quello sintetico di SIDI.

In particolare, come si vedrà nei successivi paragrafi, sono state individuate aggregazioni della classificazione SIDI, realizzate tabelle di transcodifica tra i codici dei due sistemi e sviluppate procedure di caricamento delle informazioni sui contatti e sulle modalità di risposta delle imprese rilevate.

3. Gli indicatori standard di qualità di SIDI sulla fase di rilevazione

Gli indicatori SIDI di copertura e mancata risposta totale sono stati pensati come sottoprodotti del processo produttivo (Brancato *et al.*, 2004). Essi misurano alcune componenti della copertura e cioè della qualità dell'archivio di estrazione delle unità di rilevazione attraverso le informazioni ottenute dalla fase di contatto di una indagine che utilizza l'archivio. Ovviamente si tratta di indicatori che riflettono la sottocopertura e gli errori nelle informazioni contenute nell'archivio (tasso di errore di lista, con le sue componenti: unità non più esistenti, unità con variazioni di stato, unità fuori target), mentre nessun indicatore sulla sovracopertura è ovviamente desumibile dalla fase di rilevazione. Relativamente agli indicatori di mancata risposta totale, sono presenti in SIDI diversi indicatori: un tasso di mancata risposta totale (tutte le unità per le quali non è pervenuta alcuna risposta rispetto a tutte le unità), un tasso di mancata risposta al netto degli errori di lista che pregiudicano il contatto, e infine un tasso di mancata risposta riferito alle sole unità accertate come eleggibili. Oltre a questi tassi, in SIDI viene calcolato un tasso, denominato tasso di unità risolte, ossia un indicatore della quota di unità per le quali è stato possibile accertare lo stato di eleggibilità. Per le indagini prevalentemente postali, come le indagini strutturali sulle imprese, questo indicatore risulta molto basso a causa dei numerosi questionari inviati e per i quali non torna indietro all'Istat alcuna informazione.

Il calcolo di questi indicatori, da effettuarsi alla fine della fase di raccolta dei dati, richiede la sola disponibilità degli esiti del contatto (unità rispondente; unità non rispondente con il motivo; unità non contattata con il motivo, come per esempio trasferimento o indirizzo errato) opportunamente ricodificati nelle classi omogenee coinvolte nei numeratori e denominatori dei tassi. Uno dei principali problemi esistenti per il calcolo di questi indicatori è rappresentato dalla non disponibilità o dalla non registrazione degli esiti della fase di contatto.

Già nel 1996, in occasione della ristrutturazione della Rilevazione sulle Piccole e Medie Imprese e sull'esercizio di arti e professioni (PMI), era stata condotta una esperienza di definizione di una classificazione standard degli esiti della fase di contatto (Hidioglou, 1993) e sua acquisizione insieme al codice dell'impresa mediante penna ottica (Brancato G. *et al.*, 1997; Brancato G. e Signore M, 1998). Questo aveva consentito non solo di calcolare indicatori a fine periodo di rilevazione, ma anche di effettuare delle analisi durante la fase di acquisizione dei dati per valutare possibili effetti selettivi nelle risposte. A partire da quella esperienza pilota, il sistema inizialmente sviluppato per la sola indagine PMI è stato esteso a numerose altre indagini strutturali sulle imprese, come illustrato nel paragrafo precedente.

Contemporaneamente, il sistema dei codici definito per PMI è stato ampliato per rispondere alle esigenze di monitoraggio proprie del sistema e alcuni nuovi codici specifici per singole indagini sono stati aggiunti. Tuttavia, come già accennato, più recentemente si è proceduto ad una armonizzazione di questi codici e alla creazione di una tabella di transcodifica che riportava i codici propri di ciascuna indagine in quelli standard e più aggregati di SIDI. Ciò anche in considerazione del fatto che, come precedentemente accennato, è stato avviato un flusso di informazioni tra il sistema di monitoraggio del pervenuto e l'archivio ASIA, per il quale è stata adottata la classificazione di SIDI.

In dettaglio, la classificazione di SIDI si basa sui seguenti aggregati:

CLASSIFICAZIONE DELLE UNITA' IN SIDI	CODICE SIDI
Totale unità	01
Unità risolte	02
Unità non risolte	03
Unità eleggibili	04
Unità non eleggibili	05
Unità non più esistenti	06
Unità non eleggibili a causa di variazioni di stato	07
Unità fuori target	61
Unità rispondenti	08
Unità rispondenti al primo contatto	09
Unità rispondenti a contatti successivi	10
Unità non rispondenti	11
Unità non rispondenti per rifiuto	12
Unità non rispondenti per mancato contatto	13
Unità non rispondenti per altri motivi (esclusi rifiuti e mancati contatti)	14
Unità non contattate per errori di lista	15
Unità non contattate per altri motivi	16

Il gruppo SIDI si è impegnato per definire e sviluppare progetti in collaborazione con le direzioni produttive per diminuire il carico sui responsabili di indagine derivante dall'attività di calcolo e archiviazione degli indicatori in SIDI. Il paragrafo che segue riporta un esempio di tali progetti.

4. L'integrazione tra il sistema di monitoraggio del pervenuto e SIDI

Dopo un'accurata analisi, che ha coinvolto i responsabili statistici e tecnici dei due sistemi, è stata verificata la fattibilità di integrazione tra i sistemi, soprattutto perché entrambi adottano un DB relazionale Oracle e sono realizzati su piattaforma Oracle Forms.

L'analisi svolta a livello dei dati ha portato alla realizzazione, nel sistema del monitoraggio del pervenuto, di una *subject area* che permettesse di ricostruire la tabella adottata dal sistema SIDI dei valori assoluti utilizzati nelle formule di calcolo degli indicatori. A tal fine sono state create:

- Tabelle di transcodifica relativamente alla codifica delle indagini e ai codici degli esiti con i codici adottati in SIDI
- Una tabella di aggregazione territoriale a livello di ripartizione geografica a 5 modalità e totale Italia, livello di aggregazione territoriale standard adottato in SIDI per la rappresentazione degli indicatori di qualità
- Una tabella relativa ai valori assoluti, per replicazione e indagine, che rispecchiasse la struttura richiesta in SIDI.

Tra i due database è stato creato un DB link per permettere un accesso bidirezionale tra i due sistemi alle tabelle d'interesse coinvolte in questa attività.

Il dettaglio della tabella di transcodifica tra i codici dei valori assoluti SIDI e quelli degli esiti del monitoraggio del pervenuto per singola indagine è riportato nella tabella che segue. Solo alcuni codici SIDI sono riportati, in quanto tutti gli altri sono derivabili come somma di questi.

Descrizione pervenuto	Codici	Rilevazioni										
Unità non risolte (03)												
Questionario Teleform	TT	-	-	-	-	-	-	-	-	-	SPVA	-
Sollecito affidato a ditta esterna	ZZ	-	-	-	-	-	TEC	CIS4	-	-	SPVA	-
Risposta non pervenuta		CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM

Unità non più esistenti (06)												
Cessata,fallita (anno in corso)	40	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Cessata,fallita (anno precedente)	41	-	-	-	-	-	TEC	-	-	-	SPVA	-
Unità con variazione di stato (07)												
Esigua (fino a 94 addetti)	12	-	-	-	-	SCI	TEC	-	RCL	-	SPVA	-
Esigua (da 95 a 99 addetti)	13	-	-	-	-	-	-	-	RCL	-	SPVA	-
Esigua (fino a 9 addetti) TELEC	16	-	-	-	-	-	TEC	-	RCL	-	-	-
Questionario utilizzabile: ditta modificata successivamente (PRODCOM)	18	-	-	-	-	-	-	-	-	-	-	PRODCOM
PMI con 100 addetti e oltre	19	-	-	-	PMI	-	-	-	-	-	-	-
Inattiva (anno in corso)	50	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Inattiva (anno precedente)	51	-	-	-	-	-	TEC	-	-	ESES	SPVA	-
In liquidazione (anno in corso)	70	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
In liquidazione (anno precedente)	71	-	-	-	-	-	TEC	-	-	-	SPVA	-
Amministrazione controllata, concordato preventivo, concordato fallimentare (anno in corso)	80	CVTS3	-	RS1	PMI	SCI	TEC	CIS4	RCL	-	SPVA	PRODCOM
Amministrazione controllata, concordato preventivo, concordato fallimentare (anno precedente)	81	-	-	-	-	-	TEC	-	-	-	SPVA	-
Fusa o scorporata (anno in corso)	90	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Fusa o scorporata (anno precedente)	91	-	-	-	-	-	TEC	-	-	ESES	SPVA	-
Unità fuori target (61)												
Fuori campo di osservazione(anno in corso)	BB	CVTS3	-	RS1	PMI	-	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Ha risposto in modo negativo ma farà ricerca successivamente	B1	-	-	RS1	-	-	-	-	-	-	-	-
Fuori campo di osservazione(permanente)	CC	CVTS3	-	RS1	-	SCI	TEC	-	RCL	-	-	-
Rispondenti (08)												
L'impresa fornisce le sole previsioni di spesa	02	-	-	RS1	-	-	-	-	-	-	-	-
Pervenuta e utilizzabile per il solo anno in corso	03	-	-	RS1	-	-	-	-	-	-	-	-
Rispondente in cui i dati RS sono nel consolidato di altra impresa	04	CVTS3	-	RS1	-	-	-	-	-	-	-	-
Pervenuta utilizzabile	11	CVTS3	FATS	-	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Pervenuto e utilizzabile solo dati definitivi (anno precedente) SPVA	14	-	-	-	-	-	-	-	-	-	SPVA	-
Pervenuto e utilizzabile solo dati provvisori (anno in corso) SPVA	15	-	-	-	-	-	-	-	-	-	SPVA	-
Attività RS extramuros oppure extra e previsioni	21	CVTS3	-	RS1	-	-	-	-	-	-	-	-
Pervenuto questionario elettronico	95	-	-	RS1	-	SCI	TEC	-	RCL	-	SPVA	PRODCOM

Rifiuti (12)												
Rifiuto esplicito	20	CVTS3	-	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Non rispondenti per altri motivi (14)												
Pervenuta non utilizzabile(dati incompleti,in accertamento,richiesta chiarimenti ecc.	AA	-	FATS	-	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM
Questionario rispedito	10	CVTS3	-	-	-	-	-	-	-	-	SPVA	-
Concessa proroga per 1 mese	DD	-	-	-	-	-	-	CIS4	RCL	-	-	-
Concessa proroga per 2 mesi	EE	-	-	-	-	-	-	-	-	-	SPVA	-
Concessa proroga per 3 mesi	FF	-	-	-	-	-	-	-	-	-	SPVA	-
Modelli in sospenso>249(RCL ESES)	LL	-	-	-	-	-	-	-	RCL	ESES	-	-
Modelli in sospenso< 250(RCL ESES)	MM	-	-	-	-	-	-	-	-	ESES	-	-
Questionario restituito in bianco dietro accordi con l'impresa (PRODCOM)	PP	-	-	-	-	-	-	-	-	-	-	PRODCOM
Sede legale non produttiva (impresa plurilocalizzata) (PRODCOM)	QQ	-	-	-	-	-	-	-	-	-	-	PRODCOM
Mancati contatti per errori di lista (15)												
Respinta,trasferita,sconosciuta	31	CVTS3	FATS	RS1	PMI	SCI	TEC	CIS4	RCL	ESES	SPVA	PRODCOM

Come si può notare dalla tabella precedente nel sistema del monitoraggio del pervenuto vengono registrati tutti gli esiti a livello elementare, quindi è stato necessario attraverso la procedura realizzata ricostruire la gerarchia della classificazione adottata da SIDI, visualizzata nella figura sottostante.

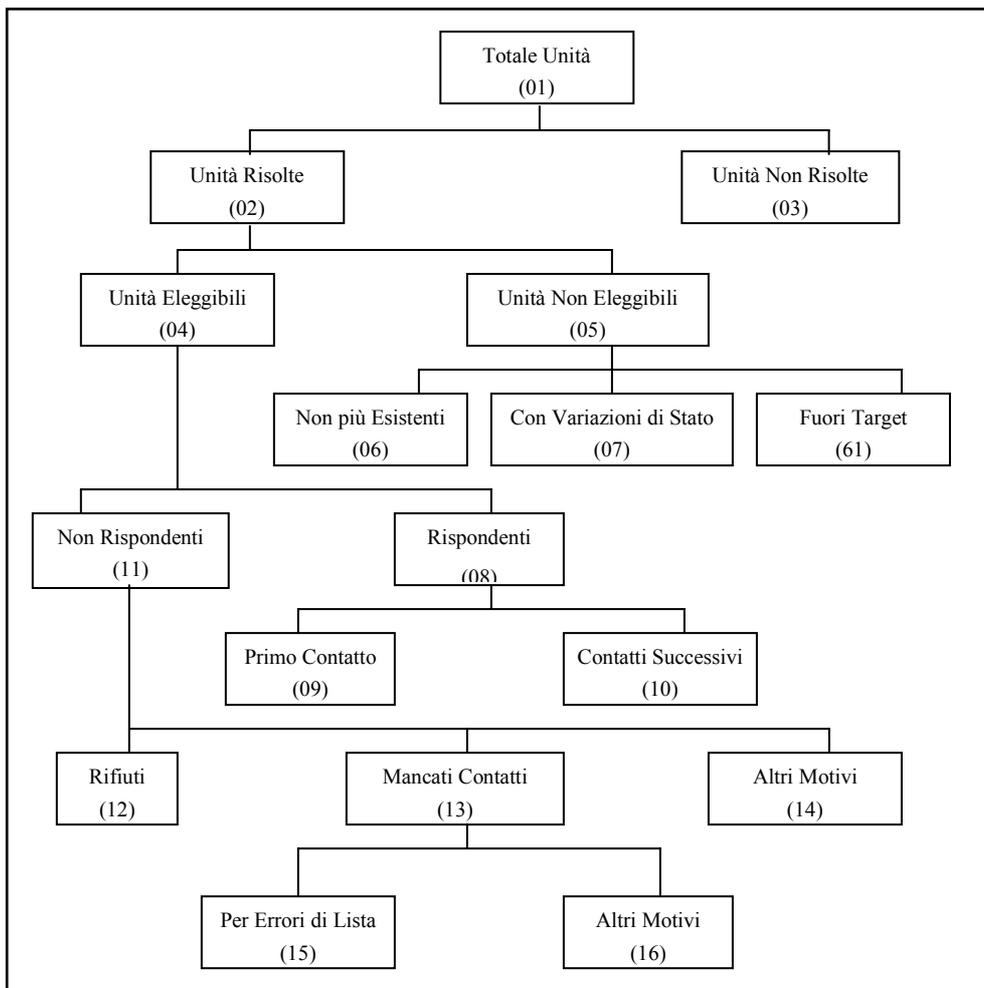


Figura 1 – Grafo della classificazione delle unità per il calcolo degli indicatori SIDI

L'analisi svolta a livello di funzioni ha portato alla realizzazione di una *stored procedure pl/sql* che permettesse di ricostruire la struttura della tabella dei valori assoluti necessari a SIDI per il calcolo degli indicatori: questa procedura riceve in input il codice indagine e l'anno di riferimento dei dati e ricostruisce, a livello di ripartizione geografica e di totale Italia, tutta la gerarchia della classificazione SIDI dei valori assoluti necessari al calcolo degli indicatori di copertura e mancata risposta in un'apposita tabella.

All'interno della procedura è stata realizzata una funzione che permette di raccordare la rappresentazione temporale dell'occasione di indagine del sistema di monitoraggio del pervenuto con la data di riferimento dei dati adottata in SIDI: infatti nel sistema del pervenuto viene riportato come riferimento temporale soltanto l'anno mentre in SIDI, per tutti i campi di tipo data, è stato adottato il formato DD-MM-YYYY.

Questa procedura, per la generazione dei valori assoluti, non comporta nessuna attività aggiuntiva per il responsabile di indagine, che anzi non si trova costretto a realizzare programmi ad-hoc per generare il file text o per immettere manualmente i dati nel sistema SIDI.

Di seguito vengono riportati i passi che il responsabile del monitoraggio del pervenuto e il referente SIDI per la data indagine, debbono svolgere per il calcolo degli indicatori.

All'ingresso del sistema (Figura 2) del monitoraggio del pervenuto al momento dell'autenticazione dell'utente, viene memorizzato il codice indagine (1) necessario per ricostruire, attraverso una tabella di transcodifica, il codice indagine del sistema SIDI.

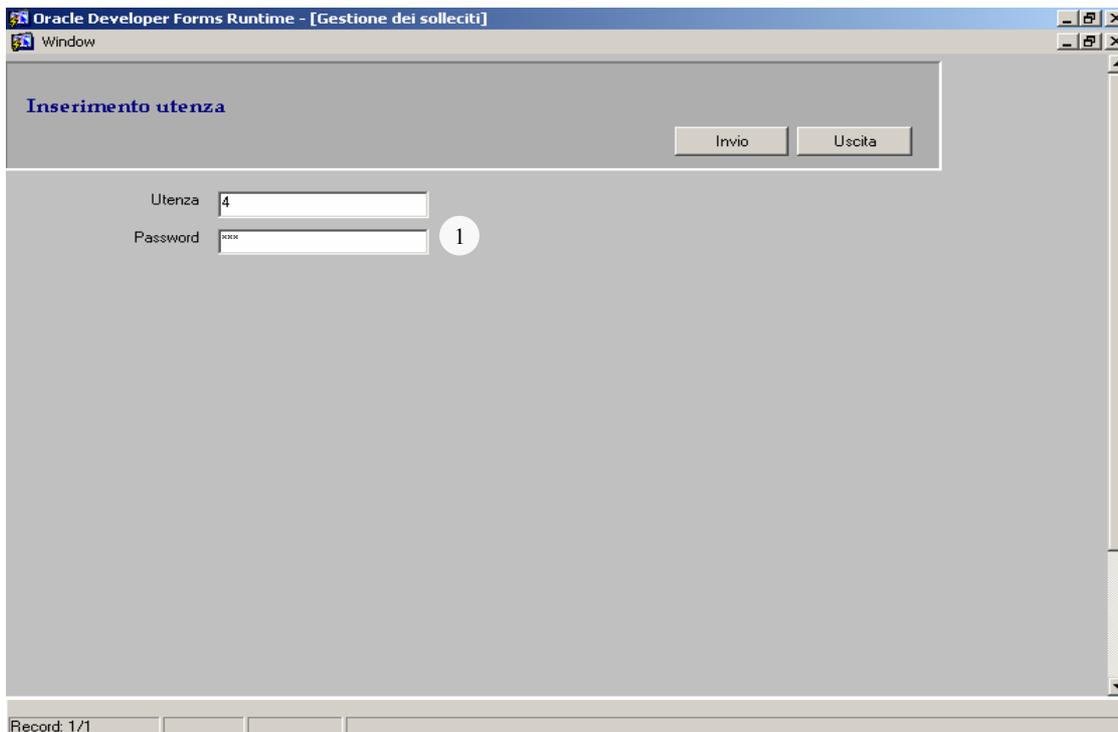


Figura 2 – Autenticazione sistema del monitoraggio

Dopo essersi autenticati nel sistema con utente e password viene visualizzata una schermata (Figura 3) dove si inserisce l'anno dell'occasione di indagine da elaborare (1), questo servirà a costruire la data di riferimento per SIDI (per le indagini annuali viene assegnato di default il 31/12 dell'anno selezionato). Quando l'utente "coordinatore" è certo che tutti gli esiti sono stati registrati nel sistema, potrà attraverso il bottone "SIDI" (2) mandare in esecuzione la procedura pl/sql per calcolare i valori assoluti. (Figura 3).

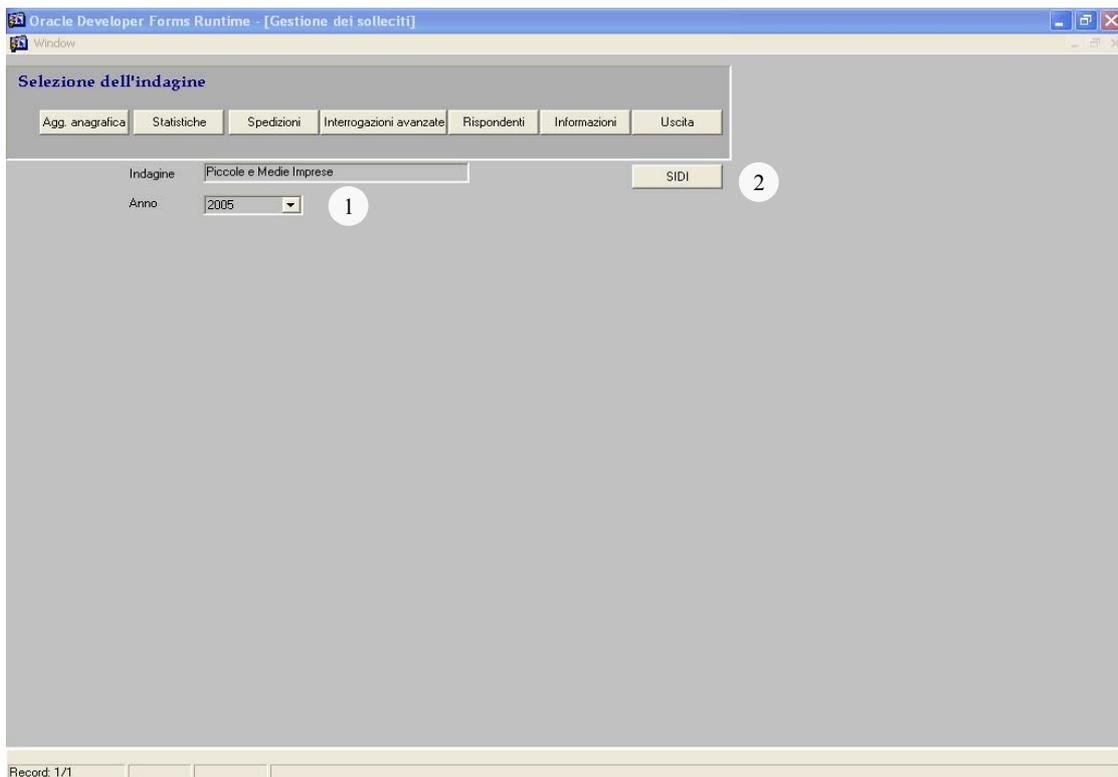


Figura 3 – Schermata per elaborazione SIDI

Nonostante nel tempo il sistema di gestione degli indicatori SIDI sia stato migrato su piattaforma Java, questo non ha comportato nessuna modifica alle procedure realizzate all'interno dei due sistemi per il reperimento dei dati.

Una volta preparati i valori assoluti, il referente di indagine per SIDI dovrà entrare nel sistema per fornire le informazioni mancanti e calcolare gli indicatori di qualità, importando i dati direttamente da DB.

Infatti, nel sistema SIDI i valori assoluti possono essere immessi in tre modi: manualmente digitando direttamente nella griglia impostata i valori dei numeratori e denominatori, attraverso il caricamento da un file testuale avente una struttura compatibile con il sistema o, quando abilitati, attraverso l'import da DB. SIDI attraverso la tabella dei profili di utenti-indagini riconosce quando l'utente, per l'indagine in linea, è abilitato ad importare direttamente i dati da un altro DB.

Sarebbe stato possibile importare direttamente i dati nella tabella SIDI dei valori assoluti, ma si è preferito adottare questa modalità per garantire la congruenza dei metadati prevalenti associati agli indicatori e per effettuare un ulteriore controllo dei dati attraverso la procedura di convalida.

Di seguito viene mostrata la schermata che appare al referente di indagine dopo l'immissione dei metadati prevalenti e l'import da DB (Figura 4).

Figura 4 - Caricamento dei valori assoluti in SIDI

5. Un esempio di applicazione

5.1. La rilevazione sulle piccole e medie imprese e sull'esercizio di arti e professioni (PMI)

La rilevazione sulle piccole e medie imprese e sull'esercizio di arti e professioni è una indagine annuale che raccoglie una serie di informazioni riguardanti la struttura e la competitività delle imprese con 1-99 addetti che esercitano attività economica compresa tra le divisioni 10 e 92 della classificazione NACE Rev.1.1 (con esclusione delle divisioni 65, 66 e 75).

Obiettivo dell'indagine è ottemperare al regolamento comunitario n. 58/97 (SBS) sulle statistiche strutturali e soddisfare le esigenze dei principali utilizzatori nazionali. Queste riguardano la stima di medie e totali relative ad alcune variabili di interesse per differenti sottopopolazioni (domini di studio). L'unità di rilevazione e di analisi è l'impresa, su cui sono raccolte informazioni base di natura quantitativa in forma molto dettagliata per il conto economico (valore e costi della produzione),

L'occupazione (per categoria professionale e per sesso), l'articolazione del costo del lavoro del personale dipendente, gli investimenti effettuati nell'esercizio per categoria di beni, oltre ad una serie di informazioni multiscopo di natura prettamente qualitativa (utilizzo di tecnologie dell'informazione e della comunicazione, il commercio elettronico, l'attività di innovazione e di ricerca e sviluppo, la struttura di controllo dell'impresa, ecc.).

L'indagine è campionaria, con un disegno di campionamento ad uno stadio stratificato con selezione delle unità con probabilità uguali; gli strati sono definiti dalla concatenazione delle modalità delle variabili 'classi di attività economica' (NACE Rev.1.1 a 4 cifre), 'classi di addetti' e 'regione'.

L'universo oggetto di indagine è costituito dalle imprese attive nell'anno di riferimento. L'archivio di estrazione della lista delle unità campionate è rappresentato dall'Archivio Statistico delle Imprese Attive (ASIA), costruito sulla base dell'integrazione di varie fonti, di carattere sia amministrativo sia statistico. L'indagine è condotta mediante invio postale del modello di rilevazione. I dati delle imprese rispondenti sono sottoposti a revisione, a controlli di coerenza e compatibilità, a tecniche di trattamento delle mancate risposte parziali e delle osservazioni anomale.

5.2. Aspetti critici della rilevazione PMI per gli indicatori di mancata risposta e copertura in SIDI

La notevole crescita degli obblighi di produzione e diffusione di statistiche strutturali ad un elevato livello di disaggregazione (settoriale, dimensionale e territoriale), impone la necessità di estrarre un campione sufficientemente numeroso (in media di 120.000 imprese all'anno) per dare una significatività statistica alle stime e soddisfare le molteplici richieste informative.

L'accuratezza di tali statistiche e in generale la qualità dell'informazione prodotta risulta tuttavia condizionata da molteplici fattori, come illustrato sinteticamente nello schema 1.

In particolare, per quanto riguarda le componenti della qualità che vengono quantificate in SIDI con il calcolo degli indicatori di mancata risposta e di copertura, si possono fare le seguenti osservazioni circa alcuni aspetti critici nella rilevazione PMI che influenzano il livello di tali indici.

La mancata risposta: negli ultimi anni la rilevazione PMI ha evidenziato un significativo e progressivo recupero del tasso di risposta: considerando i dati rilevati dall'indagine PMI 2004, il tasso di risposta, misurato dal rapporto fra la numerosità dei questionari ritornati e quelli spediti (al netto dei casi per i quali non è stato possibile instaurare un contatto), è stato del 51,8%.

L'analisi dello stesso indicatore di mancata risposta evidenzia una maggiore propensione alla risposta delle imprese localizzate nel nord-est (61,4%), fenomeno che diminuisce al crescere della dimensione dell'impresa. Nel corso degli ultimi anni si è assistito ad un miglioramento tendenziale del tasso di risposta (dal 44,4% del 1998 al 51,8% del 2004) che risulta, comunque, insoddisfacente.

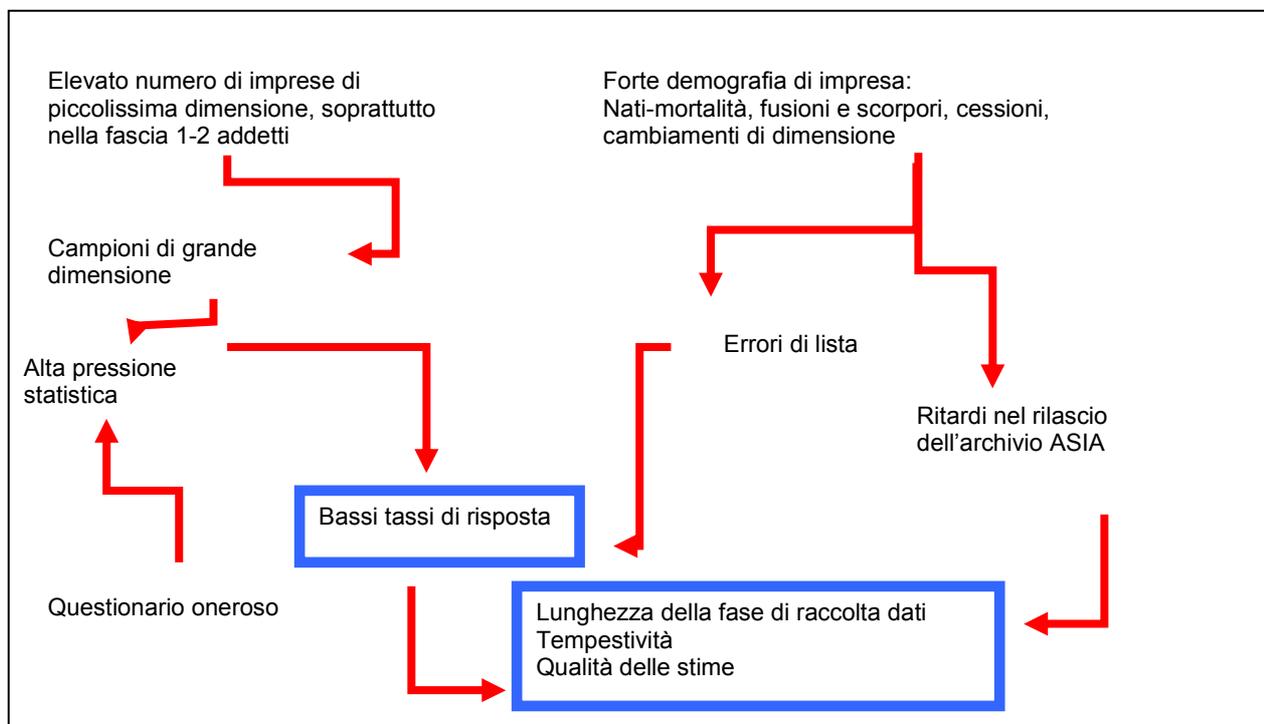
Nel corso degli anni è notevolmente aumentata la pressione statistica complessiva, ossia l'onere della risposta alle molteplici richieste d'informazioni che pervengono alle imprese da parte degli enti del Sistema Statistico Nazionale. A partire dall'anno di riferimento 1998 sono state implementate tecniche di selezione coordinata per fare in modo che nel campione estratto fossero presenti in percentuale minima unità già intervistate nelle rilevazioni PMI relative ad anni precedenti o in altre rilevazioni ISTAT aventi come oggetto lo stesso universo d'indagine.

Tuttavia, l'elevata percentuale di strati dell'universo costituiti da poche unità (circa il 44% di essi risulta composto da non più di cinque unità) rende difficile effettuare una efficiente rotazione del campione nello strato per ridurre la pressione statistica sulle unità, con conseguente possibile incremento della mancata risposta totale.

Nello stesso tempo, la prevalenza delle imprese di piccolissima dimensione (il 95% unità dell'universo appartengono alla fascia dimensionale 1-9 addetti e per circa il 35% si tratta di imprese di un solo individuo) rende difficile il contatto a causa della forte nati-mortalità.

Gli errori di lista: la disponibilità della versione aggiornata di ASIA un anno dopo il periodo di estrazione del campione determina un disallineamento tra l'archivio di selezione delle unità e la popolazione d'interesse e costituisce fonte di potenziale distorsione delle stime.

Altra fonte non campionaria di errore è legata alla demografia delle imprese, in particolare alla elevata nati-mortalità che modifica ogni anno di circa 100.000 unità il campo di osservazione, e in misura minore alle migrazioni delle imprese tra strati, dovute prevalentemente alle variazioni nel numero di addetti.



Schema 1 - Aspetti critici della rilevazione PMI per gli indicatori di mancata risposta e copertura

5.3. Gli indicatori standard di qualità di SIDI calcolati col sistema integrato

A partire dalla seconda metà degli anni '90, la rilevazione sulle piccole e medie imprese (PMI) ha attraversato una complessa fase di ristrutturazione volta a migliorare la qualità complessiva dell'informazione statistica fornita dall'indagine, tenendo conto da un lato della complessità delle esigenze informative che deve soddisfare, e dall'altra della necessità di ridurre l'onere e fastidio statistico che sempre meno le imprese sono disposte a tollerare.

Le azioni intraprese per il miglioramento della qualità si sono tradotte in una serie di interventi che hanno riguardato: la ristrutturazione di procedure e strumenti di indagine, la ridefinizione del questionario, la progettazione del disegno di campionamento.

Nello stesso tempo, l'implementazione del sistema di monitoraggio del pervenuto per le indagini strutturali sulle imprese, con l'acquisizione dei motivi della non risposta già durante la fase di rilevazione, ha fornito la possibilità di calcolare in maniera più precisa ed univoca gli indicatori standard sullo stato delle liste di rilevazione gestite da SIDI.

L'andamento temporale degli indicatori di mancata risposta e copertura fornisce una misura indiretta dell'efficacia dei controlli e delle innovazioni introdotte nel processo di produzione, consentendo di quantificare con maggiore esattezza l'incidenza delle varie componenti dell'errore non campionario e quindi l'importanza dei fattori critici descritti sulla qualità complessiva della rilevazione.

Dal punto di vista del referente d'indagine, l'immissione in SIDI degli indicatori di qualità a partire dagli esiti di risposta registrati tramite il sistema di monitoraggio del pervenuto, ha richiesto fino ad ora:

- la predisposizione di un programma SAS per il calcolo del totale delle imprese (rispondenti e non), secondo la modalità di risposta (o non risposta) e alla suddivisione in macro regioni geografiche;

- l'esportazione del data set con i valori assoluti precedentemente determinati in un file ASCII *.txt* con uno specifico formato per il caricamento in SIDI;
- la validazione dei dati immessi in SIDI e l'eventuale controllo della quadratura dei totali e della coerenza degli indicatori.

Particolare attenzione è richiesta nella fase di aggregazione dei totali per codici di risposta da cui si determinano i valori assoluti dei vari indicatori di mancata risposta e copertura, secondo le definizioni adottate in SIDI.

Nella costruzione degli aggregati, infatti, possono essere erroneamente inseriti totali relativi a codici di risposta che in realtà non concorrono alla definizione dell'indicatore corrispondente, essenzialmente per due ordini di motivi: l'errata interpretazione dei codici o delle definizioni da parte del referente che effettua il calcolo; interventi sul sistema di monitoraggio del pervenuto e modifiche nella codifica delle modalità di risposta di cui il referente può non essere a conoscenza.

Non necessariamente gli errori nel calcolo degli indicatori sono visibili al termine dell'inserimento degli stessi in SIDI: la fase di convalida, infatti, può andare a buon fine anche quando gli aggregati non sono costruiti secondo l'esatta definizione, per effetto di compensazioni tra i vari sub-totali che li compongono.

La sperimentazione del sistema integrato di interfaccia tra sistema di monitoraggio del pervenuto e SIDI ha messo soprattutto in evidenza come la possibilità di incorrere in tali problemi è praticamente annullata dal sistema di interfaccia stesso, sia perché esso effettua in maniera automatica e trasparente la transcodifica delle modalità di risposta tra i due sistemi informativi, sia per la costruzione degli aggregati su cui si costruiscono gli indicatori di qualità dell'indagine secondo una definizione univoca adottata da tutte le indagini.

Ciò permette una standardizzazione del calcolo degli indicatori che facilita sia l'analisi del loro andamento nel tempo per una particolare indagine, sia il confronto degli stessi tra indagini diverse.

6. Conclusioni

L'adozione di sistemi generalizzati di gestione dei processi di indagine condivisi da più indagini, come il sistema per il monitoraggio del pervenuto in uso presso l'SSI, è una opportunità per sviluppare progetti di integrazione con il sistema di documentazione delle indagini SIDI.

La gestione centralizzata e univoca della predisposizione dei dati per il calcolo degli indicatori di qualità in SIDI ha come principali ricadute i seguenti due vantaggi:

- previene possibili errori nel calcolo degli indicatori;
- agevola il lavoro dei referenti SIDI incaricati del calcolo degli indicatori;

Infatti, considerando che chi gestisce il monitoraggio del pervenuto (utente e coordinatore) e chi calcola gli indicatori di qualità in SIDI potrebbero non essere la stessa persona, avere dei sistemi di calcolo gestiti a livello centrale, contribuisce a prevenire possibili errori soprattutto in relazione ad eventuali cambi nel sistema di codifica.

L'ampliamento del sistema di monitoraggio del pervenuto per la predisposizione degli indicatori in SIDI è stato sviluppato per quasi tutte le rilevazioni che lo utilizzano e verificato per la Rilevazione sulle piccole e medie imprese e sull'esercizio delle arti e professioni.

Il progetto quindi non si può considerare concluso. Da una parte c'è da determinare come comportarsi per l'indagine PRODCOM e cosa fare relativamente all'inclusione di alcuni indicatori non obbligatori sulla risposta al primo contatto e ai contatti successivi, dall'altra è necessario un periodo di test per verificarne il corretto funzionamento su tutte le rilevazioni.

Infatti, l'indagine PRODCOM al momento non può usufruire dell'integrazione in quanto, diversamente dalle altre, invia i questionari non direttamente alle imprese ma alle unità locali produttive, ed è pertanto necessario valutare se calcolare gli indicatori per unità locale o per impresa e, in questo ultimo caso,

introdurre una regola per stabilire quando l'impresa diventa rispondente. Questa variazione non presenta particolari difficoltà e sarà implementata se la fase di test sulle indagini darà risultati positivi. Inoltre, al momento, nel sistema di monitoraggio del pervenuto, esiste un campo che in via teorica potrebbe essere sfruttato per calcolare il tasso di risposta al primo contatto e quello ai contatti successivi (indicatori facoltativi in SIDI). Bisognerà però valutare la qualità dell'informazione di questo campo per tutte le indagini, prima di considerarlo utile per il calcolo dei due tassi di risposta (al primo contatto e a contatti successivi).

Relativamente al monitoraggio sull'utilizzo dell'applicazione integrata, bisogna considerare che la correttezza dei risultati dipende dall'accuratezza che viene posta nel codificare le risposte e dall'utilizzo del data set corretto, ossia quello ottenibile alla fine della fase di rilevazione. Se viene attivata la funzione di creazione dei dati per gli indicatori di qualità, ogni modifica sullo stato delle imprese apportata al sistema successivamente a tale attivazione non sarà automaticamente acquisita nell'indicatore. Sarà quindi necessario delineare una semplice procedura che stabilisca i tempi e i ruoli per le persone coinvolte nelle attività di monitoraggio del pervenuto e per i referenti SIDI, in modo che l'attività di calcolo degli indicatori sia coordinata.

7. Bibliografia

- Brancato G, Fortini M, Signore M. (1997) "Il controllo delle mancate risposte totali nell'indagine sulle piccole imprese". *Atti del convegno La Statistica per le Imprese*, Comunicazioni Spontanee, volume 2. Convegno SIS. Torino 2-4 Aprile 1997.
- Brancato G, Pellegrini C, Simeoni G. e Signore M. (2004) "Standardising, Evaluating and Documenting quality: the implementation of Istat information system for survey documentation – SIDI". *Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004)*. Mainz, 24-26 May 2004. CD ROM.
- Brancato G, Signore M. (1988) "Il controllo della fase di rilevazione per l'indagine sulle Piccole Imprese e Professioni Liberali". Istat, *Contributi* n.15, Anno 1998.
- EUROSTAT (1996), "The Council Regulation (EC, EURATOM) No 58/97 concerning structural business statistics".
- Hidiroglou MA, Drew JD, Gray GB (1993) "A Framework for Measuring and Reducing Nonresponse in Surveys". *Survey Methodology*, 19, 1, 81-94.
- ISTAT (1991), "Classificazione delle attività economiche", Metodi e norme, serie C-n.11.
- ISTAT, (1999), "Conti economici delle imprese", Informazioni.

Misure della qualità dei dati per una statistica di fonte amministrativa: i condannati con sentenza irrevocabile

Daniela Capelli (*), Laura Rossetti(**), Franco Turetta(**)

(*) ISTAT, DCIS/A (Organizzazione dei processi e sistemi informativi)

(**) ISTAT, DCIS/GIU/A (Servizio Giustizia)

Introduzione

La statistica dei condannati per delitto con sentenza irrevocabile è una tra le numerose indagini di tipo amministrativo svolta all'interno del Servizio giustizia dell'Istat. Spesso si incontrano delle difficoltà ove si cerchi di valutare il livello di qualità che caratterizza questo tipo di indagini: infatti non sempre i principali indicatori riguardanti la qualità adottati in sede europea possono essere calcolati per le statistiche non campionarie, ed in particolar modo per quelle di fonte amministrativa.

Allo scopo di ottenere indicazioni il più possibile complete sulla qualità di tale statistica, da fornire agli utilizzatori finali, si sono utilizzati due approcci. Il primo, essenzialmente qualitativo, consiste in una checklist approntata dall'ONS (Office for National Statistics), della quale è stata considerata la sola sezione riguardante le statistiche di fonte amministrativa. La checklist è basata sulle sei componenti della qualità adottate in sede europea (Rilevanza; Accuratezza; Tempestività e Puntualità; Accessibilità e Chiarezza; Comparabilità; Coerenza). Il secondo approccio, quantitativo, ha comportato invece l'utilizzo di un software specifico per valutare l'impatto delle procedure di controllo e correzione, di natura deterministica, realizzate tramite programmi SAS. A tal fine sono state modificate in parte le procedure di trattamento dei dati per potere ottenere due data set contenenti record in forma sequenziale, l'uno con i dati acquisiti dal casellario, l'altro con i medesimi record modificati dalle procedure di controllo e correzione applicate. A scopo conoscitivo sono state incluse anche alcune variabili di tipo sociale non considerate ai fini della diffusione in quanto largamente carenti.

Sui due data set è stato applicato il software IDEA (Indices for Data Editing Assessment), allo scopo di valutare la qualità delle procedure di controllo e correzione ed il loro impatto.

1. La statistica dei condannati con sentenza irrevocabile

La statistica degli imputati condannati riguarda l'insieme degli individui condannati in qualsiasi fase o tipo di giudizio, con riferimento al momento in cui, divenuto irrevocabile il provvedimento di condanna, vi è l'iscrizione al Casellario giudiziale centrale.

Dal 1996 i dati si riferiscono ai condannati per reati (delitti e contravvenzioni) tentati o consumati, previsti dal codice penale e dalle altre leggi, iscritti presso il Casellario giudiziale centrale. Negli anni precedenti la rilevazione riguardava anche i prosciolti, era limitata ai soli delitti ed era condotta presso le Cancellerie degli uffici giudiziari giudicanti.

In caso di concorso di delitti, il condannato viene preso in considerazione con riferimento al delitto per il quale è prevista dal Codice penale e dalle altre leggi la pena più grave; nel caso invece di concorso di contravvenzioni, il condannato viene classificato in base alla prima contravvenzione iscritta nel Casellario.

In presenza infine di più reati non in concorso tra di loro, l'individuo figura tante volte quante sono le sentenze irrevocabili di condanna a suo carico pronunciate nel periodo considerato.

Nell'anno 2005 sono stati iscritti nel Casellario giudiziale centrale 221.381 condannati con sentenza divenuta irrevocabile per delitto e 81.380 condannati per contravvenzione (i condannati sia per delitto che per contravvenzione sono computati una sola volta tra i condannati per delitto).

La statistica fornisce principalmente indicazioni di natura giudiziaria relative al condannato, con una limitata presenza di variabili di tipo socio-demografico.

I dati estratti dal Casellario giudiziale centrale consentono, in un'ottica statistica, due approcci principali:

- analisi delle caratteristiche delle persone condannate;
- analisi dei delitti commessi dalle persone condannate.

Normalmente si privilegia la prima chiave di lettura, classificando preliminarmente i condannati secondo il delitto più grave commesso, utilizzando il criterio della pena edittale media¹, ma per studi

¹ La pena edittale media viene calcolata come la semisomma delle pene minima e massima previste dalla Legge per un determinato delitto. La classificazione in base alla pena media edittale (uno dei possibili criteri) consente di identificare

mirati su reati specifici, è possibile prendere in considerazione sia l'insieme dei condannati che ha commesso un particolare delitto (ed eventualmente altri delitti concorrenti, anche più gravi), sia le tipologie dei reati concorrenti con il delitto oggetto di studio. L'archivio amministrativo del Casellario giudiziale centrale è caratterizzato da una architettura di tipo relazionale e contiene, per ogni condannato, fino a sette tipi di record, come riportato nel seguente prospetto 1.

Prospetto 1 – Tipologia e contenuto dei record dell'archivio dei condannati, presso il Casellario giudiziale centrale del Ministero della giustizia

TIPO RECORD	CONTENUTO
DECISIONE	Tipo e sede dell'Autorità che ha emesso la sentenza; Data del passaggio in giudicato; Presenza di precedenti penali del condannato; ...
ANAGRAFICO	Sesso; Data e luogo di nascita; Stato civile; Istruzione; Professione; (variabili scarsamente alimentate) ...
REATO (max 199)	Estremi identificativi del reato (Anno e numero della legge, articolo ecc.); Tipo di reato (delitto o contravvenzione); Natura del reato (tentato, consumato, continuato ecc.); Data e luogo del commesso reato; ...
CIRCOSTANZE DEL REATO (max 14)	Estremi identificativi della circostanza (Anno e numero della legge, articolo ecc.) esempio: art. 62 del Codice penale = circostanze attenuanti comuni;
PENE INFLITTE / SANZIONI SOSTITUTIVE (1)	Pena base del delitto (o contravvenzione); Durata reclusione inflitta (durata arresto per le contravvenzioni); Ammontare multa inflitta (ammontare ammenda per le contravvenzioni); ... Tipo di diminuzione di pena (es. applicazione pena su richiesta delle parti); Tipo di sanzione sostitutiva (es. semidetenzione); ...
PENE ACCESSORIE	Tipo di pena accessoria (es. interdizione dai pubblici uffici); Durata della pena accessoria.
MISURE DI SICUREZZA O PREVENZIONE	Tipo di misura di sicurezza o prevenzione (es. libertà vigilata; sorveglianza con obbligo di soggiorno); Durata della misura di sicurezza o prevenzione.
(1) Un solo record per condannato in caso di pena inflitta per il complesso dei reati o ritenendo la continuazione tra i reati, altrimenti un record per ogni reato ascritto	

il delitto "più grave" commesso dal condannato per più delitti. E' pertanto possibile predisporre tavole statistiche nelle quali ogni condannato è computato una sola volta, in relazione al delitto più grave commesso.

Secondo le linee guida dell'Istat², "si definiscono dati amministrativi le informazioni raccolte e conservate da istituzioni pubbliche ai fini di controllo o di intervento nei confronti di singoli individui o entità di altro tipo (es. persone giuridiche). Il fine, amministrativo, della raccolta di tali dati ne contraddistingue quindi la natura."

Si può di conseguenza definire "indagine di fonte amministrativa" l'indagine che utilizza i dati - precedentemente raccolti a scopo gestionale-amministrativo - con finalità di tipo statistico.

Un'indagine di fonte amministrativa può essere condotta dallo stesso Ente che effettua la raccolta dei dati a fini amministrativi o da altri Enti, utilizzando in tutto o in parte tale patrimonio informativo, e ponendolo eventualmente in relazione con dati provenienti da altre fonti, amministrative o meno.

Il dato di fonte amministrativa rappresenta una risorsa conoscitiva di estremo interesse, in quanto la sua trasformazione in dato statistico comporta normalmente un costo estremamente contenuto, non dovendosi in questo contesto computare i costi sostenuti per la raccolta delle informazioni, già effettuata per altri scopi.

Proprio perché la raccolta di dati è già stata effettuata con altri fini, tuttavia, vi possono essere dei limiti anche seri rispetto ai risultati che si potrebbero avere rispetto ad un'indagine progettata ed eseguita a fini statistici.

A riprova di quanto possano influire sulla qualità del dato rilevato modifiche apparentemente poco rilevanti, si sono esaminate a posteriori le implicazioni del passaggio tra la rilevazione cartacea condotta precedentemente (fino al 1995) e quella attuale.

La rilevazione cartacea, pur utilizzando un modello progettato con un'ottica statistica, era anch'essa un'indagine amministrativa, avendo come rispondenti i cancellieri degli uffici giudiziari, che attingevano i dati richiesti dai registri riportanti le risultanze processuali (fonte più ricca e varia rispetto alle notizie presenti nel Casellario giudiziale, che ne sono un sottoprodotto parziale, ancorché strutturato).

Per la statistica dei condannati per delitto con sentenza irrevocabile l'abbandono del modello cartaceo ha comportato essenzialmente:

- riflessi sulla copertura: la popolazione di riferimento dell'indagine attuale (i condannati per delitto con sentenza irrevocabile) è un sottoinsieme di quella indagata dal modello cartaceo (Istat M.315). Non sono infatti più compresi gli imputati giudicati e prosciolti per tutti i delitti loro ascritti, per i quali non è prevista l'annotazione nel Casellario.

Per contro, dalla fonte amministrativa è stato possibile trarre contestualmente anche i condannati per contravvenzione (delitti e contravvenzioni formano il complesso dei reati) in precedenza non rilevati. Si è quindi di fatto passati da una rilevazione esaustiva degli imputati per delitto condannati o prosciolti ad una rilevazione - anch'essa esaustiva - dei condannati per reato, distinti secondo il tipo (delitto o contravvenzione) di reato commesso.

- riflessi sull'accuratezza: vi è una mancanza pressoché totale di informazioni per alcune variabili di carattere sociale, marginali ai fini giudiziari e della tenuta amministrativa dell'archivio, ma la cui presenza sarebbe auspicabile a fini statistici. Attualmente, se si considerano le sette variabili socio-demografiche presenti nel record anagrafico di ogni condannato (sesso, data di nascita, luogo di nascita, cittadinanza, stato civile, istruzione, professione) solo le prime tre sono alimentate in proporzione tale da fornire un'informazione valida e poter essere divulgate. Precedentemente, si aveva una buona alimentazione (80% o superiore) delle prime 5 variabili. E' opportuno ricordare che con l'acquisizione dei dati dall'archivio informatizzato è venuto meno il contatto diretto tra il personale dell'Istat ed i cancellieri che trascrivono le notizie richieste dai registri.

Un importante aspetto positivo, invece, è costituito dal fatto che la modificazione del sistema di rilevazione ha consentito di acquisire tutte le diverse qualificazioni giuridiche del reato. Infatti il modello cartaceo, che recava un elenco dei principali delitti da indicare, consentiva la distinzione solo di alcune tipologie. Si è avuto quindi un notevole incremento conoscitivo potendo correttamente dettagliare anche i delitti che precedentemente ricadevano nella categoria indifferenziata "altri delitti". Inoltre, l'affrancamento da una lista di delitti predisposta e quindi statica, consente di meglio gestire la

² <http://www.istat.it/strumenti/metodi/lineeguida/22.htm>

classificazione dei delitti previsti da nuove leggi intervenute nel corso del tempo, a condizione di tenere costantemente aggiornato un archivio delle qualificazioni giuridiche di reato. Quest'ultima attività, alquanto onerosa stante la quantità di leggi in materia penale approvate o modificate annualmente, viene svolta all'interno del Servizio giustizia dell'Istat.

- riflessi sulla tempestività: l'acquisizione dei modelli cartacei dalle singole cancellerie degli uffici giudiziari era previsto mensilmente. Questo avveniva, più che per fini divulgativi, per poter meglio cadenzare le fasi di lavorazione dei modelli, molto onerose. Per l'estrazione dei record dal Casellario giudiziale centrale è stata invece concordata una cadenza annuale. Va tuttavia precisato che non ci sarebbero motivi tecnici ostativi ad un'acquisizione dei dati più frequente.

- riflessi sui costi: accanto alle ripercussioni sulla qualità, è corretto evidenziare che l'utilizzo di una fonte amministrativa esistente ha comportato nel caso dei condannati un risparmio economico molto consistente³.

2. Misure della qualità dei dati nelle indagini amministrative: l'approccio ONS

La "qualità del dato statistico" può utilmente essere definita in relazione alla capacità che il dato stesso ha di soddisfare le esigenze e le aspettative dell'utilizzatore. Tale tipo di approccio ha come corollario l'importanza del fornire all'utente ogni informazione che possa aiutarlo a valutare autonomamente se il dato statistico prodotto possiede o meno i requisiti qualitativi necessari per i propri intenti.

Non sempre gli indicatori più in uso riguardanti la qualità possono essere calcolati per le statistiche non campionarie, ed in particolar modo per quelle di fonte amministrativa. Nel tentativo di valutare la possibilità di produrre un'informazione più ricca e più specificatamente mirata alle statistiche amministrative, si è tratto spunto dalle linee guida per la misurazione della qualità statistica dell'ONS⁴.

Viene proposta, tra i vari approcci, una checklist di indicatori e informazioni sulla qualità della rilevazione, distinta per fasi dell'indagine (prospetto 2), e che dall'aprile 2006 si è arricchita di una sezione relativa ai dati di origine amministrativa. Le informazioni attengono alle sei dimensioni qualitative adottate in sede europea: Rilevanza (R); Accuratezza (A); Tempestività e Puntualità (T); Accessibilità e Chiarezza (AC); Comparabilità (C); Coerenza (CH).

³ Nell'ottica di semplice esercizio matematico, si è calcolata la differenza tra i costi imputabili alle sole risorse umane impiegate (utilizzando i salari di ingresso per livello professionale) nell'indagine cartacea ed in quella attuale, registrando nel passaggio un risparmio pari a circa l'80% del totale.

⁴ Office for National Statistics: Guidelines for measuring statistical quality, version 3.0, 2006, London

Prospetto 2 – Numero di indicatori della checklist proposta dall’ONS, per fase dell’indagine e dimensioni della qualità

FASI DELL'INDAGINE	Indicatori						Totale	
	Dimensioni della qualità							
	R	A	T	AC	C	CH		
B1	Disegno dell'indagine	10	12	-	-	4	-	26
B2	Dati di origine amministrativa	6	7	3	3	2	3	24
B3	Raccolta dei dati	-	19	-	-	-	-	19
B4	Trattamento dei dati	-	13	-	-	-	-	13
B5	Pesi campionari e stima degli errori	-	13	-	-	-	-	13
B6	Analisi delle serie storiche	-	11	-	-	7	-	18
B7	Misure prese per assicurare la riservatezza del dato	3	2	-	8	1	1	15
B8	Divulgazione dei dati	-	-	8	11	1	-	20
Totale		19	77	11	22	15	4	148

Fonte: ONS

Viene qui di seguito esaminata in dettaglio la sola sezione (B2) riguardante le statistiche di fonte amministrativa, ai cui quesiti si è tentato di dare una risposta di massima, relativamente alla statistica dei condannati per delitto con sentenza irrevocabile. In alcuni casi, ove una trattazione esauriente avrebbe richiesto uno spazio eccessivo, le risposte non sono esaustive.

SEZIONE B2 – DATI DI FONTE AMMINISTRATIVA

INFORMAZIONI SULLA FONTE DATI AMMINISTRATIVA (B2.1-B2.12)⁵

RILEVANZA (B2.1-B2.3)

B2.1 - Descrizione dei principali utilizzi (a fini statistici) dell'Archivio Amministrativo

I dati del Sistema Informativo del Casellario Giudiziale (SIC) possono essere utilizzati a fini statistici per l'analisi della criminalità giudicata, per l'analisi di impatto di leggi o regolamenti, per la valutazione dell'efficacia del sistema sanzionatorio italiano, per l'analisi dell'esecuzione delle pene.

I dati vengono forniti in modo anonimo e a fini statistici (art. 19 D.P.R. 313/2002) al Ministero della Giustizia; all'ISTAT per la redazione delle Statistiche sui condannati con sentenza passata in giudicato; alla Presidenza del Consiglio dei Ministri per la redazione della Relazione al Parlamento sullo stato delle tossicodipendenze; alle Pubbliche Amministrazioni che ne facciano richiesta.

⁵ Si ringrazia l'Ufficio Centrale del Casellario - Ministero della Giustizia - Dipartimento per gli Affari di Giustizia - Direzione Generale della Giustizia Penale, per la collaborazione prestata nella stesura delle risposte da B2.1 a B2.12.

B2.2 - Motivo principale della raccolta dei dati nell'Archivio Amministrativo

Il principale fine della raccolta dei dati del Casellario è quello di garantire il servizio della certificazione regolamentato prima dal Regio decreto n. 778 del 1931 e ora dal D.P.R. n.313 del 2002. Il servizio è assicurato su base territoriale dai Casellari giudiziari (ora Uffici Locali) presso le Procure della Repubblica, i quali interrogano il SIC per estrarre certificati (Generale, Penale, Civile) su richiesta dell'Autorità Giudiziaria, della Pubblica Amministrazione o del Privato cittadino.

B2.3 - Concetti, definizioni e classificazioni della popolazione amministrativa e delle variabili nell'Archivio Amministrativo

Definizione delle popolazioni di riferimento: i soggetti condannati con sentenza passata in giudicato, o i provvedimenti iscritti (a seconda del tipo di analisi statistica che si vuol fare).

Classificazioni: Sui soggetti non ci sono classificazioni definite a priori, sui provvedimenti ci sono due tipi di classificazioni:

- a) principali (sentenze) ed esecutivi (provvedimenti dell'esecuzione riferiti alle sentenze)
- b) per tipologia di provvedimento: in funzione della materia (decisione penale, misura di sicurezza, misura di prevenzione, interdizione, inabilitazione, ...).

ACCESSIBILITA' (B2.4 – B2.6)

B2.4 - Metadati forniti insieme all'Archivio Amministrativo e metadati non forniti

Quando vengono forniti dei dati vengono fornite le descrizioni strettamente attinenti alla lettura degli stessi (ad esempio all'ISTAT viene fornita in modo codificato l'autorità giudicante e insieme la descrizione dell'autorità).

B2.5 Procedure di raccolta dei dati amministrativi

I dati sono costituiti da estratti di provvedimenti (es. sentenze di condanna passate in giudicato e provvedimenti dell'esecuzione riferiti alle stesse, decreti penali di condanna...) che, riportati su appositi moduli (sentenza di condanna = scheda, provvedimento dell'esecuzione = foglio complementare) dall'ufficio giudiziario che ha emesso il provvedimento, sono inviati al Casellario presso la Procura del luogo di nascita del condannato (per gli stranieri è competente il Casellario di Roma).

I Casellari, 165 uffici incardinati presso le Procure della Repubblica, provvedono all'inserimento dei dati, contenuti negli estratti nel Sistema Informativo del Casellario (SIC).

La recente normativa (D.P.R. n. 313/2002) ha introdotto una nuova regolamentazione del servizio, che prevede l'iscrizione dei dati nel sistema informativo centralizzato del casellario direttamente dagli uffici giudiziari che emettono i provvedimenti giudiziari iscrivibili (es. le sentenze o i provvedimenti dell'esecuzione riferiti alle stesse).

Il nuovo sistema informativo centralizzato, che da attuazione al citato D.P.R. sarà in esercizio a decorrere dal febbraio 2007.

B2.6 - Formato in cui i dati amministrativi sono resi disponibili

Generalmente in formato ASCII, Excel o Access.

ACCURATEZZA (B2.7 – B2.10)

B2.7 – Descrizione di eventuali problemi di copertura dell' Archivio Amministrativo accompagnata se possibile da stime: frazione di copertura; frazione di unità duplicate; frazione di unità erroneamente classificate

La copertura dovrebbe essere totale, ogni anno si registra dell'arretrato, ma con il tempo si perviene alla completezza dei dati.

Non siamo in grado di definire una frazione di copertura; non esistono unità duplicate; non ci sono unità classificate erroneamente in quanto non esiste una classificazione a monte.

B2.8 - Descrizione delle fonti di errori note dei dati amministrativi (accompagnata se possibile da tassi: stime tassi errori di codifica, stime tassi errori di lettura ottica;...)

Gli errori sono quelli legati alla digitazione degli operatori in un sistema basato sulla codifica alfanumerica.

Il Nuovo Sistema Informativo, progettato in attuazione del recente Regolamento di cui al D.P.R. 313/2002, prossimo all'avvio consentirà una notevole riduzione di tali errori in quanto le maschere di data entry riducono al minimo la digitazione di dati e offrono controlli ben strutturati a livello di campi dove avviene la digitazione. Non abbiamo una stima media degli errori di codifica.

B2.9 - Frazione di record che presentano almeno un valore mancante

I dati risultano abbastanza esaustivi per le variabili amministrative importanti, molto carenti per le variabili sociali.

B2.10 Proporzioni di valori mancanti per le variabili più importanti

Non ci sono valori mancanti per le variabili più importanti.

TEMPESTIVITA' (B2.11)

B2.11 - Informazioni sui tempi di aggiornamento dei dati dell'archivio Amministrativo

Non siamo in grado di fornire una stima precisa di questi tempi.

COERENZA (B2.12)

B2.12 - Presenza di variabili identificative delle unità che possono essere utilizzate in procedure di linkage

Dati anagrafici, codice fiscale e codice unico identificativo per gli stranieri nel nuovo sistema.

INFORMAZIONI SULLE RISULTANZE STATISTICHE DERIVATE IN TUTTO O IN PARTE DA FONTE DATI AMMINISTRATIVA (B2.13-B2.24)

RILEVANZA (B2.13-B2.15)

B2.13 - Elenco di tutte le fonti di dati di tipo amministrativo utilizzate come input dalla statistica in oggetto.

FONTI PRINCIPALI

- Condannati iscritti presso il Casellario Giudiziale Centrale

Natura: Archivio amministrativo informatizzato

Ente titolare: Casellario Giudiziale Centrale (Ministero della Giustizia)

- Classificazione dei delitti

Natura: Classificazione amministrativa ricavata da testi normativi essenzialmente cartacei

Ente titolare: Istat (Servizio giustizia, di concerto con il Ministero della Giustizia)

FONTI COMPLEMENTARI

- Classificazione territoriale

Natura: Classificazione amministrativa (regioni, province, comuni)

Ente titolare: Istat

- Classificazione territoriale

Natura: Classificazione amministrativa (Corti d'appello, Tribunali, ecc.)

Ente titolare: Ministero della Giustizia

- Classificazioni giudiziarie di vario tipo

Natura: Classificazioni amministrative (es. modalità delle singole variabili, come la tipologia degli Uffici giudiziari, ecc.)

Ente titolare: Ministero della Giustizia

- ...

B2.14.) - Indicare in che misura i dati di fonte amministrativa rispondono ai requisiti statistici

I dati rispondono soddisfacentemente alle esigenze statistiche, ma si possono evidenziare le seguenti problematiche:

- alcune notizie di tipo sociale (cittadinanza, stato civile, istruzione, professione) risultano estremamente carenti e non possono essere utilizzate a fini statistici;

- ci può essere una velocità di afflusso differenziale delle annotazioni provenienti dai Casellari giudiziari locali che riduce la confrontabilità temporale e territoriale.

B2.15 - Indicare eventuali limitazioni che impediscano di ottenere il dato amministrativo con il dettaglio richiesto

Non si evidenziano problematiche di questo tipo.

ACCURATEZZA (B2.16-B2.18)

B2.16 - Descrizione dei processi applicati agli archivi amministrativi

Vengono utilizzate alcune procedure informatiche atte a verificare e/o migliorare la qualità del dato amministrativo. Senza affrontare una descrizione troppo dettagliata, sono attualmente previsti programmi SAS di check che segnalano incongruenze (pene irrogate sproporzionate rispetto al delitto commesso, età del condannato incompatibile con i limiti di legge o in contrasto con la tipologia dell'ufficio giudicante ecc.) ed è prevista in un immediato futuro la trasmissione sistematica di tali incongruenze al Casellario. E' tuttavia da rimarcare la laboriosità dell'operazione di verifica tramite ritorno ai rispondenti, e l'impossibilità di procedere a correzioni ove il valore "anomalo" non sia frutto di un'errata trascrizione, ma compaia nel documento amministrativo originale.

B2.17 e B2.18 - Descrizione delle procedure di linkage utilizzati sugli archivi amministrativi / Proporzione di agganci corretti, falsi positivi, falsi negativi

Poiché non esiste una rilevazione statistica apposita cui l'archivio amministrativo sia di integrazione, ma tutte le informazioni sui condannati per delitto vengono direttamente tratte da un unico archivio amministrativo, non viene effettuato un linkage nel senso indicato. Si può tuttavia scegliere di considerare in questo ambito la corretta rispondenza del contenuto dell'archivio con le varie liste (normalmente di classificazione). Le classificazioni territoriali (a livello di codice comune) trovano rispondenza nella totalità dei casi (anno 2005). Sempre nel 2005, i mancati agganci tra estremi identificativi del delitto commesso dal condannato inseriti nel Casellario, e gli archivi predisposti dal Servizio giustizia, che consentono di associare ai singoli delitti altre notizie non contenute nel Casellario come la pena edittale media, sono stati lo 0,1% del totale.

TEMPESTIVITA' (B2.19 – B2.20)

B2.19 - Periodo di tempo che intercorre tra la raccolta dei dati a fini amministrativi e raccolta a fini statistici

I dati relativi ai condannati iscritti nel Casellario giudiziale centrale in un anno di calendario vengono estratti dall'archivio del Casellario giudiziale centrale – mediamente – alla fine del primo trimestre dell'anno successivo.

L'iscrizione del condannato con sentenza definitiva nel Casellario è l'ultima fase di un'iter burocratico complesso, e dipende dal volume degli inserimenti effettuati dai singoli Casellari giudiziali locali. Ciò può avere ripercussioni sulla qualità dell'informazione statistica ove si voglia riferire esattamente il dato rilevato al contingente di condannati per i quali la sentenza è divenuta esecutiva in un dato arco temporale. Questo rimane teoricamente possibile effettuando un'analisi longitudinale, in quanto nei record è presente l'indicazione della data di passaggio in giudicato della sentenza, ma a forte danno della tempestività del dato.

B2.20 – Ritardi nel rilascio dei dati amministrativi rispetto al termine previsto

Non ci sono ritardi. Il rilascio prevede preliminarmente una corrispondenza ufficiale tra l'Istat ed il Casellario giudiziale centrale, che formalizza la richiesta.

COMPARABILITA' (B2.21 – B2.22)

B2.21 - Indicare l'impatto sul dato statistico dei cambiamenti legislativi riguardanti l'archivio amministrativo

Le modifiche normative non hanno comportato particolari problemi di tipo statistico nel recente passato. Ovviamente, qualsiasi cambiamento che influisca sulle modalità di una variabile (es. istituzione o soppressione di una tipologia di ufficio giudicante) ha un impatto sulla comparabilità nel tempo di quella data variabile. Non è invece ancora possibile valutare quale sarà l'impatto dell'obbligo di alimentare l'archivio amministrativo ad opera dei singoli Uffici giudicanti, previsto nel D.P.R. n.313 del 2002.

B2.22 - Indicare le modifiche nel corso del tempo dei dati dell'archivio amministrativo e il loro effetto sul dato statistico

Le modifiche di lieve entità che si sono avute nei dati dell'archivio amministrativo a partire dal 1996 non hanno comportato particolari problemi di tipo statistico, se non negli aspetti già evidenziati precedentemente.

COERENZA (B2.23 – B2.24)

B2.23 - Indicare le differenze concettuali, nelle definizioni e nella classificazioni tra la fonte amministrativa e l'output statistico

Trattandosi di una materia estremamente disciplinata da leggi (ad es. molte classificazioni di variabili hanno come modalità tutti i possibili esiti che la legge prevede) sotto questi aspetti non vi sono differenze di rilievo.

B2.24 - Indicare gli interventi adottati a causa di differenze concettuali, nelle definizioni e nella classificazioni tra la fonte amministrativa e l'output statistico

Nel caso di modificazioni normative ci si limita ad adeguare la classificazione all'interno dell'archivio statistico. Non sono stati finora necessari interventi di maggior rilevanza.

L'approccio raccomandato dall'ONS che è stato utilizzato ha dato, nel caso della rilevazione dei condannati per delitto con sentenza irrevocabile, buoni risultati in termini di fattibilità e di chiarificazione di alcuni aspetti riguardanti la qualità, evidenziando determinate caratteristiche della componente amministrativa della rilevazione. La checklist ha un suo interesse ulteriore, in quanto nella sua compilazione è stato interessato direttamente l'Ente che conduce l'indagine a fini amministrativi, ottenendo quindi particolareggiate indicazioni sugli aspetti qualitativi delle fasi precedenti all'acquisizione delle notizie da parte dell'Istat.

3. L'utilizzo di IDEA per la valutazione della qualità dei dati dell'indagine sui condannati per delitto con sentenza irrevocabile

3.1. IDEA: finalità di impiego e modalità di utilizzo

IDEA (*Indices for Data Editing Assessment*) è un software generalizzato sviluppato all'interno dell'ISTAT per valutare, in primo luogo, l'impatto che l'applicazione di una procedura di controllo e correzione può avere sulla qualità dei dati. Oltre a questo, lo strumento permette anche di calcolare una serie di indicatori previsti all'interno del Sistema Informativo di Documentazione delle Indagini (SIDI), per la valutazione della componente non campionaria degli errori⁶.

I dati, una volta raccolti, normalmente sono affetti da errori e da valori mancanti. Prima di procedere all'elaborazione e al calcolo degli aggregati, è pratica corrente quella di effettuare una fase di individuazione dei valori errati. Una volta localizzati, questi vengono imputati assieme ai valori mancanti, in modo da ottenere un insieme di dati il più possibile completo e coerente.

Secondo una terminologia consolidata, i dati *originali* (o "grezzi") vengono *validati* (o "puliti"), ed è di grande interesse studiare la trasformazione che i dati subiscono in tale passaggio.

Se, infatti, l'impatto della procedura è limitato, questo può dipendere da due fattori:

limitata presenza di valori mancanti e/o errati nei dati originali;

limitata capacità della procedura di individuare i valori errati o mancanti e di correggerli

Se, invece, l'impatto è rilevante, anche questo può derivare da due situazioni alternative:

elevata presenza di errori e valori mancanti nei dati, che la procedura di controllo e correzione rileva correttamente e provvede a correggere o imputare

non correttezza della procedura di controllo e correzione che individua valori esatti come errati e provvede a modificarli, introducendo in questo modo nuovi errori nei dati.

L'ideale sarebbe poter disporre di una procedura di controllo e correzione di accertata validità, in modo da poter limitare le situazioni derivanti da una sua possibile inadeguatezza. Questo non sempre è possibile, per cui è necessario analizzare accuratamente i risultati dell'applicazione di IDEA per evitare interpretazioni errate.

L'impiego di IDEA è operativamente molto semplice: occorre disporre delle due versioni dei dati, quella originale e quella validata. Da un punto di vista informatico, devono essere rispettati due requisiti:

entrambi i file (SAS) devono avere gli stessi nomi delle variabili che devono essere poste a confronto;

deve essere disponibile una chiave univoca su entrambi i file, per consentire il corretto accoppiamento.

Attraverso una interazione molto semplice, l'utente dichiara quali sono i file dei dati originali e dei dati validati, ne indica la chiave di accoppiamento e le variabili rispetto alle quali è interessato a condurre l'analisi. Queste possono essere di tre diverse tipologie: nominali, ordinali o continue. Per ogni tipologia di variabile sono definiti diversi set di indicatori.

3.2. Gli indicatori di valutazione in IDEA

Gli indicatori ottenibili mediante IDEA sono, come già detto, da una parte relativi alla valutazione dell'impatto della procedura di controllo e correzione, dall'altra sono quelli di interesse per SIDI.

3.2.1. Valutazione dell'impatto della procedura

Per quanto riguarda gli indicatori relativi alla valutazione dell'impatto della procedura, questi vanno distinti relativamente a diversi aspetti:

1. *valori individuali*: per ogni variabile vengono misurate le modifiche subite individualmente, cioè record per record;

⁶ Per una descrizione esaustiva, cfr "IDEA (Indices for Data Editing Assessment) – Sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI" di G.Della Rocca, M.Di Zio, O.Luzi, G.Simeoni

2. *distribuzione semplice*: sempre per ogni variabile, viene studiato l'impatto subito dalla relativa distribuzione dei valori, cioè nel complesso del dataset;
3. *distribuzione multivariata*: viene studiato l'impatto della procedura sulla distribuzione congiunta di più variabili;
4. *associazioni tra variabili*: viene misurato l'impatto sulla struttura associativa dei dati.

3.2.1.1. Modifica dei valori individuali

Sia per le variabili nominali che per le ordinali e le continue, viene calcolata la *percentuale di imputazione*:

$$D_1 = \frac{\sum_{i=1}^n w_i I(Y_i, Y_i^*)}{\sum_{i=1}^n w_i} \times 100 \quad (1)$$

dove

Y_i : valore originale

Y_i^* : valore validato

$I(Y_i, Y_i^*) = 1$ se $Y_i \neq Y_i^*$ e $I(Y_i, Y_i^*) = 0$ se $Y_i = Y_i^*$.

Nel caso di indagini campionarie, w_i rappresenta il peso associato alla i -esima osservazione, che vale sempre 1 nel caso di rilevazioni universali.

Questo indicatore è tra i più importanti, in quanto fornisce un'idea immediata sia dell'impatto della procedura di controllo e correzione, che della qualità dei dati originali.

E' di sicuro interesse andare a specificare ulteriormente questo indicatore. Se nella formula (1) all'interno delle sommatorie anziché il numero totale dei casi n vengono presi in considerazione:

1. n_k : il numero di casi in cui la variabile Y ha inizialmente valori mancanti, imputati con un valore diverso da blank;
2. n_k : il numero di casi in cui la variabile Y ha inizialmente valori ammissibili, imputati con altri valori ammissibili;
3. n_k : il numero di casi in cui la variabile Y ha inizialmente valori ammissibili, imputati con un valore blank;

otterremmo, rispettivamente:

1. D_{1r} : percentuale di imputazione netta (ci da una misura dell'imputazione delle mancate risposte parziali);
2. D_{1m} : percentuale di modificazione (misura dell'imputazione dei valori giudicati errati dalla procedura, ed imputati di conseguenza);
3. D_{1c} : percentuale di cancellazione (misura degli errori derivanti da percorsi errati nel questionario).

Per le variabili continue sono anche disponibili una serie di altri indicatori, che qui ci limitiamo a citare:

1. gli *indicatori di distanza semplice, quadratica* e in L^∞ ;
2. il parametro β della retta di regressione $Y = \beta Y^*$ col relativo valore di R^2 .

Un cenno al secondo gruppo di indicatori: stimando un modello di regressione tra i valori originali e quelli validati di una stessa variabile Y , quanto più il parametro β è vicino a 1 e quanto più la retta "spiega" bene la distribuzione (quindi anche il valore di R^2 è vicino a 1), tanto meno i valori di Y sono stati modificati dalla procedura di controllo e correzione.

3.2.1.2. Modifica delle distribuzioni monovariate

Per le sole variabili nominali ed ordinali, la valutazione del grado di modificazione della distribuzione di Y viene valutato attraverso gli *indici di dissomiglianza*:

$$I_{m1} = \frac{1}{2} \sum_{k=1}^K |f_{Y_k} - f_{Y_k^*}| \quad (2a)$$

$$I_{m2} = \left[\frac{1}{2} \sum_{k=1}^K |f_{Y_k} - f_{Y_k^*}|^2 \right]^{\frac{1}{2}} \quad (2b)$$

dove f_{Y_k} e $f_{Y_k^*}$ sono le frequenze marginali della variabile Y rispettivamente nel dataset dei dati originali ed in quello dei dati validati. Tali indicatori assumono un valore tra 0 (nessun impatto) e 1. Il secondo, rispetto al primo, è meno sensibile alle singole differenze tra le frequenze relative alle diverse modalità. Sempre nel caso delle sole variabili nominali ed ordinali, vengono prodotte le *matrici di transizione*, tabelle nella cui fiancata vengono riportate le modalità assunte dalla variabile Y nei dati originali, ed in testata le modalità assunte nei dati validati. Le frequenze nelle celle indicano le “transizioni”. Nella diagonale principale sono riportati i casi che non hanno subito modifiche. Leggendo la tabella per riga, è possibile comprendere come è stato modificato un valore originale; leggendo per colonna, si comprende da quali valori originali si ottiene un valore validato.

Per quanto riguarda le variabili continue, vengono posti a confronto indicatori sintetici univariati relativi alle distribuzioni “prima” e “dopo”: medie, scostamenti quadratici medi, mediane e quartili. Vengono inoltre calcolati gli *indici di distanza di Kolmogorov-Smirnov* tra le funzioni di ripartizione delle distribuzioni della variabile.

3.2.1.3. Modifica delle distribuzioni multivariate

Per le variabili nominali e ordinali, date due variabili X e Y, ai fini della valutazione del grado di modifica della relazione bivariata tra X e Y sono definiti i seguenti indicatori:

$$I_{j1} = \frac{1}{2} \sum_{Y, X} |f_{YX} - \tilde{f}_{YX}| \quad (3a)$$

$$I_{j2} = \left[\frac{1}{2} \sum_{Y, X} |f_{YX} - \tilde{f}_{YX}|^2 \right]^{\frac{1}{2}} \quad (3b)$$

che variano nell'intervallo tra 0 (nessun impatto) e 1, e possono essere estesi al caso di k variabili (con un massimo di k=5).

3.2.1.4. Modifica delle struttura associativa

Per le variabili nominali ed ordinali, per valutare il grado di modifica della struttura associativa tra le variabili (a due a due), viene utilizzato il *coefficiente di contingenza di Cramer*:

$$Cramer = \left[\frac{\chi^2}{n \times \min(r-1, c-1)} \right]^{\frac{1}{2}} \quad (4)$$

che assume i valori tra -1 (associazione negativa), 0 (indipendenza) e 1 (associazione positiva).

Per le variabili continue, vengono calcolati gli *indici di correlazione di Pearson* e le *covarianze* tra coppie di variabili nei dati grezzi ed in quelli puliti.

3.2.2 Indicatori per SIDI

IDEA permette anche di produrre un set di indicatori che possono essere esportati per il caricamento all'interno del Sistema Informativo di Documentazione delle Indagini, per essere visualizzati con SIDI-TOP.

Questi possono essere così raggruppati:

indicatori sull'ammontare dei dati sottoposti al processo di imputazione;

indicatori di valutazione dell'effetto della procedura di imputazione;
 indicatori di sintesi sul tasso di imputazione per variabile e per record.

Per quanto riguarda il primo gruppo, gli indicatori riguardano il totale dei record, il numero delle variabili nel complesso e quello delle variabili sottoposte al processo.

Nel secondo gruppo sono contenuti indicatori riguardanti:

il tasso di imputazione, con il dettaglio: indicatori di modificazione, di imputazione netta e di cancellazione;

il tasso di non imputazione, anche in questo caso col dettaglio: tasso di valori blank immutati e tasso di valori non blank immutati.

Infine, per quanto riguarda gli indicatori di sintesi, essi riguardano il complesso delle variabili:

- *primo quartile del tasso di imputazione delle variabili*: il valore del tasso di imputazione che lascia a sinistra il 25% delle variabili ordinate in senso crescente rispetto al tasso stesso;
- *terzo quartile del tasso di imputazione delle variabili*: il valore del tasso di imputazione che lascia a sinistra il 75% delle variabili ordinate in senso crescente rispetto al tasso stesso;

numero di variabili con tasso di imputazione superiore al 5%;

numero di variabili con tasso di imputazione superiore al 2%;

ed il complesso dei record:

- *primo quartile del tasso di imputazione per record*: il valore del tasso di imputazione che lascia a sinistra il 25% dei record ordinati in senso crescente rispetto al tasso stesso;
- *terzo quartile del tasso di imputazione per record*: il valore del tasso di imputazione che lascia a sinistra il 75% dei record ordinati in senso crescente rispetto al tasso stesso;

numero di record con tasso di imputazione superiore al 5%;

numero di record con tasso di imputazione superiore al 2%.

3.3. L'applicazione di IDEA ai dati di indagine ed analisi dei risultati

I dati dell'Indagine sui condannati dell'anno 2005 riguardano 61 diverse variabili, quasi tutte di tipo nominale, per un complesso di 221.381 record.

La prima attività effettuata è consistita nell'individuare quali di queste siano state effettivamente coinvolte nell'applicazione della procedura di controllo e correzione. Dal calcolo degli indici relativi alla modifica dei valori individuali, risulta che 13 di queste hanno subito un impatto (vedi tabella seguente).

Nome variabile	D1	D1i	D1m	D1c
Anno di nascita(ANNON)	0,016	0,016		
Mese di nascita (MESEN)	0,319	0,319		
Giorno di nascita (GIORNON)	0,384	0,325	0,059	
Luogo di nascita (LUOGON)	0,004	0,004		
Mese del delitto (MESEIR1)	4,449	4,448	0,001	
Giorno del delitto (GIORNOIR1)	10,626	10,565	0,062	
Provincia del delitto (PROVCR1)	7,324	7,324		
Comune del delitto (COMUNECR1)	7,324	7,187	0,138	
Natura del delitto (NATREAT1)	0,002	0,002		
Anno del provvedimento (DATAG)	0,065	0,001	0,064	
Ergastolo (ERG)	0,002			0,002
Giorni detenzione (RECLUG)	0,902	0,877	0,025	
Sesso (SESSO)	0,009		0,009	

E' da precisare che prima di applicare IDEA, i valori grezzi eguali a 0 della variabili sopraelencate sono stati posti a blank, per permettere il calcolo corretto degli indici D1i e D1m, sostituendo così gli "0"

presenti nei record per indicare la mancata risposta. Il tasso di imputazione globale D1 va da un minimo di 0,002% per la natura del delitto, ad un massimo del 10,626% per il giorno del delitto. In realtà, delle 13 variabili, solo per 4 di esse si può parlare di una effettiva attività di imputazione, con percentuali superiori al 4%: si tratta di mese e giorno del delitto, provincia e comune del delitto.

Questo non significa però che la qualità dei dati sia del tutto accettabile, e che per tale ragione l'impatto della procedura sia così limitato. Consideriamo ad esempio le variabili demografiche sesso, stato civile, istruzione, professione e cittadinanza: la distribuzione dei relativi valori mancanti nei dati grezzi, coincidente con quella nei dati puliti, è riportata nella seguente tabella:

Variabile	% valori mancanti
Sesso	0,00
Stato civile	87,92
Istruzione	94,57
Professione	94,65
Cittadinanza	77,43

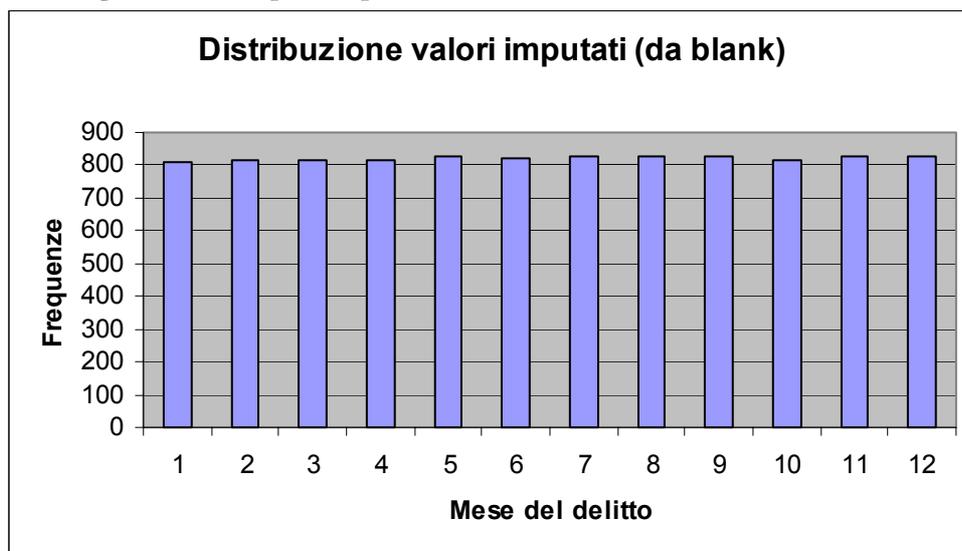
In sostanza, l'unica variabile "completa" è quella relativa al sesso: tutte le altre sono affette da un'alta o altissima percentuale di valori mancanti.

E' questo un caso frequente nei dati amministrativi: la completezza e correttezza sono normalmente assicurate relativamente alle informazioni che costituiscono l'oggetto delle procedure amministrative che utilizzano o producono i dati. Le informazioni utili solo indirettamente a tali procedure sono spesso carenti sia sotto il profilo della correttezza che sotto quello della completezza, come nel caso in esame.

In una situazione come quella evidenziata non si può d'altronde pretendere da una procedura di controllo e correzione di imputare un così alto numero di valori mancanti (tra il 77% ed il 94%).

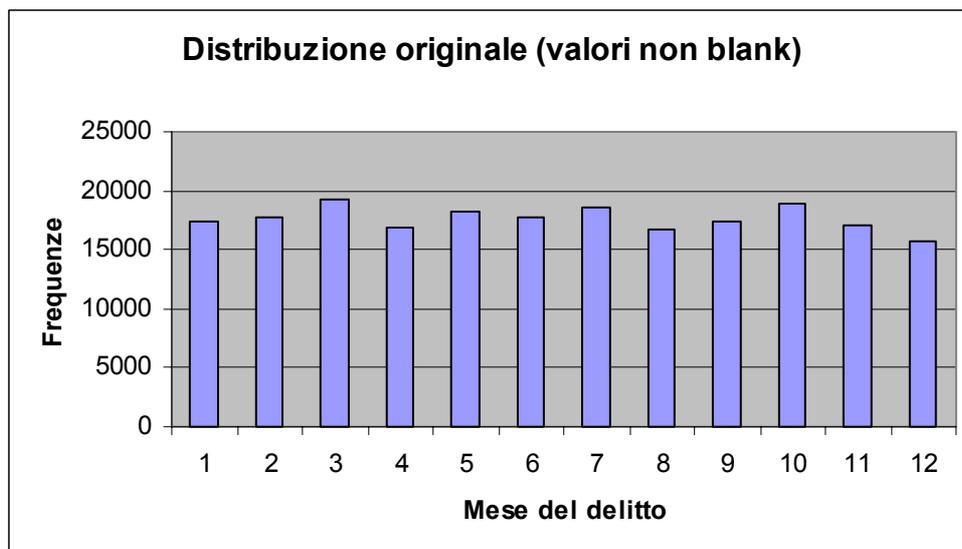
Imputare ha un senso quando la frazione di errori e/o valori mancanti è piccola, e l'informazione presente consente di recuperare quella mancante senza stravolgere la distribuzione di partenza.

Vediamo se questo è avvenuto nei casi in cui si è provveduto ad imputare, ad esempio nella variabile "mese del delitto", per una frazione dei casi pari a circa il 4,5%. Gli indicatori ci dicono che la quasi totalità delle imputazioni ha comportato l'attribuzione di un valore ammissibile al posto di un valore mancante. IDEA ci permette di ottenere la matrice di transizione relativa alla variabile, dalla quale è possibile immediatamente comprendere la distribuzione dei valori nei quali sono stati trasformati i blank, riportata nel seguente grafico:



L'attribuzione dei valori è stata effettuata mediante un programma che, a rotazione, ha attribuito i valori da 1 a 12: si tratta di una distribuzione di tipo sostanzialmente costante, sia pure con leggere imperfezioni.

Se però consideriamo la distribuzione originale presente nei dati grezzi, al netto dei valori mancanti, questa non è di tipo costante, in quanto caratterizzata da picchi a marzo e ottobre e da minimi in agosto e dicembre:



Possiamo pertanto dire che l'imputazione effettuata ha in qualche modo provocato una distorsione. Per capire l'entità dell'impatto facciamo riferimento agli indicatori Im1 e Im2, pari entrambi a 0,001: la distorsione è quindi del tutto trascurabile. Tuttavia un risultato migliore, in termini di qualità dell'imputazione, si sarebbe potuto ottenere utilizzando un metodo quale quello del donatore, che permette di replicare una distribuzione dei dati imputati molto simile a quella dei dati originali. Considerazioni del tutto analoghe si possono fare per la variabile "giorno del delitto".

Per quanto riguarda la provincia e il comune in cui sono stati effettuati i delitti (PROVCR1 e COMUNECR1), il tasso globale di imputazione è identico (7,324%). Mentre però per la provincia queste hanno riguardato totalmente l'imputazione di valori mancanti, nel caso del comune vi sono state anche modifica di valori, sia pure in numero limitato (0,138%).

Consideriamo ora una variabile continua, "giorni di detenzione" (RECLUG). La gran parte delle imputazioni riguarda i valori mancanti (imputati 1.941 volte, lo 0,9% del totale). Per analizzare l'impatto sulla distribuzione marginale, IDEA calcola media analitica e medie di posizione sia sul dataset originale che su quello validato. In questo caso, il confronto è inficiato dalla presenza di valori "777777" (codifica della pena dell'ergastolo). Pertanto si è deciso di porre a confronto i seguenti insiemi di dati:

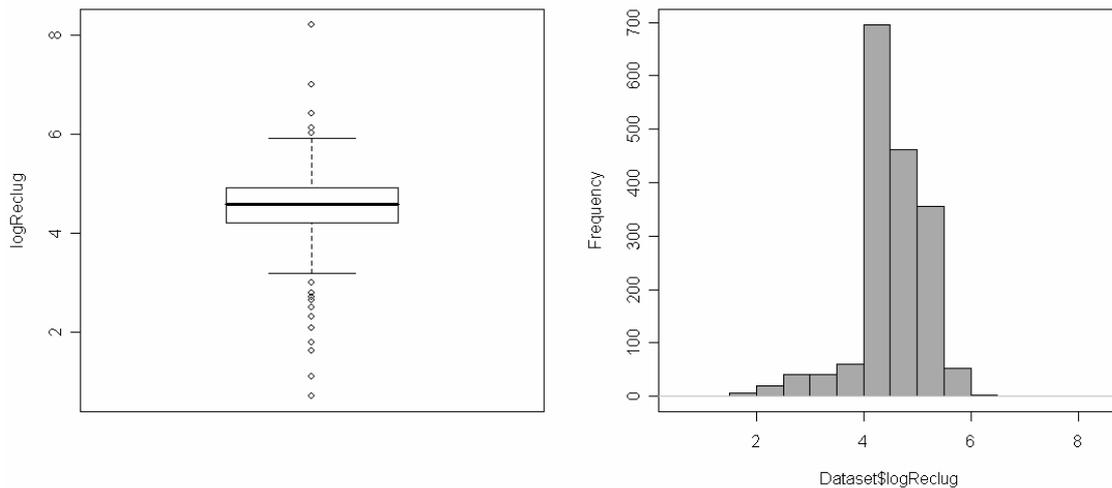
la distribuzione dei 1.941 valori imputati con esclusione dei condannati all'ergastolo (in totale 1.745 casi);

la distribuzione dei valori originali diversi da missing, sempre con l'esclusione dei condannati all'ergastolo (in totale 142.952 casi),

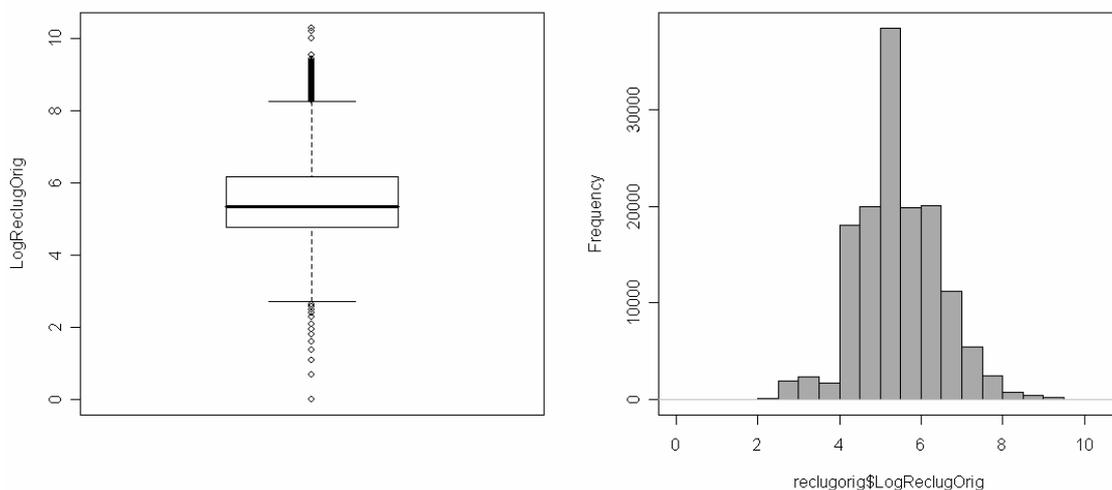
per verificare che i primi siano distribuiti rispettando la distribuzione dei secondi, ed evitando in tal modo l'introduzione di distorsioni nella distribuzione finale di questa variabile.

Per garantire una migliore rappresentazione, in entrambi i casi i valori sono stati trasformati in scala logaritmica, anche per verificare l'approssimazione alla normale.

Il box-plot e l'istogramma della prima distribuzione:



confrontati con il box-plot e l'istogramma della seconda distribuzione:



evidenziano delle differenze, dovute ad esempio ad una presenza marcata di valori nella coda destra della distribuzione originale, molto meno presenti in quella dei valori imputati, al contrario caratterizzati da una maggior presenza di valori nella coda sinistra. D'altra parte il rispetto della forma a campana per entrambe le distribuzioni, e la limitata numerosità dei valori imputati rispetto al totale, portano a concludere che l'imputazione non possa aver causato distorsioni significative nella distribuzione finale.

Infine, gli indicatori prodotti per SIDI permettono di avere un quadro d'insieme della situazione, in quanto questi vengono calcolati sul complesso delle variabili modificate dalla procedura.

Una prima considerazione è l'impatto è abbastanza contenuto, avendo riguardato il 2,6% dei valori delle variabili coinvolte. La seconda, immediata, è che la gran parte delle imputazioni è consistita nell'attribuire valori ammissibili a quelli mancanti, mentre la correzione vera e propria (da valore a valore) è molto limitata.

Per quanto riguarda la variabilità territoriale (per ripartizione) degli indicatori, il tasso medio di imputazione sembra essere uniforme, con la significativa eccezione delle Isole, con un picco di imputazione: un aspetto che andrebbe indagato per individuare eventuali fattori sistematici.

Ripartizione	Tasso imputazione	Tasso modificazione	Tasso imputazione netta	Tasso cancellazione
Nord Ovest	2,33	0,21	2,12	0,00
Nord Est	2,57	0,24	2,32	0,00
Centro	2,08	0,19	1,89	0,00
Sud	2,35	0,26	2,09	0,00
Isole	4,72	0,32	4,40	0,00
ITALIA	2,63	0,24	2,39	0,00

In conclusione, l'analisi dei risultati evidenzia un limitato impatto della procedura di controllo sui dati dell'indagine, derivante essenzialmente dall'elevata accuratezza con la quale la gran parte delle informazioni è prodotta dalle procedure amministrative di origine.

Nei limitati casi in cui la procedura di controllo e correzione interviene, l'analisi ha dimostrato che l'impatto sulle distribuzioni dei dati è trascurabile.

Gli indicatori di copertura e mancata risposta nella nuova indagine campionaria sui servizi idrici

Grazia Di Bella, Silvia Capuano, Marilena Angela Ciarallo, Antonino Laganà, Simona Ramberti
DPTS/AMB

1. Introduzione
 2. Il sistema delle indagini sulle acque 2005
 3. L'indagine preliminare: strategie per la minimizzazione degli errori di lista
 4. Valutazione della qualità dell'indagine preliminare
 5. L'indagine effettiva: alcune riflessioni sugli indicatori di copertura
 6. Conclusioni e ulteriori sviluppi
- Bibliografia

1. Introduzione

Il Sistema delle indagini sulle acque (SIA) predisposto dall'Istat ha l'obiettivo di fornire informazioni statistiche sull'uso delle risorse idriche a scopo potabile, sul trattamento delle acque reflue urbane e sulle principali caratteristiche dei servizi idrici presenti in Italia.

Il settore dei servizi idrici include diverse infrastrutture (acquedotti, reti di distribuzione dell'acqua potabile, reti fognarie e impianti di depurazione) preposte al servizio di prelievo, trasporto, distribuzione e depurazione dell'acqua. Le unità di rilevazione del SIA sono gli Enti gestori di tali infrastrutture.

La normativa che disciplina la gestione dei servizi idrici, nell'ultimo decennio, è stata oggetto di profondi cambiamenti, a partire dalla legge Galli emanata nel 1994¹, che hanno reso necessaria un'approfondita analisi del mutato contesto di riferimento.

Uno degli obiettivi di questa riforma è stato quello di rimediare alla elevata frammentazione della gestione, caratterizzata dalla presenza di una molteplicità di enti (soprattutto comuni) responsabili soltanto di una parte del servizio (ad esempio, enti che gestiscono soltanto reti di distribuzione, oppure soltanto reti fognarie).

La riorganizzazione prevista richiede che un unico soggetto debba assicurare l'integrazione della gestione di tutti i servizi idrici, su una dimensione territoriale sufficientemente ampia. A questo proposito la legge Galli ha individuato gli Ambiti territoriali ottimali (Ato), che rappresentano appunto le aggregazioni territoriali ottimali all'interno delle quali dovrebbe operare il gestore del Servizio idrico integrato (Sii) definito come *“insieme dei servizi pubblici di captazione, adduzione e distribuzione di acqua ad usi civili, di fognatura e di depurazione delle acque reflue”*.

Questi mutamenti istituzionali hanno determinato una notevole dinamicità del contesto in cui operano gli Enti gestori: alle molte gestioni prevalentemente in economia del passato, rivolte per lo più a pochi impianti, si stanno sostituendo gestioni di forme giuridiche diverse e di maggiori dimensioni; ossia pochi soggetti, responsabili di tutte le funzioni del servizio, dal prelievo alla depurazione.

Le trasformazioni dettate dalla legge Galli sono tuttora in corso; infatti, mentre l'individuazione degli Ato sul territorio è stata completata², l'affidamento della gestione ad un ente e l'operatività dello stesso presentano situazioni diverse nei vari Ato: in alcuni casi la procedura di affidamento del Sii è stata espletata e l'ente affidatario è operativo, in altri Ato la procedura è stata ultimata ma l'ente affidatario non è ancora operativo, in altri Ato l'affidamento non è stato ancora realizzato.

Attraverso l'attività di individuazione delle unità di rilevazione il SIA ha fotografato questa trasformazione in atto e proprio l'operazione di aggiornamento della lista, attraverso l'individuazione dei gestori non più esistenti e dei nuovi gestori Sii subentrati, si è rilevata non priva di difficoltà.

Nei paragrafi che seguono, dopo avere descritto il fenomeno osservato e i cambiamenti che lo hanno interessato, vengono analizzate le strategie adottate al fine di minimizzare gli errori di lista; sono inoltre descritte le caratteristiche principali del processo di rilevazione che hanno consentito la prevenzione

¹ Legge del 5 gennaio 1994, n. 36 “Disposizione in materia di risorse idriche”.

² La regione Veneto e la regione Friuli-Venezia Giulia hanno previsto la costituzione, tuttora in corso, di un Ato interregionale.

degli errori in fase di raccolta del dato. Infine sono presentati alcuni indicatori di qualità relativi alla copertura e alla mancata risposta.

2. Il sistema delle indagini sulle acque 2005

L'Istat rileva dati sulle acque urbane sin dal 1951, attraverso diverse indagini di tipo censuario volte a raccogliere informazioni su specifiche fasi del ciclo dell'acqua. Nella prima edizione del 1951 sono stati rilevati dati sugli acquedotti e sulle reti fognarie, nelle edizioni successive sono stati di volta in volta aggiunti moduli indirizzati alla rilevazione di dati su altre fasi del ciclo di uso dell'acqua, fino ad arrivare alla versione completa realizzata nel 1999 attraverso il Sistema delle indagini sulle acque (SIA).

Il SIA è un sistema di rilevazioni integrate, suddiviso in due fasi. Nella prima fase viene condotta un'*Indagine preliminare* presso gli Uffici tecnici dei comuni con l'obiettivo di aggiornare le liste di base necessarie per la rilevazione: la lista delle infrastrutture dei servizi idrici (acquedotti, reti di distribuzione, reti fognarie e impianti di depurazione) e la lista delle unità di rilevazione, ovvero gli Enti gestori delle infrastrutture stesse. Una volta costituite le liste, nella seconda fase viene condotta l'indagine cosiddetta *effettiva*, allo scopo di rilevare le informazioni afferenti le singole tipologie di impianto.

Con il SIA 99, realizzato a livello censuario, è stata costituita per la prima volta la lista degli Enti gestori operanti in Italia e dei relativi impianti che rappresenta un patrimonio informativo unico e completo; in particolare le informazioni raccolte per ciascun gestore riguardano:

- denominazione e indirizzo dell'Ente gestore dei servizi idrici;
- acquedotti gestiti (denominazione e comune di prelievo);
- reti di distribuzione dell'acqua potabile gestite (comune servito);
- reti fognarie gestite (comune servito);
- impianti di depurazione delle acque reflue urbane gestiti (denominazione e indirizzo).

Tali liste sono state la base per l'implementazione dell'edizione successiva del SIA, riferita al 2005, con l'obiettivo di aggiornare ed integrare le informazioni rilevate nell'edizione precedente.

A causa dell'evoluzione dell'assetto delle gestioni che ha riguardato il settore, come già detto, è stato necessario riorganizzare il piano di rilevazione ai fini del corretto aggiornamento delle liste. Oltre ai comuni e agli Enti gestori, unità di rilevazione rispettivamente dell'*Indagine preliminare* e dell'*Indagine effettiva*, la struttura del SIA ha dovuto considerare un altro soggetto rispondente: l'Autorità d'ambito territoriale ottimale (Aato). Tale Autorità, prevista presso ciascun Ato, tra gli altri compiti, deve provvedere alla organizzazione del Sii, disponendo l'affidamento della prestazione ad un unico Gestore. Allo scopo di monitorare ed aggiornare lo stato di attuazione della riforma dei servizi idrici³, le Aato insediate sono state le unità di riferimento di una ulteriore rilevazione, che ha preceduto l'*Indagine preliminare*, denominata *Ricognizione sullo stato di attuazione del Sii* (Prospetto 1).

In particolare tale *Ricognizione*, per ciascun Ato, ha conseguito i seguenti obiettivi:

- verifica o conferma dell'avvenuto affidamento del Sii;
- aggiornamento della lista degli Enti gestori a cui è stato affidato il Sii;
- acquisizione della lista dei comuni in cui i gestori Sii sono operativi e dei relativi servizi gestiti (verificando la presenza di situazioni di transizione);
- individuazione degli enti sovrambito e, laddove presenti, acquisizione della relativa denominazione e degli altri dati anagrafici.

Da questa *Ricognizione* è emerso che al 30 giugno 2005 in 54 Ato, dei 91 individuati sul territorio nazionale, è stato realizzato l'affidamento del Sii ad uno o più gestori. I gestori del Sii risultano 93. Per i restanti 37 Ato si è ancora in una fase intermedia di attuazione della riforma.

³ Cfr "*Indagine sui servizi idrici: ricognizione sullo stato di attuazione del Servizio idrico integrato al 30 giugno 2005*" - Istat - 4 novembre 2005 - Nota informativa.

Negli Ato in cui il Sii è stato affidato, l'operatività dei gestori è graduale: in alcuni comuni la gestione integrata è completa, ossia l'ente affidatario gestisce tutti i servizi del comune, in altri casi è parziale, vale a dire che l'ente ne gestisce soltanto una parte (ad esempio solo l'acquedotto, mentre la depurazione mantiene la gestione preesistente).

Complessivamente è risultato che i comuni con il Sii sono 3.541 (43,7 per cento del totale) e rappresentano il 58,4 per cento della popolazione totale residente.

La *Ricognizione* ha rivelato la presenza di due realtà: circa la metà dei comuni italiani è gestita da enti affidatari del Sii, solo 93 Enti gestiscono i servizi idrici su 3.541 comuni, i rimanenti 4.559 comuni mantengono una elevata frammentazione delle gestioni. Tale dicotomia ha richiesto una opportuna strategia operativa per la rilevazione dei dati. Si è previsto un piano di campionamento a due stadi. Al primo stadio, la lista di campionamento è costituita dall'insieme dei comuni: dai 4.559 comuni con una elevata polverizzazione delle gestioni sono state campionate 1.428 unità; i 3.541 comuni in cui era invece operativo il Sii sono stati censiti. Al secondo stadio sono stati considerati tutti gli Enti gestori operanti nei comuni campionati e i 93 gestori Sii individuati con la *Ricognizione*.

Se la *Ricognizione* ha permesso di aggiornare la lista dei gestori per la parte Sii, l'*Indagine preliminare* condotta sui 1.428 comuni del campione ha permesso di aggiornare la parte restante delle gestioni.

Prospetto 1 – Fasi del SIA 2005

Indagine	Unità di rilevazione	Numero di unità di rilevazione	Periodo di rilevazione	Metodo di rilevazione	Principale obiettivo
Ricognizione	Aato operative	88 Aato	Giugno 2005	CATI	Aggiornare la lista dei gestori del Sii e dei comuni gestiti
Preliminare	Uffici tecnici del comune	1.428 comuni	Gennaio 2006	CATI	Aggiornare la lista degli altri gestori e degli impianti gestiti
Effettiva	Enti gestori dei servizi idrici	93 gestori Sii + 1.686 altri gestori = 1.779 gestori	Maggio-Luglio 2006	CATI/CAPI	Rilevare i dati quantitativi sui flussi di acqua

3. L'indagine preliminare: strategie per la minimizzazione degli errori di lista

Il SIA 2005, rispetto alla precedente indagine, ha apportato numerose innovazioni strutturali e di processo che hanno riguardato essenzialmente la fase di rilevazione.

La novità principale riguarda la scelta della tecnica di rilevazione: per il SIA99 sono stati predisposti cinque questionari cartacei per ottenere notizie sulle caratteristiche degli Enti gestori e sugli impianti. Per il SIA 2005 sono state adottate tecniche di acquisizione dati assistite da computer CATI (*Computer Assisted Telephone Interviewing*) e CAPI (*Computer Assisted Personal Interviewing*); i questionari elettronici utilizzati nelle varie fasi del SIA 2005 sono stati sviluppati in Istat con *software* Blaise.

In particolare, l'utilizzazione di tale strumento di rilevazione si è rivelato utile nella minimizzazione degli errori di lista e nella massimizzazione dei tassi di risposta. Infatti esso ha consentito, grazie alla sua specificità, di ottenere molteplici vantaggi:

- l'acquisizione dei dati alfanumerici identificativi (denominazione del gestore, dell'acquedotto, dell'impianto di depurazione) è stata effettuata attraverso la procedura di *codifica assistita*, disponibile con il software Blaise, sulla base delle liste di archivio riducendo il rischio di duplicazione e di errore;
- l'utilizzo dei dati di archivio per la personalizzazione dei quesiti ha facilitato l'intervista diminuendo i tempi di compilazione;

- L'attivazione di domande filtro ha indirizzato automaticamente il percorso dell'intervista (modularità del questionario);
- il confronto con i dati di archivio ha consentito di inserire dei controlli *on line* nell'intervista.

Un pre-test realizzato presso il laboratorio CATI dell'Istat su un campione di comuni ha permesso di valutare la fluidità dell'intervista; la sequenza dei quesiti da sottoporre al rispondente è riportata nel diagramma di Figura 1.

Il percorso del questionario è sinteticamente articolato in tre passi:

1. presenza dei servizi idrici nel comune;
2. elenco dei gestori operanti nel comune e relative notizie anagrafiche;
3. elenco degli impianti associati ad ogni gestore dichiarato.

Nella prima parte del questionario si è chiesto se nei comuni sono presenti i servizi di acquedotto, di distribuzione, di fognatura e di depurazione. Tali quesiti hanno agito da filtro permettendo di modulare il percorso del questionario con lo scopo di eliminare qualsiasi rischio di inesattezza o non accuratezza della rilevazione. Le informazioni dell'archivio SIA99, relative ai servizi presenti in tutti i comuni d'Italia, sono state utilizzate come strumento di controllo. Se un comune intervistato ha fornito una risposta difforme dal dato presente in archivio (ad esempio ha dichiarato che nel suo comune non è presente il servizio di acquedotto e tale servizio risultava presente nel 1999) viene attivata una domanda di approfondimento volta ad accertare che la situazione sia veramente mutata rispetto al 1999 (Figura 2).

Figura 1 - Diagramma del questionario per l'indagine preliminare sui servizi idrici

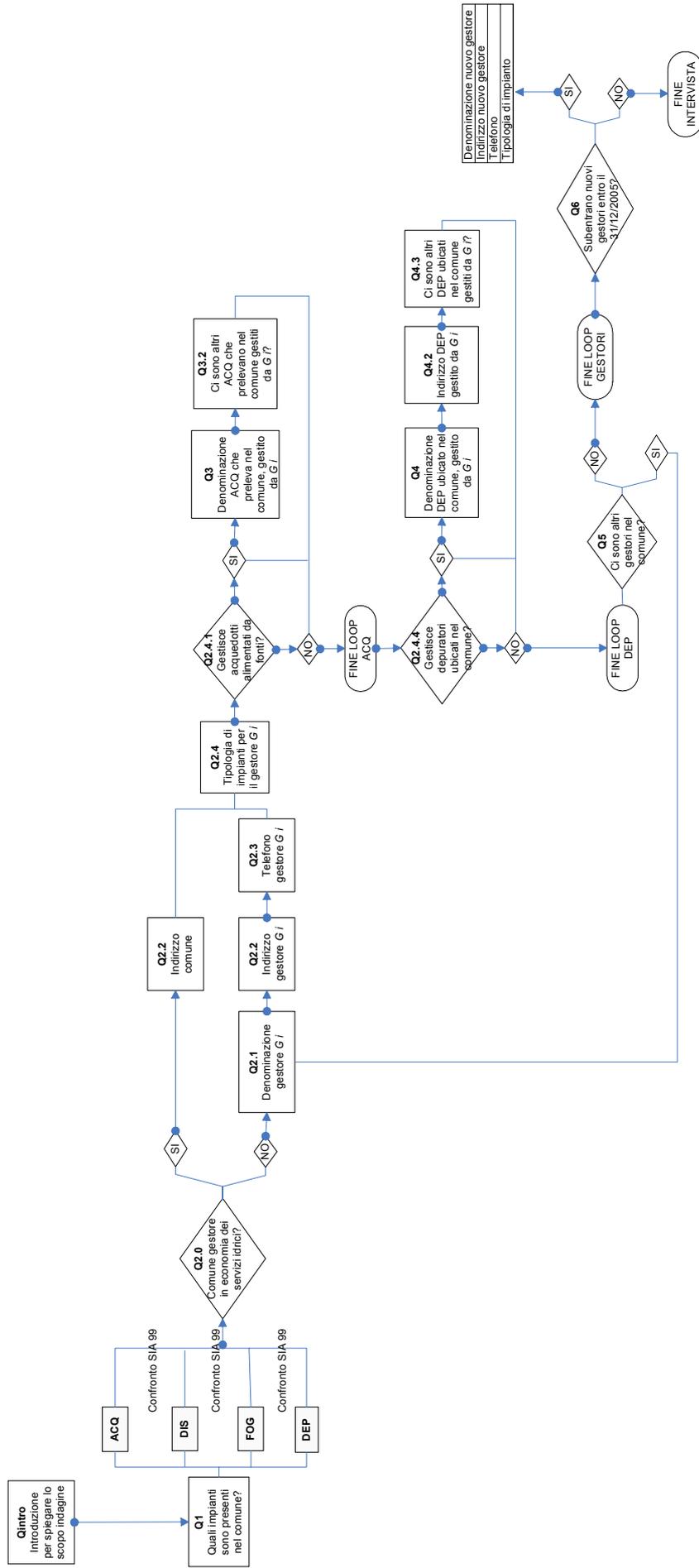


Figura 2 - Domanda di approfondimento

Q1.2.A Ma nel 1999 c'erano fonti che alimentavano acquedotti?

1. Sì, prima c'erano e adesso non più

2. No, non c'erano neanche nel 1999

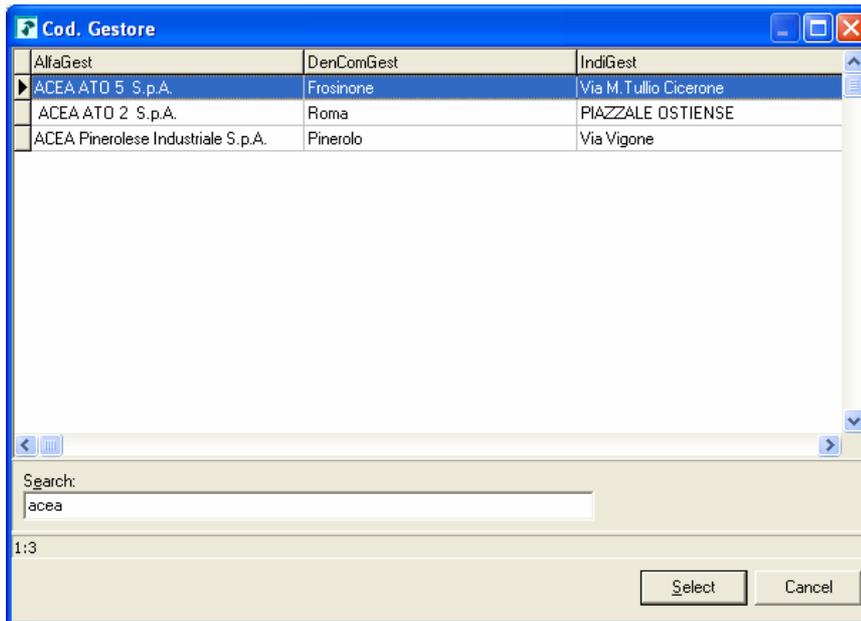
Se il rispondente ha dichiarato che il servizio non era presente neanche nel 1999 viene attivato un ulteriore strumento di controllo, il *soft message*, che ha indotto il rispondente a riflettere sulla bontà del dato di archivio e sulla correttezza del dato appena fornito e, se lo ha ritenuto necessario, anche di modificarlo (Figura 3).

Figura 3 – Soft message collegato alla domanda di approfondimento

Questions involved	Value
Non c'erano fonti nel '99	@No, non c'erano neanche nel 1...
Fonti	No

Domande di approfondimento sono state progettate per tutte e quattro le tipologie di servizio. La parte centrale del questionario, da cui dipende l'aggiornamento della lista degli Enti gestori, si è aperta con una domanda filtro: "Il comune è gestore in economia, cioè gestisce direttamente tutti i servizi idrici?" In base alla risposta, il percorso è stato articolato in quesiti diversi: in caso affermativo è stato chiesto l'indirizzo, altrimenti la denominazione del gestore che opera nel comune. Utilizzando la lista degli Enti gestori aggiornata con la *Ricognizione*, le operazioni di acquisizione della denominazione del gestore sono state effettuate tramite la funzione di *codifica assistita*. Il rilevatore ha scritto la denominazione nell'apposito campo e, indipendentemente dalla modalità di scrittura (ad esempio ACEA, acea, A.C.E.A.), il *software* ha confrontato la stringa digitata con la lista in archivio attraverso l'uguaglianza delle permutazioni dei trigrammi (gruppi di tre lettere) e si è avviata la procedura di "*look up*" (Figura 4). Questa procedura ha permesso di visualizzare la lista delle possibili unità codificate che possono apparirsi alla stringa.

Figura 4 – Procedura di codifica assistita



Se la denominazione fornita è troppo generica, e in elenco appaiono più alternative riconducibili alla stessa, facendo scorrere la barra orizzontale l'intervistatore ha potuto consultare alcune informazioni aggiuntive quali ad esempio il comune in cui è ubicata la sede principale dell'ente e il suo indirizzo, come illustrato nella schermata di *look up* riportata in Figura 4. Questo ha consentito di valutare, insieme al rispondente, quale stringa selezionare per garantire una corretta identificazione.

Il sistema di codifica assistita ha permesso di ottenere una riduzione degli errori di digitazione poiché una volta individuato il nominativo della lista il *software* ha registrato soltanto il codice Istat ad esso associato.

Questa procedura è risultata utile nel caso di gestori che operano in diversi comuni intervistati: nel caso in cui questi abbiano fornito la denominazione del gestore in modo differente, il sistema di codifica ha evitato la creazione di duplicati.

Il passo successivo ha riguardato l'acquisizione dell'indirizzo: se il gestore dichiarato dal rispondente è presente nell'archivio, il quesito relativo all'indirizzo viene personalizzato con le seguenti informazioni :

- il comune della sede principale;
- il nome della via;
- il civico;
- il CAP.

La personalizzazione dell'indirizzo dell'Ente gestore sulla base dei dati derivanti dal SIA99 ha permesso di ridurre ulteriormente l'errore di lista e massimizzare la copertura dell'indagine, nonché di diminuire i tempi di compilazione dell'intervista: infatti, se i dati sono stati confermati dal rispondente, è stato immediatamente sottoposto il quesito successivo, altrimenti è stato possibile un loro aggiornamento in tempi rapidi perché il rilevatore ha potuto posizionarsi sui singoli campi interessati dalla modifica. Se il gestore non era presente in archivio la denominazione è stata inserita *ex novo* e tutti i dati relativi alle notizie anagrafiche sono stati richiesti con domande apposite ed inseriti manualmente.

L'acquisizione del dato relativo al numero di telefono ha richiesto particolare attenzione poiché esso è stato successivamente utilizzato per contattare gli Enti gestori nell'*Indagine effettiva* CATI. Per tale motivo si è posta particolare attenzione ai quesiti interessati analizzando le possibili implicazioni. Innanzitutto si è evitato, qualora i comuni fossero essi stessi gestori, di sottoporre al rispondente il quesito relativo al numero di telefono grazie all'attivazione di un filtro automatico. Per tutti gli altri gestori si è scelto di richiedere il dato non tramite conferma, come per l'indirizzo, ma tramite domanda aperta, per sottolineare la necessità di una collaborazione attiva con il rispondente e per aumentare l'esattezza del dato. Come ulteriore elemento di verifica si è previsto un confronto automatico tra il

dato fornito in corso d'intervista e quello presente in SIA99: se i due numeri di telefono differivano si attivava una domanda di controllo allo scopo di indurre l'intervistato a riflettere su quale dei due fosse corretto e confermarne uno dei due.

Come già detto, oltre alla lista degli Enti gestori, l'*Indagine preliminare* ha aggiornato la lista degli impianti da essi gestiti. Dopo aver registrato le informazioni anagrafiche di un gestore, il questionario elettronico è stato suddiviso in quattro moduli, uno per ogni tipologia di impianto. All'intervistato sono state sottoposte quattro domande filtro per sapere quali tipologie di impianto sono gestite da ciascun ente operante sul suo territorio.

In caso di risposta affermativa per il modulo degli acquedotti e dei depuratori, è stato necessario procedere all'identificazione degli impianti gestiti tramite apposite domande. Per facilitare il lavoro dell'intervistatore, i quesiti relativi a queste due tipologie di impianti sono stati personalizzati con i dati in archivio: la domanda che l'intervistatore ha sottoposto al rispondente conteneva già le informazioni relative alla denominazione degli acquedotti, alla denominazione e all'indirizzo dei depuratori.

Nel caso in cui i dati contenuti nel quesito personalizzato non erano corretti, si è proceduto all'acquisizione del dato corretto attraverso la procedura di codifica assistita e, quando necessario, con l'acquisizione del nuovo dato.

Ulteriore strumento di minimizzazione degli errori sono stati i controlli *on line* inseriti all'interno del questionario elettronico. Il principale controllo volto a evitare errori di cancellazione di lista ha riguardato la verifica tra i gestori dichiarati e quelli presenti in archivio: se il rispondente non ha dichiarato un ente gestore presente nella lista di archivio associato al comune è stato attivato un messaggio di errore *soft* con lo scopo di accertare se il gestore non opera più nel comune o di sanare una eventuale dimenticanza del rispondente (Figura 5).

Figura 5 – Controllo di fine intervista

The screenshot shows a window titled "Active Signal" with a close button in the top right corner. The main text area contains the following message:

SG2 - Dai dati in nostro possesso risulta operare nel suo comune anche il gestore COMUNE DI SAN COSMO ALBANESE, che non mi ha elencato, conferma che questo ente non gestisce impianti sul suo comune?
se l'intervistato conferma cliccare su 'Suppress', altrimenti cliccare su 'Goto' per inserire un altro ente gestore.

Below the text is a table with two columns: "Questions involved" and "Value".

Questions involved	Value
Altri gestori	No

At the bottom of the window are three buttons: "Suppress", "Close", and "Goto".

4. Valutazione della qualità dell'indagine preliminare

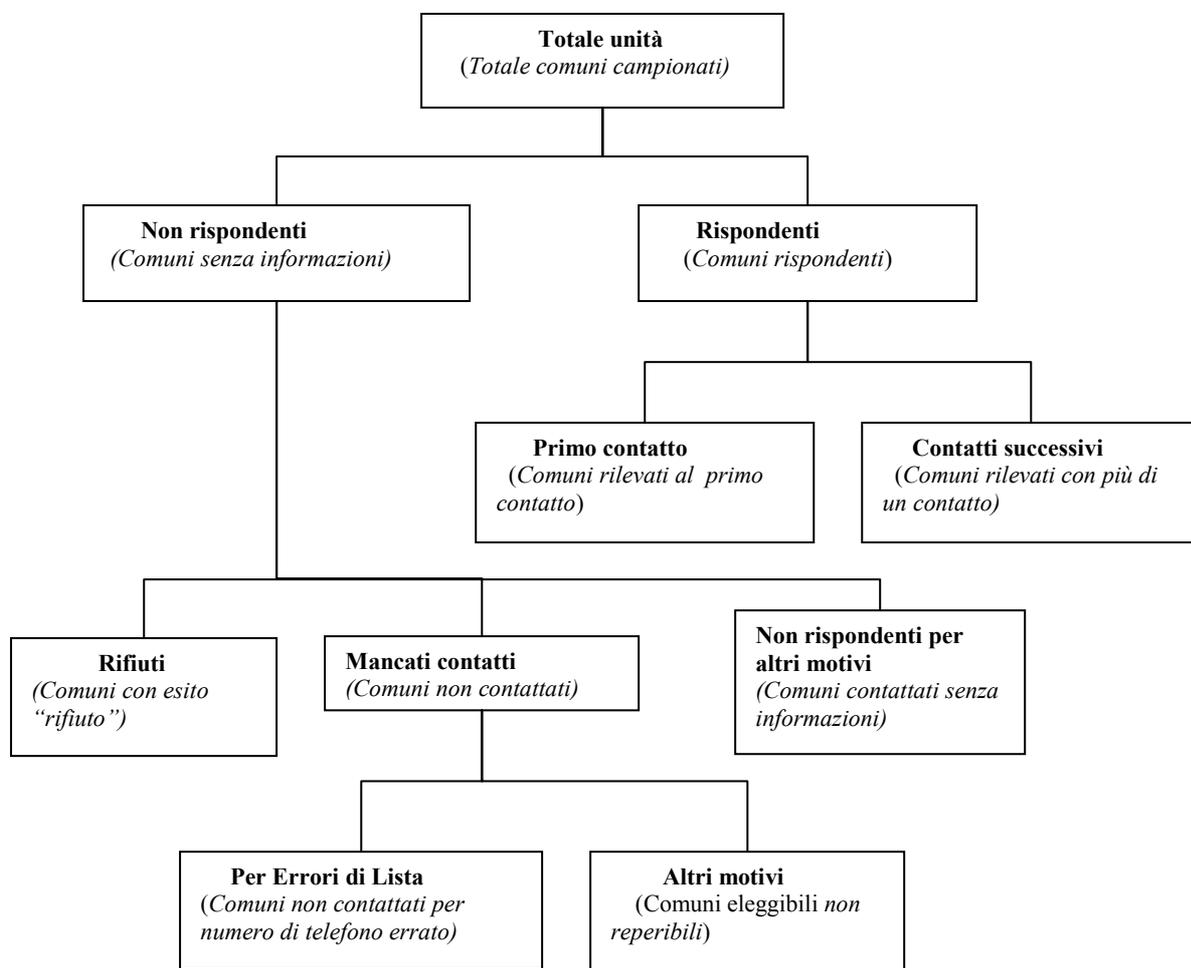
L'*Indagine preliminare* è stata effettuata in modalità CATI da una società di servizi di *contact center* selezionata dall'Istat tramite gara.

La rilevazione telefonica presso i responsabili degli uffici tecnici comunali si è svolta dal 13 Gennaio al 2 Febbraio 2006 per un totale di sedici giorni lavorativi; le telefonate sono state eseguite dal lunedì al venerdì dalle ore 9:00 alle 13:00, orario di maggiore reperibilità degli addetti comunali, salvo nei casi di appuntamenti pomeridiani stabiliti con gli interessati. Il monitoraggio costante da parte dell'Istat delle attività della società incaricata ha garantito un'elevata qualità della rilevazione.

Nella Figura 6 è riportato lo schema delle unità di rilevazione, ovvero dei comuni campionati, sulla base del quale sono stati calcolati gli indicatori di copertura e gli indicatori sulle mancate risposte presenti nel sistema Sidi (Sistema informativo di documentazione delle indagini) dell'Istat (Tavole 1 e 2).

L'obiettivo dell'*Indagine preliminare* è stato quello di aggiornare la lista degli Enti gestori dei servizi idrici e, di conseguenza, l'eleggibilità del comune è definita in base alla presenza di almeno una delle infrastrutture del servizio idrico (acquedotto, rete di distribuzione, rete fognaria, impianto di depurazione). Tutte le unità contattate sono risultate eleggibili e si può ipotizzare che, in base a informazioni *a priori*, anche i comuni non contattati siano eleggibili (non esistevano nel 1999 comuni sprovvisti di tutte e quattro le tipologie di servizio idrico e si presuppone non ne esistano nel 2005); ciò equivale a considerare tutte le unità come risolte.

Figura 6 - Classificazione delle unità per esito della rilevazione nell'Indagine preliminare sui gestori dei servizi idrici



Le unità eleggibili non contattate per errore di lista, ovvero numero di telefono errato, sono risultate solo tre (0,2 per cento), sia grazie alla buona qualità della lista fornitaci da Ancitel, la società di servizi telematici dell'Associazione nazionale dei comuni italiani (Anci), ma soprattutto perché si è adottata una procedura di recupero dei numeri telefonici in corso di indagine. Ad una settimana dall'inizio della rilevazione, e poi ad intervalli di tempo più brevi, si è previsto di estrarre l'elenco dei comuni che fino a quel momento avevano un esito negativo di contatto con modalità "numero errato" (libero, fax, segreteria, cellulare staccato). Sono state effettuate apposite ricerche dal personale Istat che, dopo aver recuperato i corretti numeri di telefono, ha provveduto ad inserirli nuovamente all'interno del *software* Blaise.

Il totale dei comuni campionati è pari a 1.428 unità che, come già detto, sono state selezionate dalla lista di campionamento costituita dai comuni in cui la gestione dei servizi idrici non è stata ancora affidata ad un gestore unico secondo le indicazioni della legge Galli.

Tavola 1- Indicatori di copertura per l'Indagine preliminare sugli Enti Gestori dei servizi idrici
(tassi espressi in percentuale)

Totale unità	1428
Tasso di unità risolte	100,0
Tasso di errori di lista	0,2
Tasso di unità non eleggibili	0
Tasso di unità non più esistenti	-
Tasso di unità con variazioni di stato	-
Tasso di unità fuori target	0
Tasso di unità eleggibili non contattate per errore di lista	0,2

Le interviste telefoniche concluse sono state 1.361 su un totale di 1.428, quindi il tasso di risposta è pari al 95,3 per cento (Tavola 2).

Tra le 67 unità non rispondenti si registrano 2 rifiuti, 62 Mancati contatti e 3 unità contattate ma non in grado di fornire informazioni. La maggior parte dei Mancati contatti (59) è dovuta al superamento della soglia prefissata per il numero di tentativi di chiamata; tale soglia è uno dei parametri disponibili all'interno dello schedulatore delle telefonate del software Blaise.

Tavola 2 - Indicatori di mancata risposta per l'Indagine preliminare sugli Enti Gestori dei servizi idrici

Totale Unità	1428
Tasso di unità risolte	100,0
Tasso di risposta	95,3
Tasso di mancata risposta totale	4,7
Tasso di rifiuto	0,1
Tasso di mancato contatto	4,3
Tasso di mancato contatto per errori di lista riferito alle unità eleggibili	0,2
Tasso di mancato contatto per altri motivi	4,1
Tasso di mancata risposta per altri motivi	0,2
Tasso di mancata risposta totale al netto degli errori di lista che pregiudicano il contatto	4,4
Tasso di mancata risposta totale riferito alle unità eleggibili	4,7

La fluidità del questionario elettronico, la buona preparazione professionale del personale del *call center*, nonché l'assidua presenza del personale Istat durante tutto il periodo della rilevazione presso la società, hanno consentito di ottenere ottimi livelli di qualità e ridurre al minimo il fastidio statistico per il rispondente.

Nella Tavola 3 sono presentati gli indicatori per ripartizione geografica. Si nota un seppur lieve andamento decrescente nei tassi di risposta procedendo da nord a sud.

La percentuale di rispondenti al primo contatto è pari al 22,4 per cento. Questo dato riveste particolare importanza perché rivela il grado di disponibilità delle informazioni richieste, considerando l'elevato contenuto tecnico delle variabili da rilevare.

Tavola 3 - Tasso di risposta dell'Indagine preliminare sugli Enti Gestori dei servizi idrici per ripartizione geografica

RIPARTIZIONI	Rispondenti	%	Non rispondenti	%	Rispondenti primo contatto	%	Rispondenti successivi	%	Totale Unità
GEOGRAFICHE									
Nord-Ovest	317	97,5	8	2,5	92	29,0	225	71,0	325
Nord-Est	371	95,9	16	4,1	97	26,2	274	73,8	387
Centro	100	97,1	3	3,0	22	22,0	103	78,0	103
Sud	253	93,7	17	6,3	40	15,8	213	84,2	270
Isole	320	93,3	23	6,7	54	16,9	266	83,1	343
ITALIA	1361	95,3	67	4,7	305	22,4	1056	77,6	1428

5. L'indagine effettiva: alcune riflessioni sugli indicatori di copertura

L'*Indagine effettiva* del SIA2005 ha avuto come unità di rilevazione i responsabili tecnici degli Enti gestori dei servizi idrici con lo scopo di raccogliere le informazioni finali della rilevazione.

La complessità delle informazioni richieste ha reso necessaria la sperimentazione di un questionario elettronico sviluppato per moduli, realizzato per la prima volta in Istat attraverso un'applicazione che utilizza più *data model* Blaise collegati tra di loro con il linguaggio Visual Basic. Ciò ha permesso di passare da un modulo all'altro dell'intervista senza dover seguire un prefissato percorso, facilitando così il lavoro del rispondente e garantendo una maggiore flessibilità ed elasticità.

Il questionario è stato somministrato attraverso la tecnica CATI per i gestori in possesso di un numero relativamente ridotto di informazioni ed in tecnica CAPI per gli Enti gestori di un elevato numero di impianti (Tavola 4). La mancata risposta di un ente gestore di numerosi impianti e, conseguentemente di una vasta porzione di territorio, ovviamente costituisce una significativa perdita di informazioni.

Allo scopo di minimizzare il rischio di mancata risposta per queste unità, a cui si richiede oltretutto una maggiore disponibilità di tempo per l'intervista, è stata realizzata una raccolta preliminare di dati precedente all'intervista CAPI. Con l'ausilio degli Uffici regionali dell'Istat sono stati inviati agli Enti gestori quattro questionari (uno per ogni tipologia di impianto: acquedotti, reti di distribuzione, reti fognarie e impianti di depurazione) personalizzati con le informazioni provenienti dall'archivio, predisposti in formato Excel e strutturati in modo tale da consentire l'autocompilazione guidata da parte dei rispondenti. Sulle incompatibilità emerse dai dati forniti, si è proceduto con la somministrazione delle domande di controllo (*hard e soft message*) direttamente presso gli Enti gestori, utilizzando la tecnica CAPI.

La rilevazione presso il *call center* in modalità CATI si è svolta nei giorni feriali dal 22 Maggio 2006 al 23 Giugno 2006, dalle 9:00 alle 13:00 e dalle 15:30 alle 17:30. Rispetto all'*Indagine preliminare* le telefonate sono state effettuate anche nel pomeriggio per garantire una maggiore reperibilità dei soggetti rispondenti, rappresentati anche da società o consorzi.

I dati relativi ai gestori intervistati in modalità CAPI sono stati rilevati nel periodo Maggio-Luglio 2006.

Tavola 4 –Unità di rilevazione per tecnica di indagine e per ripartizione geografica (a)

RIPARTIZIONI GEOGRAFICHE	CATI	CAPI	Totale
Nord-ovest	616	21	637
Nord-est	448	16	464
Centro	94	19	113
Sud	281	12	293
Isole	268	3	271
Italia	1707	71	1778

(a) Esiste un gestore con sede all'estero e il totale dei gestori, classificati per ubicazione della sede centrale, è pertanto pari a 1.779.

Di seguito si presenta l'analisi degli indicatori del sistema SIDI riferiti all'*Indagine effettiva* SIA 2005. La Figura 7 mostra la classificazione delle unità di rilevazione e la Tavola 5 presenta gli indicatori di mancata risposta.

Grazie alla elevata qualità della lista delle unità di rilevazione, aggiornata - come già detto - dalla *Ricognizione* e dall'*Indagine preliminare*, il tasso di eleggibilità delle unità contattate è pari al 100,0 per cento. L'eleggibilità delle unità di rilevazione può essere connessa al concetto di operatività dell'ente gestore. Nei casi in cui la gestione è stata affidata in tempi recenti e l'ente non è ancora in grado di fornire le informazioni richieste, ancora in possesso dei precedenti gestori, esso potrebbe essere considerato non eleggibile. Nel corso della rilevazione si è proceduto adottando una strategia pragmatica per governare questa fase transitoria. Tale strategia è consistita nel mantenere l'eleggibilità del gestore pur richiedendo i dati ai gestori precedenti, nell'attesa di verificare l'avvenuta transizione e aggiornare la lista.

Anche gli Enti gestori non contattati sono stati considerati totalmente "Risolti" ed eleggibili in base alle informazioni rilevate nell'*Indagine preliminare* o nella *Ricognizione*.

Figura 7 - Classificazione delle unità per esito della rilevazione nell'indagine sui gestori dei servizi idrici 2005

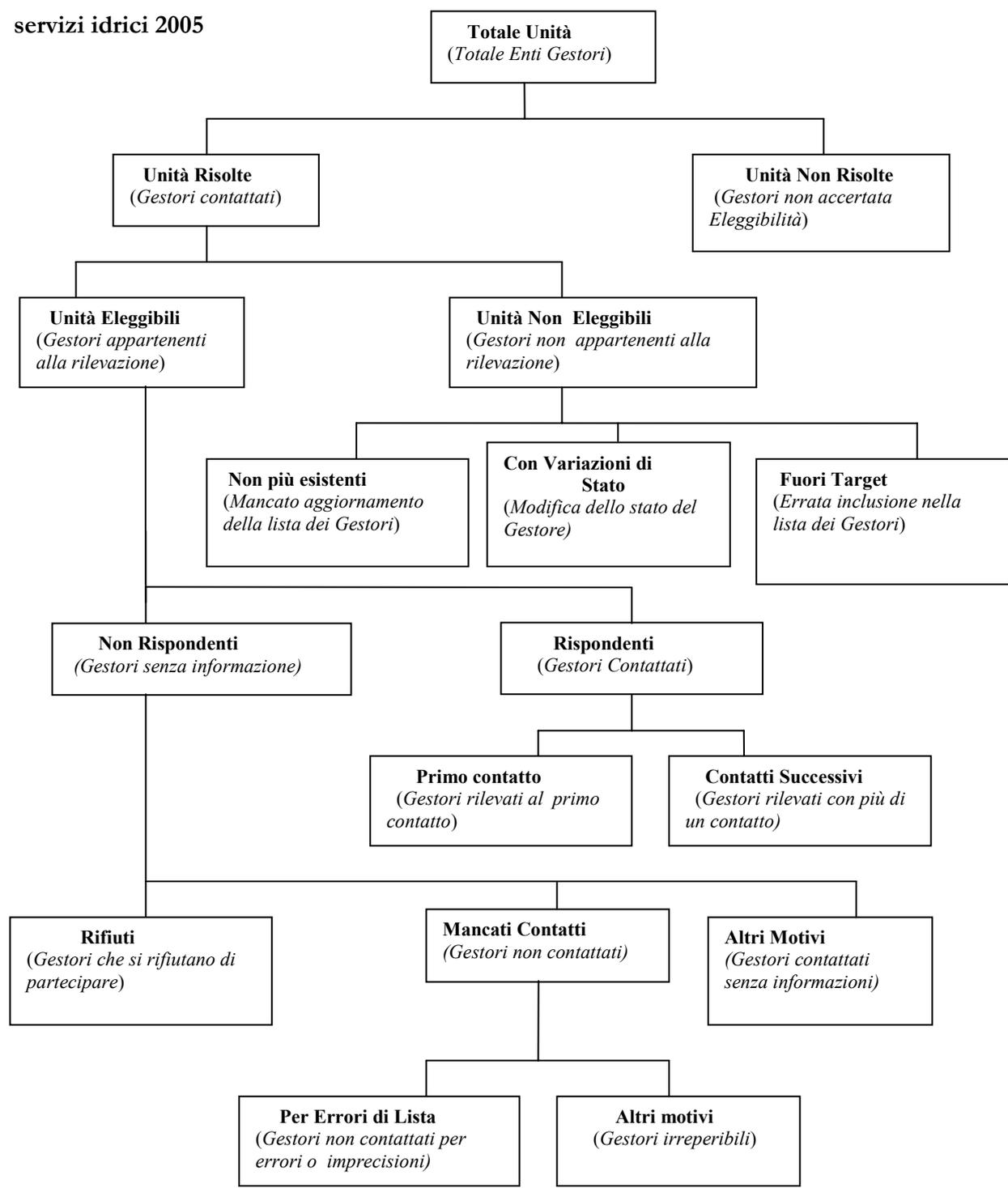


Tavola 5 - Indicatori di mancata risposta. Indagine effettiva

Indicatori di mancata risposta

Totale Unità	1779
Tasso di Unità Risolte	100,0
Tasso di Risposta	93,3
Tasso di Mancata risposta totale	6,7
Tasso di Rifiuto	4,5
Tasso di Mancato contatto	2,2
Tasso di Mancato Contatto per Errori di Lista (riferito alle Unità Eleggibili)	1,9
Tasso di Mancato Contatto per Altri Motivi	0,3
Tasso di Mancata risposta per Altri Motivi	0,0
Tasso di Mancata risposta Totale al netto degli Errori di lista che pregiudicano il contatto	4,8
Tasso di Mancata Risposta Totale riferito alle Unità Eleggibili	6,7

Gli indicatori sono stati calcolati anche per area geografica e tecnica di rilevazione CAPI/CATI (Tavola 6).

Gli Enti gestori dei servizi idrici rilevati sono 1.779; di questi 1.708 (il 96,0 per cento del totale) sono stati intervistati con tecnica CATI e i restanti 71 (4,0 per cento) con tecnica CAPI (Tavola 4).

Se la strategia utilizzata per l'aggiornamento della lista delle unità ha permesso di ridurre al minimo gli errori di lista, il tasso di risposta finale, pari al 93,3 per cento, avvalorava le scelte adottate nel corso dell'intera rilevazione.

L'articolazione territoriale del tasso di risposta mostra che il Centro è la ripartizione geografica con la percentuale maggiore di unità rispondenti (96,5 per cento); di contro il Nord-est registra il valore più basso dell'indicatore (89,9 per cento) (Tavola 6).

Tavola 6 – Tasso di risposta per tecnica di indagine e ripartizione geografica

RIPARTIZIONI GEOGRAFICHE	CATI	CAPI	Totale
Nord-ovest	94,0	95,2	94,0
Nord-est	89,5	100,0	89,9
Centro	96,8	94,7	96,5
Sud	95,0	91,7	94,9
Isole	94,4	100,0	94,5
ITALIA	93,2	95,8	93,3

L'analisi del tasso di mancata risposta totale, pari al 6,7 per cento, fa emergere come causa principale della non risposta il rifiuto a rispondere, con 80 unità pari al 67,2 per cento del totale delle unità non rispondenti, seguito dalla modalità mancato contatto 39 unità pari al 32,8 per cento (Tavola 7).

Per il sottoinsieme dei gestori intervistati con tecnica CAPI, le tre mancate risposte derivano dal rifiuto a rispondere.

Tavola 7 – Unità non rispondenti per motivo della non risposta e ripartizione geografica

RIPARTIZIONI GEOGRAFICHE	Rifiuto	Mancato contatto	Altri motivi	Totale
Nord-ovest	25	13	0	38
Nord-est	33	14	0	47
Centro	3	1	0	4
Sud	13	2	0	15
Isole	6	9	0	15
ITALIA	80	39	0	119

Le unità non contattate per errore di lista, numero di telefono errato, sono risultate 33 e sono presenti esclusivamente nel sottoinsieme di unità CATI. Il tasso di mancato contatto per errore di lista è dunque pari all'1,9 per cento.

Come per l'indagine preliminare, anche nel corso dell'*Indagine effettiva* gli eventuali numeri di telefono errati sono stati ricercati ed aggiornati nel corso della rilevazione.

6. Conclusioni e ulteriori sviluppi

La strategia di rilevazione volta alla minimizzazione degli errori di lista e alla massimizzazione del tasso di risposta ha portato degli ottimi risultati. La sequenza delle rilevazioni (*Ricognizione sullo stato di attuazione del Sii e Indagine Preliminare*), volte all'aggiornamento delle liste che compongono il SIA, e la successiva *Indagine Effettiva* presentano un bilancio costi/benefici che avvalorano le scelte fatte. La progressiva concentrazione delle gestioni nell'ambito di pochi soggetti, in seguito all'applicazione della legge Galli, rafforza la necessità di adottare una strategia di tipo misto e flessibile che accompagni la transitorietà del processo gestionale.

Ulteriori elementi di riflessione da sviluppare per il futuro riguardano la definizione di indicatori di qualità riferiti non solo agli Enti gestori, quanto piuttosto alle unità di primo stadio, cioè i comuni attraverso opportune ponderazioni dei valori. Laddove la gestione comprenda un elevato numero di comuni, ad esempio un Ato, la mancata risposta del gestore costituisce una notevole perdita di informazioni.

Bibliografia

Istat - *Indagine sui servizi idrici: ricognizione sullo stato di attuazione del servizio idrico integrato (30 giugno 2005)*. Roma: Istat, (Nota informativa 4 Novembre 2005). <http://www.istat.it/ambiente/ambterr>

Istat - *Il sistema delle indagini sulle acque (Anno 2005)*. Roma: Istat, (Statistiche in breve 29 Novembre 2006). <http://www.istat.it/ambiente/ambterr>

Istat - *Il sistema delle indagini sulle acque Anno 1999*. Roma: Istat, 2006 (Informazioni, n. 16).

Pagine web per la consultazione dei dati del Sistema delle indagini sulle acque 1999 <http://acqua.istat.it>

La nuova Indagine sulle unità locali delle grandi imprese: contributi al miglioramento della qualità dell'informazione statistica sulla struttura delle imprese

Luigi Costanzo (Dcar/B), Barbara Gentili (Dcar/Reg/A), Serena Migliardo (Dcar/Reg/B)¹

1. Motivazioni e obiettivi della nuova indagine

L'Indagine sulle unità locali delle grandi imprese (IULGI), tenutasi per la prima volta nel 2004, è un'indagine totale e diretta a frequenza annuale, svolta per via postale e finalizzata all'implementazione – nell'Archivio statistico delle imprese attive (ASIA) – di un nuovo livello informativo sulle unità locali, destinato a produrre dati sull'insediamento territoriale delle imprese. L'indagine è nata per sopperire alla carenza di dati amministrativi sulle unità locali d'impresa, in particolare per quanto riguarda la distribuzione territoriale degli addetti. Secondo il Censimento del 2001, le imprese unilocalizzate (per le quali si può assumere l'identità tra unità locale e sede d'impresa) rappresentavano il 95,2% del totale, ma coprivano soltanto il 66,4% degli addetti: la copertura della quota rimanente, necessaria per una completa territorializzazione dei dati di ASIA, è stata affidata alla nuova indagine, il cui flusso informativo confluisce dunque nel processo di produzione dell'Archivio statistico.

IULGI rileva tutti i principali caratteri di struttura e di localizzazione rilevati dal Censimento dell'industria e dei servizi ma, grazie al campo di osservazione ridotto e ad alcune innovazioni di processo, è in grado di restituire dati territoriali sulla struttura delle imprese in tempi molto più brevi di quelli richiesti dal Censimento. Uno studio sperimentale sulle imprese manifatturiere, basato sul confronto fra i risultati dell'Indagine 2005 e le due ultime tornate censuarie (il Censimento intermedio del 1996 e quello generale del 2001) è stato pubblicato nel Rapporto annuale Istat del 2005, uscito a maggio 2006 e quindi a meno di un anno dall'epoca di riferimento dei dati (giugno 2005).

Questa comunicazione intende mettere in evidenza le ripercussioni positive che l'istituzione di IULGI e le innovazioni di processo introdotte nella filiera di produzione dei dati sulla struttura delle imprese hanno avuto su tre aspetti fondamentali della qualità dell'informazione statistica sulle imprese: *completezza*, *accuratezza* e *tempestività*. Vantaggi significativi in termini di accuratezza possono essere associati, come di consueto, all'utilizzazione di un questionario *on-line* con sistema di registrazione controllata e alla creazione di un sistema di monitoraggio della rilevazione in tempo reale, ma anche alla flessibilità del modello organizzativo della rete di rilevazione e alla scelta di "avvicinare" la gestione dell'indagine ai rispondenti, valorizzando il ruolo degli Uffici regionali. IULGI ha conseguito, fin dalla prima tornata, tassi di risposta insolitamente elevati e i suoi risultati – immessi nel processo produttivo di ASIA – hanno un ritorno positivo sulla qualità dell'archivio statistico. I vantaggi in termini di tempestività e completezza sono ancora più evidenti: la realizzazione di ASIA-Unità locali, resa possibile dalla nuova indagine, non solo consente di misurare a breve termine (ma con la stessa ricchezza di dettaglio territoriale del Censimento) le ripercussioni dei mutamenti economici sulla struttura dell'apparato produttivo, ma anche di arricchire e completare l'offerta informativa sull'universo delle imprese, andando incontro alla necessità, molto avvertita dagli utenti della statistica ufficiale, di disporre di dati territoriali di dettaglio sulle dinamiche delle imprese e dell'occupazione.

¹ A Luigi Costanzo sono da attribuire i capitoli 1 e 3; a Barbara Gentili i paragrafi 4.2 e 4.3; a Serena Migliardo i capitoli 2 e 5 e il paragrafo 4.1.

L'aggiornamento annuale dei dati sulle unità locali delle imprese, finora disponibili solo ogni dieci anni, in occasione del Censimento generale dell'industria e dei servizi, risponde a due esigenze fondamentali:

- a) *Cogliere tempestivamente le ripercussioni locali dei cambiamenti della struttura economica del Paese.* Statistiche locali aggiornate e di buona qualità rappresentano, infatti, la base informativa necessaria non solo per la pianificazione territoriale ma anche per i sempre più numerosi processi di programmazione e di valutazione di livello locale attivati, ad esempio, dalle politiche comunitarie.
- b) *Conformarsi ai regolamenti europei².* In materia di Business Register, il Regolamento Europeo n. 2186/1993, chiede agli stati membri dell'Unione di sviluppare, nei registri di imprese, anche il livello delle unità locali.

IULGI viene dunque a colmare la lacuna informativa che impediva la realizzazione di un registro statistico delle unità locali, e in questo senso contribuisce in misura decisiva al conseguimento di entrambi gli obiettivi. L'avvio della nuova indagine, tuttavia, era condizione necessaria ma non sufficiente per la costruzione e l'aggiornamento di ASIA-Unità locali: per questo nuovo prodotto è stato disegnato, pertanto, un processo che integra *input* di fonti diverse:

- a) *La nuova indagine annuale*, totale e diretta, rivolta alle unità locali delle imprese plurilocalizzate di grandi dimensioni (IULGI);
- b) *L'applicazione di modelli statistici*, che utilizzano fonti amministrative per la stima dello stato di attività e per la scelta ottimale dei caratteri delle unità locali;
- c) *L'utilizzazione del sistema informativo di ASIA-Imprese* per la produzione dell'archivio-base di IULGI e per l'integrazione delle stime finali.

Poiché IULGI nasce con l'obiettivo di colmare una carenza di dati amministrativi sulla localizzazione delle attività produttive, per servire a questo scopo doveva prefiggersi di garantire tassi di risposta molto elevati, adeguati a un'indagine chiamata a costituire una fonte di *input* di un archivio statistico. In vista di questo obiettivo, nella progettazione dell'indagine sono state seguite due linee strategiche:

- a) *Ridurre il più possibile il carico statistico per i rispondenti*, in modo da assicurarsi il massimo di collaborazione;
- b) *Avvicinare la gestione dell'indagine ai rispondenti*, coinvolgendo gli Uffici regionali dell'Istituto in tutte le fasi dell'indagine e valorizzando la loro conoscenza dei contesti locali e la loro capacità di contatto diretto con le imprese sul territorio.

La tabella seguente mostra come l'obiettivo di una copertura molto elevata dell'universo d'indagine (con particolare riferimento alle imprese di maggiori dimensioni) sia stato sostanzialmente raggiunto fin dalla prima tornata (2004).

² Andrebbe, seppur brevemente, rilevato che la mancanza di un registro delle unità locali riduce la qualità di molte indagini sulle imprese che hanno come unità di analisi le KAU o le Local KAU. Queste criticità inficiano la stessa qualità dei conti nazionali.

Tab. 1 – Indagine IULGI: Questionari inviati e raccolti per ripartizione geografica e per classe di addetti dell'impresa. Anno 2004 (Valori assoluti e percentuali)

RIPARTIZIONI CLASSI DI ADDETTI	QUESTIONARI		TASSI DI RISPOSTA (raccolti/inviati)
	<i>Inviati</i>	<i>Raccolti</i>	
Nord-ovest	1.451	1.374	94,7
Nord-est	1.746	1.514	86,7
Centro	1.148	982	85,5
Mezzogiorno	2.217	2.087	94,1
Imprese con 100 addetti e più	4.889	4.417	90,3
Imprese con meno di 100 addetti	5.000	4.211	84,2
Totale	9.889	8.628	87,2

2. La struttura dell'indagine

Il campo di osservazione dell'indagine, giunta nel 2006 alla sua terza tornata, include tutte le imprese con più di 99 addetti e le imprese plurilocalizzate con almeno 50 addetti che svolgono la loro attività economica principale nei settori dell'industria e dei servizi. IULGI rileva tutte le principali variabili di struttura (addetti e attività economica) e di localizzazione rilevate dal Censimento dell'industria e dei servizi e riferibili alle unità locali delle imprese. Il questionario è unico a livello d'impresa e parzialmente precompilato grazie alle informazioni ricavate dall'archivio ASIA. I rispondenti devono quindi limitarsi a confermare o aggiornare i dati prestampati, e indicare – per ciascuna unità locale – il dato degli addetti per il mese di giugno dell'anno a cui fa riferimento l'indagine.

Nella progettazione del questionario si è optato per la massima semplicità e brevità al fine di facilitare il compito dei rispondenti, ben conoscendo i problemi di carico statistico delle grandi imprese. Il questionario si compone di cinque sezioni, le prime tre rivolte alla verifica delle variabili strutturali d'impresa: addetti (sia in media annua sia come dato puntuale al 30 giugno dell'anno di riferimento), attività economica prevalente e localizzazione. IULGI rileva anche il personale esterno all'impresa, al fine di analizzare l'incidenza delle nuove forme di lavoro atipico. Le ultime due sezioni sono dedicate alle unità locali, distinte fra quelle già rilevate in precedenza (sez. 4) e quelle nuove o comunque non presenti nell'elenco prodotto da ASIA (sez. 5). I dati della sezione 4 sono prestampati, grazie alle informazioni rilevate nelle precedenti tornate e confluite nell'Archivio: i rispondenti devono quindi limitarsi, anche in questo caso, a confermarli o correggerli.

Le principali variabili di struttura rilevate sono l'attività economica svolta nella singola unità locale e il numero di addetti impegnati nell'unità nel mese di giugno dell'anno oggetto di indagine. Viene rilevata, inoltre, la tipologia dell'unità locale: viene chiesto, cioè, all'impresa di specificare se nelle singole unità locali si svolga un'attività *operativa*, *amministrativa/legale* o *ausiliaria* (cioè di supporto all'attività principale dell'impresa), oppure se si tratti di unità non presidiate (magazzini, depositi e simili).

3. Le principali innovazioni

Un primo elemento di novità introdotto da IULGI (almeno rispetto all'impianto degli ultimi due Censimenti dell'industria e dei servizi, di cui la nuova indagine può ritenersi una filiazione) risiede nell'*organizzazione del flusso informativo*, dalla rilevazione alla registrazione dei dati elementari. L'acquisizione delle informazioni avviene attraverso quattro canali, cui corrispondono *quattro linee di lavorazione parallele*. Questa differenziazione del processo produttivo ha lo scopo di agevolare i rispondenti, offrendo loro la possibilità di scegliere fra diverse modalità di collaborazione. In questo senso, se pure non si può parlare di una vera e propria riduzione del carico statistico, questa misura

rappresenta un tentativo di “venire incontro” al rispondente, che è libero di scegliere la modalità meno onerosa in rapporto alle caratteristiche della propria impresa.

- 1) La prima linea di lavorazione prevede l’invio del questionario cartaceo per posta prioritaria a tutte le imprese che contano fino a 36 unità locali e la successiva restituzione (sempre a mezzo posta) ai centri di rilevazione, che provvedono alla revisione del questionario e alla sua registrazione attraverso un sistema di *data entry* controllato.
- 2) La seconda linea, alternativa alla prima, prevede la possibilità da parte delle stesse imprese (fino a 36 unità locali) di compilare e restituire il questionario via Internet. Questa modalità è stata scelta dal 23% delle imprese rispondenti, con il 17% delle unità locali. Nella prima rilevazione (2004), aveva optato per la compilazione *on-line* il 18,5% delle imprese.
- 3) La terza linea, riservata alle imprese con più di 36 unità locali, prevede la possibilità di fornire le informazioni richieste non attraverso la compilazione di un questionario ma in modo semplificato, attraverso la compilazione e l’invio via *e-mail* di una tabella su foglio elettronico Excel. La tabella predisposta per la compilazione viene recapitata alle imprese per posta, su CD-Rom.
- 4) La quarta linea, riservata alle poche imprese con un grandissimo numero di unità locali (come Poste italiane, Enel, Trenitalia) prevede contatti diretti con i referenti per concordare l’acquisizione diretta di archivi proprietari.

L’adeguamento della modalità di acquisizione delle informazioni alle caratteristiche e alle esigenze dei rispondenti è stato, a nostro parere, un fattore decisivo per il conseguimento di alti tassi di risposta e ha consentito, inoltre, di stabilire e consolidare proficui rapporti di collaborazione con i referenti delle grandi imprese.

Un altro elemento d’innovazione sul quale si è investito per aumentare il livello di qualità dei dati raccolti, inteso sia in termini di accuratezza sia in termini di tempestività, è *il coinvolgimento degli Uffici regionali dell’Istat (e dell’Astat di Bolzano) in tutte le fasi dell’indagine*, a partire dalla progettazione (dal disegno del questionario alla definizione delle modalità di acquisizione e controllo delle informazioni, alla creazione e gestione dei rapporti con i referenti delle imprese sul territorio). Il ruolo svolto dagli Uffici regionali è risultato strategico per diversi motivi, tutti riconducibili alla loro capacità operativa sul territorio e alla conoscenza diretta delle realtà locali: la prossimità ai rispondenti permette un maggiore coinvolgimento delle imprese indagate (ad es., attraverso l’intervista diretta alle imprese più restie a rispondere), mentre le imprese sembrano – dal canto loro – più propense a rispondere ai revisori locali rispetto ai revisori centrali.

La comunicazione fra i centri di revisione locali e la Direzione centrale degli archivi amministrativi e registri statistici (DCAR), cui fa capo l’indagine, è agevolata da diversi strumenti, con l’obiettivo di garantire una rapida risoluzione dei diversi problemi che si possono presentare nel corso della rilevazione. Particolarmente in quest’ambito, molte scelte organizzative derivano direttamente dall’esperienza maturata negli ultimi due Censimenti dell’industria e dei servizi (1996 e 2001). Si è puntato, in particolare, ad omogeneizzare le informazioni fornite dal centro agli Uffici regionali attraverso:

- a) *Un indirizzo di posta elettronica unico*, al quale gli operatori inviano i quesiti e dal quale ricevono le risposte;
- b) *La costituzione, a livello centrale, di referenti per gruppi di regioni*, in grado di assicurare risposte tempestive (entro tre giorni) sugli aspetti metodologici, organizzativi e informatici dell’indagine;
- c) *L’organizzazione di un flusso di frequently asked questions (FAQ)*: ogni 15 giorni, i quesiti posti dai centri regionali vengono risolti, standardizzati in casi di studio e inoltrati per conoscenza a tutti i revisori;

- d) *L'organizzazione di giornate di formazione nazionali*, integrate – ove se ne presenti la necessità – da interventi formativi dedicati a singole sedi regionali.

In generale, il continuo scambio di informazioni tra le strutture periferiche e la Direzione centrale consente di intervenire tempestivamente su eventuali problemi organizzativi dell'indagine, ed è servito anche a perfezionare l'impianto organizzativo delle rilevazioni successive.

A riprova della maggiore efficacia dimostrata dalle strutture periferiche nella rilevazione, la tabella che segue mostra un significativo divario fra i tassi di risposta conseguiti dai revisori degli Uffici regionali e quelli conseguiti dai revisori della Direzione centrale:

Tab. 2 – Indagine IULGI: Tassi di risposta conseguiti dai revisori degli Uffici regionali (per ripartizione geografica) e dai revisori della DCAR, per le imprese con 100 addetti e oltre. Anno 2005 (Valori percentuali)

	TASSO DI RISPOSTA (questionari raccolti/ questionari inviati)
Uffici regionali + Astat	91,6
Uffici regionali del Nord-ovest	84,2
Uffici regionali del Nord-est + Astat	96,2
Uffici regionali del Centro	90,6
Uffici regionali del Sud	93,8
Uffici regionali delle Isole	96,7
DCAR	84,8
Totale	88,3

Un terzo ambito d'innovazione, anch'esso in continuità con l'esperienza del Censimento, è costituito dal *sistema di monitoraggio della rilevazione*. Per IULGI è stato attivato un sistema telematico, accessibile dal sito <http://iulgi.istat.it/UL>, comprendente sei funzioni:

- a) *Questionari respinti*: la registrazione delle informazioni relative ai questionari respinti, curata dalla DCAR e dall'Astat di Bolzano, permette di aggiornare il sistema con le informazioni relative all'esito della spedizione dei questionari (cartacei o su CD-Rom) restituiti al mittente per impossibilità di consegna da parte del servizio postale. I questionari respinti sono classificati secondo la motivazione della mancata consegna, annotata sulla busta secondo la prassi delle Poste. La gestione dei questionari respinti consente di rispedire in tempi brevi i questionari che presentavano indirizzi o nominativi errati o non aggiornati. I dati aggiornati vengono inseriti contestualmente nell'archivio ASIA.
- b) *Arrivo Questionari*: questa funzione segnala l'acquisizione del questionario. La tempestività dell'aggiornamento, attraverso l'inserimento giornaliero da parte dei centri di rilevazione dei codici d'impresa di tutti i questionari restituiti, in forma cartacea o elettronica, è fondamentale ai fini dell'avvio delle azioni di sollecito, previste per tutte le imprese per le quali non risultino codici di arrivo.
- c) *Altri contatti*: questa funzione aggiorna il sistema con l'informazione relativa al tipo di contatto avviato tra l'impresa e il centro di rilevazione. Sono state previste sei modalità di contatto, che permettono di testare la disponibilità dell'impresa alla compilazione del questionario. Il tentativo non riuscito di contatto telefonico è distinto da quello che ha avuto esito positivo; il caso in cui l'impresa risulti irreperibile dopo diversi tentativi di contatto è distinto dal rifiuto esplicito di compilazione del questionario. Sono state previste, inoltre, una codifica per le imprese che contattano l'Istat spontaneamente per chiedere chiarimenti sulla compilazione e una per le imprese per le quali si ritiene necessaria un'intervista diretta o una missione.
- d) *Lavorazione respinti*: questa funzione permette di aggiornare il sistema con l'informazione relativa all'esito della lavorazione dei questionari respinti non risolti attraverso la correzione/aggiornamento dell'indirizzo o della ragione sociale. Per tali questionari viene

avviata una procedura di accertamento, svolta dalla DCAR, intesa a verificare se l'impresa interessata appartenga effettivamente al campo di osservazione dell'indagine oppure se sia coinvolta in eventi quali scorporo, fusione, cessazione dall'attività, inattività temporanea, liquidazione.

- e) *Report ed estrazioni*: tutte le informazioni relative alle diverse operazioni di rilevazione (Questionari respinti, Arrivo questionari, Altri contatti, Lavorazione respinti), sono disponibili, sotto forma di tavole riepilogative, sia per i revisori degli Uffici regionali sia per quelli della DCAR, per ciascun revisore limitatamente alle imprese di propria competenza. Per i coordinatori che operano a livello centrale sono disponibili quattro super-utenze che permettono la visualizzazione di tutte le imprese, strutturate per centro di rilevazione e per singolo revisore.
- f) *Download questionario*: questa funzione permette di scaricare in formato pdf il questionario personalizzato per ciascuna impresa. Ogni revisore può quindi rinviare il questionario, via *e-mail* o via fax, all'impresa contattata in fase di sollecito, che dichiara di non averlo mai ricevuto o di averlo smarrito.

La funzionalità del monitoraggio è basata sulla registrazione tempestiva di tutte le informazioni che caratterizzano la fase di rilevazione dell'indagine: dalla registrazione dei codici di arrivo dei questionari restituiti per posta, via fax o compilati tramite intervista telefonica, alla gestione dei solleciti (attraverso la registrazione dei codici di contatto telefonico intercorso con l'impresa). Nel complesso, il sistema consente una valutazione in tempo reale dell'andamento delle operazioni di rilevazione, con la conseguente possibilità di intervenire tempestivamente nelle situazioni critiche, a diversi livelli. Al livello dei singoli centri di rilevazione, l'inserimento dei codici di pertinenza consente di organizzare nella maniera più efficiente il carico di lavoro, gestendo in modo ottimale i solleciti e monitorando in tempo reale la situazione attraverso l'analisi dei *report* e l'estrazione delle liste di imprese per tipologia di lavorazione. Al livello centrale, la registrazione tempestiva dei codici effettuata dai centri di rilevazione periferici consente di controllare l'andamento complessivo dell'indagine su tutto il territorio nazionale nelle sue diverse fasi, e di individuare le situazioni critiche attraverso l'analisi comparata dei riepiloghi dei centri di rilevazione, per disporre opportuni interventi.

4. L'integrazione dei dati rilevati da IULGI 2005 con l'archivio ASIA

4.1. L'indagine come fonte di input dell'archivio

Come per il Censimento dell'industria e dei servizi del 2001, IULGI può essere definita una rilevazione "assistita da archivio", in quanto utilizza come base un elenco completo (salvo l'errore di aggiornamento) delle sue unità di rilevazione prodotto dall'Archivio ASIA. ASIA fornisce informazioni identificative (denominazione e indirizzo) e di struttura (attività economica, addetti dipendenti e indipendenti, forma giuridica, data di inizio e fine attività, fatturato) sulle unità registrate (imprese). A sua volta, IULGI costituisce una delle fonti fondamentali per l'aggiornamento di ASIA, e i dati rilevati dall'indagine vengono periodicamente confrontati con quelli dell'Archivio per l'aggiornamento delle variabili identificative e di struttura.

Con l'edizione 2006 è stata introdotta in IULGI la "Maschera supporto ASIA", uno strumento di raccolta di informazioni aggiuntive – derivabili da quelle rilevate tramite il questionario di IULGI – necessarie per l'aggiornamento dell'archivio ASIA. Tale maschera deve essere compilata, in corso d'opera, dai revisori dell'indagine. Tali informazioni aggiuntive si rendono necessarie in presenza di determinati eventi che si verificano nel ciclo di vita di un'impresa e che possono comportare cambiamenti sia nelle variabili di struttura che in quelle di identificazione: cessioni/acquisizioni di rami d'azienda, fusioni, scorpori, incorporazioni, cessazioni dall'attività per liquidazione volontaria o a causa di procedure concorsuali. In presenza di tali eventi si rende necessario acquisire informazioni dettagliate sulle altre imprese eventualmente coinvolte (soprattutto quando all'evento si associa una forte

variazione del numero di dipendenti) o sulle eventuali procedure di accesso alla cassa integrazione/licenziamenti, nei casi di crisi aziendale. Queste informazioni possono essere acquisite in modo diretto, tramite un ritorno sulle imprese coinvolte, oppure in modo indiretto (soluzione preferibile, nell'ottica della riduzione del fastidio statistico), tramite la consultazione di fonti esterne a disposizione dell'Istat quali l'archivio telematico delle Camere di Commercio ("Telemaco"), l'archivio dell'Inps³, nonché i siti Internet delle imprese interessate, che possono contenere, a volte, informazioni utili e soprattutto aggiornate. L'acquisizione dei dati in archivio è preceduta, in ogni caso, da controlli rivolti a testare la qualità dei dati rilevati.

Di seguito si riportano i risultati dei controlli effettuati sui dati rilevati con l'indagine 2005, prima che essi venissero utilizzati per l'aggiornamento di ASIA-imprese al 2004. I controlli hanno riguardato circa il 99% delle imprese rispondenti. Sulla base delle informazioni contenute nei questionari pervenuti sono stati definiti tre sottoinsiemi di imprese:

- a) *Lista 1* comprende le imprese per le quali la lavorazione si è conclusa positivamente: tali imprese risultano attive, non hanno subito cambiamenti di codice fiscale e non hanno segnalato procedure concorsuali o altre variazioni di assetto.
- b) *Lista 2* comprende le imprese con cambiamento di codice fiscale, segnalazione di eventi strutturali (fusioni, conferimenti, scissioni) e segnalazioni di stato diverso da attivo (cessazione o inattività). Se la data dell'evento è posteriore al 2004 (e quindi le informazioni sono utili per l'aggiornamento di ASIA 2004), le imprese sono state "forzate" e trattate come quelle della Lista 1.
- c) *Lista 3* comprende le imprese risultate fuori campo di osservazione per l'indagine IULGI 2005.

Tab. 3 – Indagine IULGI: Classificazione delle imprese rispondenti sottoposte a controlli per l'aggiornamento di ASIA 2004. Anno 2005 (Valori assoluti e percentuali)

	IMPRESE	%
Imprese in campo di osservazione, attive (lista 1)	10.891	89,1
Imprese in campo di osservazione, interessate da eventi (lista 2)	1.211	9,9
<i>Cambiamento di codice fiscale</i>	57	0,5
<i>Imprese "forzate" (evento posteriore al 2004)</i>	317	2,6
<i>Imprese aggiornate dai revisori di ASIA</i>	837	6,8
Imprese fuori campo di osservazione (lista 3)	16	0,1
Totale imprese sottoposte a controlli	12.118	99,2
Totale imprese rispondenti*	12.220	100,0

* Rappresentano l'85,3% delle imprese appartenenti al campo di osservazione dell'Indagine.

L'analisi è proseguita per le imprese della Lista 1, e in particolare per le variabili "attività economica" e "dipendenti medi al 2004". Per quanto riguarda le altre variabili, dopo una prima analisi si è deciso di accettare sempre la ragione sociale, la forma giuridica e la sede amministrativa fornita dall'indagine, in quanto ritenute più aggiornate di quelle presenti in archivio.

4.2. Aggiornamento della variabile "attività economica"

Nella compilazione del questionario, l'impresa era chiamata a fornire notizie sull'attività economica principale, cioè quella attività svolta in maniera prevalente dall'impresa. Il problema maggiore è stato riscontrato quando più attività sono esercitate nell'ambito di una stessa impresa; in tal caso l'impresa

³ Consultando la sezione "Aziende", in particolare, si possono ottenere informazioni sulla situazione occupazionale dell'impresa, desunte dai modelli DM10. I dati sono mensili, il che consente di tenere sotto osservazione i periodi in cui si rilevano variazioni anomale.

avrebbe dovuto individuare la prevalente sulla base del valore aggiunto o, in mancanza di esso, sulla base del fatturato o del numero medio annuo di addetti. Questo criterio non è stato in realtà sempre applicato; spesso l'impresa ha preferito non segnalare la variazione (dovuta ad inesattezza o effettiva variazione di attività economica) mantenendo l'attività prestampata sul questionario anche se differente da quella realmente svolta.

Per l'aggiornamento della variabile "attività economica" sono state definite le seguenti regole:

- R1) Tra le imprese con meno di 100 dipendenti medi, sono state sottoposte a controllo quelle che, nel confronto fra dati d'archivio e dati d'indagine, presentavano valori diversi solo nelle prime 4 cifre del codice Ateco 2002.
- R2) Tra le imprese con 100 dipendenti e oltre, il controllo è stato effettuato in tutti i casi di discordanza fra il dato d'archivio e quello d'indagine.

I risultati di tale controllo sono illustrati nella tabella seguente:

Tab. 4 – Indagine IULGI: Frequenze dei controlli attivati sulle imprese rispondenti per l'aggiornamento di ASIA 2004, per la variabile "attività economica". Anno 2005 (Valori assoluti e percentuali)

	IMPRESE	%
Concordanza ASIA-IULGI	9.562	87,8
Discordanza ASIA-IULGI	1.329	12,2
<i>Sottoposte a controllo</i>	1.072	9,8
<i>Valore IULGI accettato</i>	387	3,6
<i>Valore IULGI respinto</i>	685	6,3
<i>Non sottoposte a controllo (Valore IULGI sempre accettato)</i>	257	2,4
Totale imprese analizzate	10.891	100,0

L'integrazione dei dati d'archivio con i dati d'indagine ha permesso di migliorare la qualità della codifica dell'attività economica in ASIA nel 6% dei casi (3,4% di dati d'indagine accettati in seguito a controllo + 2,6% di dati accettati senza controllo). Per quanto riguarda i casi di valori d'indagine accettati senza controllo, la decisione di accettare le variazioni è stata presa a fronte di dati d'archivio "deboli", cioè provenienti da fonti giudicate di qualità inferiore rispetto all'indagine. Un campione delle imprese di questo gruppo, comunque, è stato successivamente sottoposto a un'ulteriore verifica, che ha confermato la scelta del revisore nel 90% dei casi.

4.3. Aggiornamento della variabile "dipendenti medi"

Il controllo sulla variabile "dipendenti medi" è stato effettuato tra i dipendenti medi di IULGI 2004 e quelli presenti nell'archivio ASIA nello stesso periodo (o, se non presenti, con i dipendenti medi di ASIA al 2003). Nel fornire informazioni sul quadro occupazionale e in particolare all'interno della voce addetti, l'impresa ha incontrato difficoltà nel distinguere i lavoratori indipendenti da quelli dipendenti, sottostimando notevolmente quest'ultimi di fronte a casi di soci lavoratori regolarmente iscritti nei libri paga dell'impresa ma indicati dalla stessa come lavoratori indipendenti.

In presenza di fusioni, conferimenti totali o parziali, scissioni, l'impresa non ha sempre tenuto conto degli effetti contabili e fiscali di tali variazioni strutturali, commettendo errori nella stima del numero medio dei dipendenti quando le operazioni avevano efficacia retroattiva.

Per l'aggiornamento dei dipendenti medi sono state definite le seguenti regole:

- R3) Tra le imprese con meno di 100 dipendenti medi, sono state sottoposte a controllo quelle che, nel confronto fra dati d'archivio e dati d'indagine, presentavano differenze di entità superiore al 25%.
- R4) Tra le imprese da 100 a 250 dipendenti medi, sono state sottoposte a controllo quelle che, nel confronto fra dati d'archivio e dati d'indagine, presentavano differenze di entità superiore al 15%.
- R5) Tra le imprese con oltre 250 dipendenti medi, sono state sottoposte a controllo quelle che, nel confronto fra dati d'archivio e dati d'indagine, presentavano differenze di entità superiore al 10%.

Il controllo *on-line* viene effettuato da parte di revisori esperti dell'archivio ASIA. Tale controllo viene effettuato consultando le fonti amministrative *on-line* a disposizione dell'Istat (bilanci e visure camerali dell'archivio Telemaco di Infocamere, Database Inps) o grazie alle informazioni reperite dai siti Internet delle imprese interessate o da contatti telefonici con le imprese stesse.

Dopo l'analisi di un primo gruppo di imprese che violavano le regole su esposte, si è stabilito a quali condizioni il dato d'indagine potesse essere accettato senza la necessità di ulteriori verifiche. Tali condizioni sono:

- che, nel confronto fra dati d'archivio e dati d'indagine, l'impresa interessata presenti esclusivamente una differenza di dipendenti medi;
- che l'impresa interessata abbia un numero di dipendenti medi, secondo ASIA al tempo t-1, inferiore a 250;
- che il dato dei dipendenti medi di ASIA provenga da una fonte amministrativa.

Tab. 5 – Indagine IULGI: Frequenze dei controlli attivati sulle imprese rispondenti per l'aggiornamento di ASIA 2004, per la variabile "dipendenti medi". Anno 2005 (Valori assoluti e percentuali)

	IMPRESE	%
Concordanza ASIA-IULGI	9.650	88,6
Discordanza ASIA-IULGI	1.241	11,4
<i>Sottoposte a controllo</i>	955	8,8
<i>Valore IULGI accettato</i>	365	3,4
<i>Valore IULGI respinto</i>	590	5,4
<i>Non sottoposte a controllo (Valore IULGI sempre accettato)</i>	286	2,6
Totale imprese analizzate	10.891	100,0

Anche per questa variabile, l'integrazione dei dati d'archivio con i dati d'indagine ha permesso di migliorare la qualità del dato di ASIA sui dipendenti medi nel 6% dei casi di imprese analizzate.

5. Scheda qualità e indicatori

La qualità di una rilevazione statistica è valutata dalla capacità che la stessa ha di soddisfare le esigenze degli 'utenti finali'. Il raggiungimento di tale obiettivo si valuta attraverso l'analisi delle diverse dimensioni incluse nel concetto di qualità, vale a dire: *rilevanza, accuratezza, tempestività e puntualità, accessibilità, confrontabilità e coerenza*. Tali aspetti della qualità vengono misurati attraverso l'utilizzo di alcuni indicatori condivisi a livello europeo e implementati nel sistema SIDI.

In questo paragrafo si illustrano alcuni aspetti della qualità dell'indagine Iulgi analizzati attraverso l'utilizzo sia di indicatori previsti dal sistema SIDI che di alcuni indicatori aggiuntivi.

La *rilevanza* dell'indagine Iulgi è spiegata dagli obiettivi che l'indagine si è prefissata::

- a) Implementare l'Archivio statistico delle imprese attive (Asia) di un nuovo livello informativo sulle unità locali, nel rispetto della normativa comunitaria in materia di Business Register (Regolamento comunitario n. 2186/93);
- b) Cogliere tempestivamente le ripercussioni locali dei cambiamenti della struttura economica del paese;
- c) Completare il processo di integrazione di input e fonti diverse che porta alla creazione di Asia-unità locali (modello statistico, sistema informativo di Asia-imprese)

Per il raggiungimento di tali obiettivi sono necessari tassi di risposta elevati in tutte le ripetizioni dell'indagine, inoltre le informazioni raccolte devono garantire standard elevati di qualità anche in termini di completezza e di accuratezza del dato.

Iulgi è una rilevazione totale (grandi imprese) e gli errori che caratterizzano la rilevazione sono errori di tipo non campionario, in particolare, possono verificarsi errori di lista (a causa del ritardo della lista di partenza), di mancata risposta parziale o totale ed errori di misurazione.

Gli *errori di lista* sono dovuti principalmente ad errori di localizzazione della sede amministrativa dell'impresa (utilizzata per l'invio del questionario) o al mancato aggiornamento dello stato di attività dell'impresa. La lista di spedizione viene estratta dall'Archivio Asia-imprese. In generale, Asia viene pubblicata con un ritardo di 18 mesi, ma l'aggiornamento delle variabili delle grandi imprese continua però nel corso di tutto l'anno; tale aggiornamento rende possibile l'estrazione di liste di spedizione successive alla pubblicazione ufficiale che permettono di ridurre in maniera rilevante questo tipo di errore. Inoltre grazie al sistema di monitoraggio di Iulgi è possibile individuare tempestivamente i questionari respinti, ossia tornati al mittente per impossibilità di consegna da parte del servizio postale a causa di errori nella lista di partenza. La gestione dei respinti compete direttamente alla DCAR, che accerta il motivo della mancata consegna del questionario e nel caso di errori nell'indirizzo provvede alla spedizione del questionario al nuovo indirizzo corretto.

L'aggiornamento delle variabili delle grandi imprese, sia di localizzazione che di struttura, è di tipo micro e viene effettuato da revisori esperti, utilizzando le fonti amministrative on line, tra cui Infocamere (fonte informativa della CCIAA).

Gli errori di *mancata risposta totale e parziale*, sono stati tenuti sotto controllo grazie agli elementi di innovazione che caratterizzano il processo di produzione di Iulgi di cui si è esposto nei precedenti paragrafi e che possono essere così sintetizzati :

- a) La flessibilità nelle procedure di raccolta dati grazie all'utilizzo di linee di lavorazione diverse; per la nuova edizione di Iulgi è prevista una quinta linea di lavorazione basata sull'invio del questionario per posta elettronica.
- b) La vicinanza dei rilevatori all'impresa grazie all'utilizzo della rete degli Uffici Regionali dell'Istat e delle province autonome di Trento e Bolzano.
- c) Il controllo continuo delle operazioni di rilevazione.

Per gli *errori di misura* è previsto sia un controllo manuale che un controllo automatico. La procedura di controllo dei dati prevede infatti, la revisione manuale preliminare dei questionari tornati da parte dei revisori dei vari Uffici Regionali o della DCAR; nel caso di dati mancanti o incongruenti, il revisore provvede ad ricontattare l'impresa al fine di reperire le informazioni corrette.

Dopo la revisione manuale, le informazioni reperite vengono registrate tramite un sistema di data entry controllato; Il check del questionario, basato su regole deterministiche, viene attivato dallo stesso revisore tramite il sito internet dell'indagine; la correzione degli errori viene effettuata mediante imputazione di valori, determinati in base all'esperienza dei revisori, ai dati reperiti da altre indagini Istat o da fonti amministrative (archivio Infocamere, archivio Inps, bilanci delle imprese etc..).

Di seguito vengono analizzati alcuni indicatori di qualità riferiti ai dati raccolti con IULGI 2005; alcuni indicatori sono stati calcolati solo per il sottoinsieme delle imprese con almeno 100 addetti, in quanto lo scopo principale dell'indagine è di garantire la copertura di questo universo.

5.1. Tassi di risposta

Il campo di osservazione di IULGI 2005⁴ comprendeva 9.235 imprese con almeno 100 addetti, per un totale di 60.039 unità locali e oltre 3 milioni di addetti.

Le imprese rispondenti sono state 8.152 per un totale di 61.079 unità locali e di 2.729.749 addetti.

- a) *Tasso di risposta totale al netto delle unità ineleggibili*: dato dal rapporto fra il numero di unità rispondenti e il totale delle unità della popolazione-obiettivo, meno le unità ineleggibili è stato:

$$\text{Unità rispondenti} / (\text{Totale unità} - \text{Unità ineleggibili}) = 87,9\%$$

Le unità ineleggibili sono quelle per le quali è stato possibile accertare che non appartengono alla popolazione-obiettivo: imprese cessate, inattive o fuori campo. Tale indicatore è considerato una misura generica di *performance* dell'indagine nella fase di raccolta dei dati. In tab. 6 sono riportati i tassi di risposta per ripartizione geografica: i valori più bassi si sono registrati nel Nord-ovest e nel Centro.

Tab. 6 – Indagine IULGI: Tasso di risposta totale al netto delle unità ineleggibili e unità ineleggibili per ripartizione geografica. Anno 2005 (Valori percentuali)

	TASSO DI RISPOSTA (al netto delle unità ineleggibili)	UNITÀ INELEGGIBILI	
		Imprese cessate o inattive	Imprese fuori campo di osservazione
Nord-ovest	83,8	3,0	0,05
Nord-est	92,1	2,4	0,12
Centro	85,2	2,7	0,13
Sud	95,4	3,6	0,00
Isole	97,3	1,7	0,33
Totale	87,9	2,8	0,09

- b) *Tasso di risposta totale al netto degli errori di lista*: è dato dal rapporto fra il numero di unità rispondenti e il totale delle unità della popolazione-obiettivo, meno le unità ineleggibili e quelle non contattate a causa di errori di lista.

$$\text{Unità rispondenti} / (\text{Totale unità} - \text{Unità ineleggibili} - \text{Errori di lista}) = 88,2\%$$

Tale indicatore rappresenta, rispetto al precedente, una misura più specifica dell'efficacia della rete di rilevazione. Il *Tasso netto di mancata risposta totale (MRT)* ne è il complemento (tab. 7).

4 Con Iulgi 2006, conclusasi a febbraio 2007, i tassi di risposta delle imprese sopra i 100 addetti hanno superato il 90% nonostante il numero ridotto di risorse (personale) impiegate rispetto agli anni precedenti.

Tab. 7 – Indagine IULGI: Tasso di risposta totale al netto degli errori di lista e tasso netto di mancata risposta totale, per ripartizione geografica. Anno 2005 (Valori percentuali)

	TASSO DI RISPOSTA (al netto degli errori di lista)	TASSO NETTO DI MANCATA RISPOSTA TOTALE (MRT)
Nord-ovest	84,1	15,9
Nord-est	92,1	7,9
Centro	85,8	14,2
Sud	96,1	3,9
Isole	98,0	2,0
ITALIA	88,2	11,8

- c) *Tasso di copertura in termini di addetti al netto delle unità in eleggibili:* tale indicatore non è ancora previsto dal sistema SIDI, ma per le statistiche economiche dà una misura più indicativa della copertura dell'universo indagato. Gli addetti delle imprese rispondenti (con almeno 100 addetti) rappresentano i 2/3 dell'intera occupazione e oltre il 90% delle imprese di questo universo.

Tab. 8 – Indagine IULGI: Tasso di copertura degli addetti al netto degli errori di lista e tasso di mancata copertura, per ripartizione geografica. Anno 2005 (Valori percentuali)

	TASSO DI COPERTURA DEGLI ADDETTI (AL NETTO DELLE IMPRESE INELEGGIBILI)	TASSO DI MANCATA COPERTURA
Nord-ovest	86,5	13,5
Nord-est	94,9	5,1
Centro	89,4	10,6
Sud	96,4	3,6
Isole	98,3	1,7
ITALIA	90,2	9,8

- d) *Tasso di copertura in termini di unità locali:* anche quest'indicatore non è previsto dallo schema SIDI ma è stato calcolato in quanto obiettivo fondamentale dell'indagine è di implementare l'archivio delle imprese di questo nuovo livello informativo.

Tab. 9 – Indagine IULGI: Tasso di copertura in termini di unità locali sia partite che tornate. Anno 2005 (Valori percentuali)

	u.l. partite	%	u.l. tornate	%
u.l. attive partite al 2004	45.205	75,3	45.205	74,0
u.l. non attive o non appartenenti all'impresa	5.968	9,9	5.968	9,8
u.l. appartenenti ad imprese ineleggibili (cessate, inattive o f.c)	1.045	1,7	1.045	1,7
u.l. appartenenti ad imprese non rispondenti	7.821	13,0		
u.l. mancanti dalla lista di partenza			5.308	8,7
u.l. nuove nate o acquisite nel 2005			3.553	5,8
u.l. totali	60.039	100,0	61.079	100,0

5.2. Tasso d'imputazione

Il tasso d'imputazione misura l'incidenza dei valori modificati da una procedura di *editing* e imputazione (E&I), che può essere automatica, manuale o mista. La modifica è intesa come cambiamento:

- i. da valore *blank* a valore *non-blank*;
- ii. da valore *non-blank* ad altro valore *non-blank*;
- iii. da valore *non-blank* a valore *blank*.

Si tratta di un indicatore di *performance* della procedura di E&I. Un valore elevato può essere il sintomo di diverse cause: molti valori mancanti, molti errori nei dati (soprattutto in presenza di cambiamenti da *non-blank* a *non-blank*), un piano di E&I troppo invasivo o contenente errori, ecc. Il calcolo dell'indicatore richiede che sia stata tenuta traccia di tutte le modifiche intervenute tra il file dei dati grezzi ottenuto al termine della fase di raccolta e registrazione e il file dei dati finale. L'indicatore è tra quelli previsti dal sistema SIDI ed è anche uno degli indicatori standard di Eurostat.

Tale indicatore è stato calcolato su tutte le unità rispondenti, sia a livello di impresa che di unità locale. Sono state selezionate alcune variabili ritenute di fondamentale importanza per l'indagine. L'edit viene attivato ogni qualvolta si verifica un errore o un accertamento in base alle regole previste dal piano di check dell'indagine. Detto piano prevede delle regole molto rigide nel controllo della variabile dipendenti a livello di impresa sia come media che come dato puntuale al 30/06, spiegato dalla necessità di tenere sotto controllo gli eventi (fusioni, acquisizioni, etc..) che caratterizzano sempre più il ciclo di vita delle imprese.

Tab. 10 – Indagine IULGI: Frequenze di attivazione degli edit della procedura di controllo e correzione e tassi d'imputazione delle principali variabili. Anno 2005 (Valori assoluti e percentuali)

VARIABILI	FREQUENZE DI ATTIVAZIONE DEGLI EDIT		NUMERO DI RECORD	TASSO D'IMPUTAZIONE (%)
	<i>(numero di correzioni)</i>			
ATECO_IMPRESA	555	12.143		4,57
CODICE_FISCALE	50	12.143		0,41
ATECO_UL	5.152	80.407		6,41
TOTALE_ADDETTI_30_06_2005	3.109	12.143		25,60
TOTALE_ADDETTI_MEDIA_2004	1.958	12.143		16,12
NUMERO_UL	233	12.143		1,92
TIPO_UL	1.553	80.407		1,93
STATO_ATTIVITA_UL	1.314	80.407		1,63
TOTALE_ADDETTI_UL_30_06_2005	2.522	80.407		3,14

5.3. Tempestività e puntualità

La tempestività è l'adeguatezza del rilascio dei dati alle esigenze degli utenti. In genere si approssima attraverso la misura della distanza tra la data effettiva di rilascio dei risultati e quella di riferimento dei dati. La puntualità è il rispetto degli obiettivi temporali di rilascio dei dati. Può essere misurata attraverso la distanza di tempo intercorrente tra la data effettiva di rilascio dei dati e quella prevista.

IULGI produce dati non destinati alla diffusione ma solo all'aggiornamento dell'archivio ASIA. In questo senso, la tempestività delle informazioni prodotte può essere misurata solo indirettamente, sul processo di ASIA-Unità locali, che rappresenta il "prodotto finito" cui l'indagine, per parte sua, contribuisce.

I dati di IULGI, tuttavia, vengono trasmessi alla Contabilità nazionale ai fini della stima del PIL regionale: quelli dell'edizione 2005 sono stati trasmessi alla Contabilità il 31 gennaio 2006, a circa 7 mesi dalla data di riferimento dell'indagine (30 giugno 2005). Possiamo quindi calcolare l'indicatore di *Tempestività per Contabilità nazionale*:

$$\text{Tempestività per CN} = \text{Data fornitura} - \text{Data riferimento indagine} = (31/1/06) - (30/6/05) = 215 \text{ gg.}$$

L'innovazione del sistema IVR per il monitoraggio dell'indagine multiscopo "Condizioni di salute e ricorso ai servizi sanitari 2004-2005"

Laura Iannucci, Manuela Michellini, Giuseppina Pica

1 – Introduzione

Dal 1994 l'indagine Condizioni di salute e ricorso ai servizi sanitari fa parte del Sistema di indagini multiscopo sulle famiglie e rappresenta una delle principali fonti informative per valutare lo stato di salute della popolazione. Attraverso le informazioni raccolte nell'indagine è possibile costruire indicatori sulle condizioni e percezione della salute, sugli stili di vita e la prevenzione, sui servizi sanitari e trattamenti, sul percorso assistenziale di gravidanza e parto.

L'indagine è di tipo campionario, ha cadenza quinquennale, l'ultima si è svolta nel 2004-2005 e prevede che la rilevazione sia suddivisa in quattro trimestri (dicembre, marzo, giugno, settembre). Le informazioni sono raccolte con tecnica PAPI (*Paper and Pencil Interview*) mediante intervista diretta effettuata dal rilevatore comunale. Nell'ultima indagine sono state intervistate oltre 50.000 famiglie campione distribuite sul territorio nazionale.

In occasione di quest'ultima indagine sono state introdotte importanti novità nella gestione dell'intero processo di rilevazione ed in particolare nel monitoraggio del lavoro sul campo al fine di prevenire il più possibile gli errori non campionari, caratteristici di questa fase d'indagine, e quindi migliorare la qualità dei dati raccolti.

Tramite il sistema Sigif (*Sistema Informativo Gestione Indagini sulle Famiglie*) è stato possibile informatizzare e registrare su un database unico le informazioni di tutte le fasi del processo di rilevazione: estrazione dei comuni campione, estrazione delle famiglie campione, assegnazione delle famiglie ai rilevatori, formazione degli intervistatori, rilevazione sul campo, trasmissione dei materiali, revisione, controlli di qualità ecc. Inoltre con SIGIF è stato possibile realizzare un sistema di comunicazione e raccordo tra centro e territorio finalizzato alla gestione delle diverse fasi dell'indagine. L'informatizzazione ha, in parte, cambiato il ruolo e il tipo di lavoro dei tre livelli territoriali coinvolti nel processo d'indagine, gli uffici centrali dell'Istat, gli uffici territoriali regionali e i Comuni. In particolare questi ultimi hanno iniziato l'acquisizione elettronica dei nominativi delle famiglie estratte, l'assegnazione delle famiglie al rilevatore, la gestione delle anomalie e degli esiti mancanti e il monitoraggio della trasmissione del materiale di rilevazione.

Con il sistema IVR (*Interactive Voice Response*), invece, è stata informatizzata la raccolta delle informazioni relative ai contatti che i rilevatori comunali avevano con le famiglie. In particolare, le novità del nuovo sistema di monitoraggio della rilevazione sul campo risiedono soprattutto nell'acquisizione rapida delle informazioni sull'attività svolta dai rilevatori, mediante inserimento interattivo tramite tastiera telefonica, e nella possibilità di consultare on line gli indicatori di qualità dell'indagine aggiornati quotidianamente. L'introduzione dell'IVR ha consentito quindi una supervisione in tempo reale della fase di rilevazione dati.

L'efficacia del controllo in corso d'opera della rilevazione, riguardo al rispetto della metodologia, delle procedure e dei tempi stabiliti per l'esecuzione delle interviste, è infatti subordinata alla disponibilità completa e tempestiva delle informazioni.

Il presente lavoro illustra le principali modifiche e innovazioni dell'informatizzazione del processo di rilevazione, con particolare riferimento al monitoraggio della rilevazione tramite il sistema IVR, con la presentazione di alcuni specifici risultati riferiti all'indagine.

2 – La gestione del processo di rilevazione

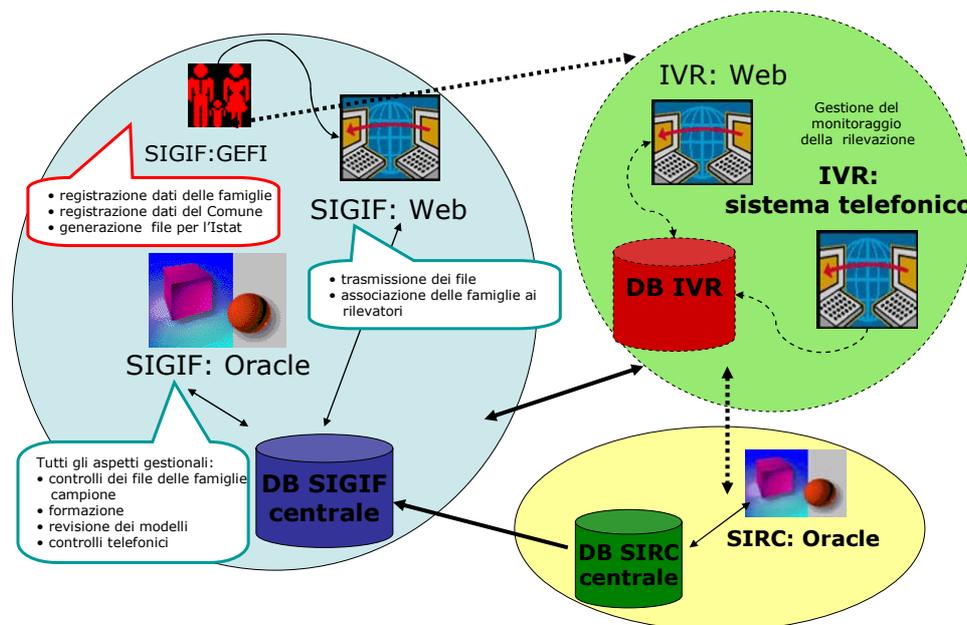
Le profonde trasformazioni di gestione del processo di indagine si basano sull'utilizzo di tre sistemi informativi integrati tra di loro: SIGIF, SIRC¹, IVR (Figura 1).

Attraverso un database unico vengono registrate tutte le informazioni del processo d'indagine provenienti dalle diverse attività dei tre livelli di competenza nelle varie fasi: ufficio centrale Istat, uffici regionali, Comuni. Durante la fase di rilevazione il database viene integrato con le informazioni provenienti dal sistema IVR. Per facilitare l'utilizzo dei dati anagrafici dei rilevatori comunali il sistema Sigif è integrato con il sistema SIRC del quale utilizza il nome e cognome e il codice rilevatore.

Al data base è affiancato il portale Web per la gestione dei rapporti tra Istat e comuni campione dalla fase di definizione ed estrazione del campione alla fase di rilevazione sul campo e alle attività di monitoraggio. Permette di gestire la trasmissione dei file, l'invio dei pacchi, le informazioni relative alle circolari, modelli e materiale di formazione ed in particolare le funzioni di assegnazione delle famiglie agli intervistatori comunali. Si tratta di un sito dedicato sviluppato internamente all'Istat. Il sito garantisce la sicurezza nello scambio delle informazioni tramite l'utilizzo di due livelli di identificazione: il primo per entrare nella pagina dedicata alle indagini sulle famiglie, il secondo specifico per l'indagine.

Figura 1

L'integrazione dei sistemi SIGIF, SIRC e IVR



Il sistema Sigif è implementato in vari sottosistemi per specifiche fasi del processo d'indagine:

- *Sottosistema dedicato alle fasi di definizione e organizzazione del campione* (individuazione e gestione dei comuni campione, acquisizione degli elenchi delle famiglie campione).

Gli uffici regionali sostituiscono i comuni campione in base alla presenza di difficoltà oggettive ad effettuare l'indagine coerentemente con lo strato di appartenenza. Attraverso il portale i comuni spediscono il file delle famiglie campione che viene sottoposto a dei controlli formali e di congruità interna da parte dei responsabili Istat.

L'acquisizione elettronica delle famiglie ha sostituito totalmente il modello cartaceo IMF/1 tradizionalmente utilizzato dai comuni nelle indagini multiscopo per elencare le famiglie estratte e assegnare le famiglie ai rilevatori. La nuova fase di acquisizione dei nominativi inizia con la memorizzazione delle informazioni di ogni singola famiglia e di tutti i suoi componenti in un unico file secondo il tracciato record allegato alla circolare per i comuni. L'Istat mette a disposizione il software

1 Sistema Informativo rete di Rilevazione Comunale

Gefi (*Gestione elenchi famiglie campione*) che può essere utilizzato o per registrare direttamente i dati delle famiglie estratte dai comuni che effettuano l'estrazione non automatizzata, o in fase di spedizione del file all'Istat per i comuni che effettuano l'estrazione automatizzata. In ambedue i casi viene utilizzato per registrare i dati del comune (indirizzo del comune, tipo di anagrafe, tipo di ordinamento, ecc.) e preparare il file per la spedizione tramite portale.

Questo metodo consente all'Istat l'immediata acquisizione degli elenchi delle famiglie campione e il conseguente utilizzo dei dati per la costruzione degli indicatori, come ad esempio quelli per il controllo della struttura della famiglia per numero dei componenti.

- *Sottosistema per l'assegnazione delle famiglie ai rilevatori*

Attraverso il portale i comuni assegnano ad ogni rilevatore le famiglie che dovrà intervistare. In questa fase i comuni hanno a disposizione l'intero elenco dei rilevatori proveniente dall'archivio SIRC mentre le famiglie sono quelle estratte dal Comune.

Sottosistemi trasversali o sviluppati per il controllo delle fasi successive del processo di produzione dei dati, ad uso dei supervisor presso gli UURR, riguardano:

- *la registrazione delle riunioni d'istruzione* sia nella parte di rendicontazione delle presenze sia nella parte qualitativa;

- *l'attività di revisione dei modelli*: si procede al controllo quantitativo dei questionari compilati e alla verifica della corretta trascrizione dei codici identificativi dei comuni e delle famiglie;

- *l'attività di "controllo"* dell'operato dei rilevatori: per ciascun intervistatore si procede alla reintervista di almeno due famiglie al fine di verificare che le interviste siano avvenute secondo le indicazioni metodologiche impartite.

Il sistema Sigif in tutte le sue componenti presenta delle funzioni per l'analisi e la reportistica delle informazioni archiviate. Per ciascuna attività prevista dal processo d'indagine è possibile monitorare il suo stato di avanzamento in tempo reale secondo un dettaglio analitico (per famiglia o per rilevatore) o sintetico (per comune o regione).

3 – Il sistema IVR

Tecnicamente il sistema IVR si fonda su di un sistema interattivo di risposta vocale a toni, operante 24 ore al giorno per tutto il periodo di rilevazione. Esso è collegato ad un numero verde, messo a disposizione dall'Istat, attraverso il quale il rilevatore comunica, tramite tastiera telefonica, le informazioni sui contatti avuti con ogni famiglia. Le informazioni rilevate vengono registrate su supporto informatico e rese disponibili all'Istat in tempo reale. La base dati così costituita viene utilizzata per il calcolo di un set di indicatori di qualità dell'indagine, aggiornati quotidianamente e consultabili, dai soggetti coinvolti nel monitoraggio della rilevazione (responsabili dell'indagine a livello centrale, regionale e comunale), accedendo al sito Web appositamente costruito.

3.1 L'accesso al sistema

L'accesso al sistema IVR per la comunicazione degli esiti dei contatti con le famiglie richiede il codice dell'indagine, il codice rilevatore SIRC e la password IVR consegnata al rilevatore durante la formazione. Solo nel caso di corretta associazione codice rilevatore–password il sistema consente di proseguire, altrimenti, dopo tre inserimenti errati, la connessione viene interrotta. Ai rilevatori durante la formazione sono state consegnate delle schede di monitoraggio sulle quali annotare le informazioni da comunicare tramite IVR: numero d'ordine della famiglia in elenco, numero di telefono della famiglia, tipologia del contatto (telefonico oppure a domicilio), esito del contatto, data e l'ora del contatto, ricezione della lettera di preavviso, durata in ore e minuti dell'intervista, numero dei componenti della

famiglia di fatto, totale delle schede individuali compilate proxy, totale dei modelli autocompilati ritirati (di cui compilati proxy e di cui compilati con l'aiuto del rilevatore).

Il sistema effettua on-line alcuni controlli di congruità tra tipologia del contatto ed esito. Ad esempio, non è possibile registrare l'esito di intervista completa se è stata scelta la tipologia corrispondente al contatto telefonico poiché le interviste debbono essere effettuate solo al domicilio del rispondente; così come non è accettata la selezione dell'esito di occupato/fax a fronte di un contatto a domicilio. Al fine di ridurre i tempi della telefonata e per non ostacolare il normale svolgimento della comunicazione da parte del rilevatore, altri controlli di congruità vengono effettuati off-line tramite una procedura programmata la quale suddivide i contatti comunicati in tre categorie: validi, anomalie e scarti. I record validi costituiscono la base per il calcolo degli indicatori di qualità, mentre i restanti vengono isolati. In particolare gli scarti, costituiti da tutti quei contatti per i quali non sono presenti sufficienti dati (rilevatore non riconosciuto, assenza di codice famiglia, assenza di esito), non possono essere utilizzati. I record anomali invece presentano incongruenze che, con l'aiuto dell'intervistatore, possono essere sanate (presenza di un esito provvisorio dopo un esito definitivo, data e ora non validi, ecc..) e quindi utilizzate per il calcolo degli indicatori di qualità insieme agli altri dati validi.

3.2 Il sito Web

Il sito Web, oltre alla consultazione degli indicatori di qualità, prevede delle funzionalità specifiche per la gestione della rilevazione a cura del responsabile dell'indagine presso il comune o dell'ufficio territoriale competente:

- *visualizzazione di tutti i contatti comunicati all'IVR;*

- *gestione dei carichi di lavoro* dei rilevatori: qualora un rilevatore non può portare a termine le sue interviste, il responsabile comunale dell'indagine può riassegnare le famiglie ad un altro rilevatore, attraverso una funzione del sito;

- *gestione e correzione dei record anomali*: il responsabile d'indagine comunale e gli uffici regionali gestiscono la correzione dei dati di monitoraggio incongruenti fra loro;

- *inserimento degli esiti definitivi mancanti* sulle famiglie per quei rilevatori che non hanno utilizzato l'IVR oppure l'inserimento degli esiti definitivi per le famiglie parzialmente monitorate tramite IVR.

Le ultime due funzioni permettono di recuperare le informazioni sugli esiti definitivi necessarie al calcolo degli indicatori e limitare in questo modo la mancata copertura del sistema di monitoraggio.

4 - L'uso dell'IVR da parte dei rilevatori

Gli intervistatori avevano il compito di comunicare entro due giorni le informazioni sui contatti intrapresi con le famiglie campione loro assegnate.

Complessivamente circa il 93% dei rilevatori comunali coinvolti nell'indagine ha utilizzato il sistema IVR per comunicare le informazioni richieste sugli esiti dei contatti con le famiglie e sulle modalità dell'intervista. Il livello di partecipazione degli intervistatori è stato elevato per tutti e quattro i trimestri di rilevazione soprattutto se consideriamo che era la prima volta in cui veniva monitorata la fase di raccolta dati nel caso di una indagine campionaria PAPI affidata ai Comuni (Tavola 1).

Tavola 1 Indicatori sull'uso del sistema IVR per trimestre di rilevazione

	Trimestre di rilevazione				Totale
	1°	2°	3°	4°	
% di rilevatori che hanno partecipato al monitoraggio	92,4	91,3	94,1	93,4	92,8
% di contatti comunicati entro 2 giorni	40,1	48,1	46,4	45,6	45,1
% di contatti comunicati entro 7 giorni	70,3	75,7	77,1	78,0	75,4
Tasso di copertura del monitoraggio	86,6	88,1	90,7	92,1	89,4

a) Il tasso di copertura del monitoraggio è calcolato come numero di famiglie monitorate con IVR sul numero famiglie campione assegnate ai rilevatori

Il 45,1% delle comunicazioni è avvenuta entro due giorni rispettando la tempistica prescritta. Nel complesso, la maggior parte degli esiti (75,4%) è stata comunicata al massimo entro sette giorni dal contatto con la famiglia. Ciò ha permesso ugualmente di intervenire in modo tempestivo sulle situazioni critiche poiché la rilevazione, in ogni trimestre, durava circa un mese. La tempestività della comunicazione è migliorata nel corso dei quattro trimestri di rilevazione: infatti, si osserva un aumento nelle rispettive quote dei contatti comunicati entro due giorni (dal 40,1% nel primo trimestre al 45,6% nel quarto trimestre) ed entro la settimana (dal 70,3% nel primo trimestre al 78% nel quarto).

La partecipazione dei rilevatori al monitoraggio mediante IVR risulta associata al carico di lavoro, ossia al numero di famiglie campione assegnate a ciascun rilevatore per le quali è richiesta la comunicazione degli esiti dei contatti al numero verde dell'IVR. All'aumentare delle famiglie da monitorare cresce infatti, la quota dei rilevatori che non effettua alcuna comunicazione (Tavola 2). Il numero ottimale di famiglie da assegnare a ciascun rilevatore per favorire il monitoraggio della rilevazione non dovrebbe superare le 20 unità. La relazione tra partecipazione al monitoraggio e carico di lavoro era prevedibile e per questo motivo in fase di progettazione è stato chiesto ai Comuni di assegnare a ciascun rilevatore non più di 15 di famiglie. La maggior parte dei comuni ha seguito questa indicazione, solo una quota residuale di comuni per problemi organizzativi ha dovuto assegnare un numero superiore di famiglie ai propri rilevatori.

Tavola 2 - Rilevatori per partecipazione al monitoraggio IVR e numero di famiglie assegnate.

N° di famiglie assegnate	Partecipazione dei rilevatori al monitoraggio		Totale
	Ha partecipato	Non ha partecipato	
1-10	90,9	9,1	100,0
11-20	95,6	4,4	100,0
21-30	87,8	12,2	100,0
31 e più	83,3	16,7	100,0
Totale	92,8	7,2	100,0

La partecipazione all'IVR è stata invogliata anche grazie ad un sistema di gestione telefonica snello. I rilevatori potevano inserire tutti i contatti effettuati nei due giorni con una sola chiamata, di durata relativamente breve. Per comunicare gli esiti di tutte le famiglie assegnate ciascun rilevatore ha effettuato mediamente 8 telefonate al numero verde dell'IVR. Nel corso di ogni telefonata sono stati comunicati in media gli esiti di 3 contatti impiegando mediamente 5 minuti. La durata della telefonata al numero verde dipendeva dal tipo di esito da comunicare: circa 90 secondi per comunicare un contatto di caduta della famiglia o di un appuntamento e circa il doppio (180 secondi) per comunicare un esito di intervista completa. In quest'ultimo caso infatti, il rilevatore doveva comunicare più informazioni: durata dell'intervista, numero di componenti la famiglia, numero di questionari compilati, numero di questionari compilati proxy, ricezione della lettera di preavviso, ecc.

L'attività di monitoraggio ha coperto circa il 90% delle famiglie campione assegnate ai rilevatori (Tavola 1). Durante la rilevazione per queste famiglie è stato possibile seguire quotidianamente sia l'esito dei contatti intrapresi di volta in volta dai rilevatori, sia la tecnica di somministrazione dei questionari e la durata complessiva dell'intervista. La quota delle famiglie monitorate è costantemente aumentata nel corso dei quattro trimestri di rilevazione: dall'86,6% nel 1° trimestre è passata al 91,6% nel 4° trimestre.

Per circa il 10% delle famiglie assegnate non è stato possibile effettuare il monitoraggio quotidiano a causa della mancata comunicazione delle informazioni da parte degli intervistatori. Si è tentato di colmare questa lacuna informativa richiedendo ai responsabili comunali e regionali dell'indagine di comunicare a fine rilevazione, attraverso il Sito Web, l'esito definitivo delle famiglie non monitorate.

Con queste informazioni aggiuntive è stato possibile calcolare gli indicatori a fine rilevazione per tutte le famiglie campione e avere le informazioni necessarie per intervenire nei trimestri successivi di rilevazione.

5 – Gli esiti dei contatti e gli indicatori di monitoraggio

Le esperienze di monitoraggio delle indagini sulle famiglie hanno portato ad una standardizzazione degli esiti dei contatti utilizzabili sia per le indagini condotte telefonicamente che per quelle faccia a faccia secondo quanto proposto da ‘The American Association for Public Research’ (AAPOR). Gli esiti dei contatti, introdotti nell’indagine ‘‘Condizione di salute e ricorso ai servizi sanitari 2004-2005’’ sono suddivisi in provvisori e definitivi e classificati come previsto nel sistema SIDI in unità risolte e non risolte, eleggibili e non eleggibili, rispondenti e non rispondenti. Per la prima volta si dispone anche per un’indagine PAPI degli esiti provvisori in analogia alle indagini CATI.

Gli indicatori di qualità calcolati per il monitoraggio della rilevazione sul campo sono stati definiti da un gruppo di lavoro costituito da persone con esperienze maturate nell’ambito delle indagini sulle famiglie con tecnica CATI, PAPI, CAPI (multiscopo, forze di lavoro e consumi).

Tali indicatori, accessibili, per la prima volta nell’ambito di indagini campionarie PAPI, on line per tutti i soggetti coinvolti nell’attività di monitoraggio (Comuni, UURR, Servizio centrale) hanno permesso di seguire quotidianamente lo svolgimento dell’indagine nei suoi diversi aspetti: lo stato di avanzamento della rilevazione, la mancata partecipazione con i suoi motivi, le modalità di compilazione dei questionari, ecc.

Il primo gruppo di indicatori raccoglie informazioni sull’andamento complessivo della rilevazione: tasso di assegnazione, tasso di copertura del monitoraggio, tasso di completezza, tasso di risposta, tasso di caduta, tasso di rifiuto, tasso di sospensione, ecc.

Il secondo gruppo di indicatori analizza in dettaglio la non partecipazione all’indagine: famiglie cadute per motivo della caduta, famiglie cadute per numero di tentativi di contatto, famiglie che hanno rifiutato per motivo del rifiuto, ecc.

Il terzo gruppo riguarda la qualità del materiale raccolto: tassi di compilazione proxy, tasso di intervista diretta per i questionari autocompilati, ecc.

Il quarto gruppo raccoglie indicatori sulla comunicazione con le famiglie: esito dei contatti telefonici e a domicilio per fasce orarie e per giorni della settimana, assenza del numero telefonico della famiglia, ricezione della lettera di preavviso, ecc.

Infine, il quinto gruppo raccoglie indicatori di survey per valutare complessivamente l’indagine ad uso esclusivo del referente centrale e del suo gruppo di lavoro.

Ciascun utente aveva la possibilità di analizzare gli indicatori secondo specifici dettagli temporali: giornaliero, settimanale, trimestrale. Il dettaglio territoriale degli indicatori disponibili on line variava a seconda dell’utente: il responsabile comunale aveva visione solo sugli indicatori relativi ai rilevatori del suo comune; il responsabile regionale aveva accesso agli indicatori relativi ai comuni e ai rilevatori impegnati nell’indagine nella regione di competenza; il responsabile centrale dell’indagine aveva accesso a tutti gli indicatori dal più sintetico al più analitico.

6 - Gli interventi sulle situazioni critiche durante la rilevazione

6.1 Lo stato di avanzamento dell’indagine

L’obiettivo del monitoraggio della rilevazione è stato quello di individuare tempestivamente le situazioni critiche, attraverso l’analisi degli indicatori di qualità disponibili sul sito WEB, e quindi adottare gli opportuni interventi correttivi prima che la rilevazione fosse conclusa.

I rilevatori che ad una settimana dalla rilevazione non avevano ancora comunicato un numero sufficiente di contatti con le famiglie, facendo quindi osservare un basso *tasso di copertura*², venivano contattati dal responsabile d'indagine regionale per capire i motivi del ritardo. In alcuni casi si trattava di rilevatori che, pur avendo contattato le famiglie, non avevano poi comunicato l'esito al numero verde. Si è cercato quindi di motivarli all'utilizzo del sistema IVR ribadendo gli obiettivi di qualità del monitoraggio e l'importanza della tempestività della comunicazione. In altri casi invece, trattandosi di rilevatori effettivamente in ritardo con le interviste, si è chiesto loro di intensificare il lavoro al fine di rispettare il periodo di rilevazione indicato dall'Istat.

Nel corso dei quattro trimestri di rilevazione si è intensificata l'attività di contatto delle famiglie nella prima settimana di rilevazione: il tasso di copertura è progressivamente aumentato passando dal 16,3% nel primo trimestre al 44,9% nel quarto (Tavola 3).

Tavola 3 Tassi di copertura e di sospensione per trimestre e settimana di rilevazione (a)

	Prima settimana		Seconda settimana		Terza settimana		Quarta settimana	
	Tasso di copertura (b)	Tasso di sospensione (c)	Tasso di copertura (b)	Tasso di sospensione (c)	Tasso di copertura (b)	Tasso di sospensione (c)	Tasso di copertura (b)	Tasso di sospensione (c)
Primo trim.	16.3	59.8	41.6	39.4	65.0	22.1	86.6	3.7
Secondo trim.	26.6	51.4	56.9	33.7	76.2	15.8	88.2	3.8
Terzo trim.	27.4	60.4	60.6	36.7	82.1	17.4	90.9	3.5
Quarto trim.	44.9	44.6	73.4	24.4	88.6	5.6	91.6	2.2

(a) I dati della seconda, terza e quarta settimana sono cumulati

(b) Il tasso di copertura del monitoraggio indica quante famiglie assegnate sono monitorate. E' calcolato come numero di famiglie monitorate su numero di famiglie assegnate per 100

(c) Il tasso di sospensione indica quante famiglie monitorate hanno solo un esito provvisorio. E' calcolato come numero di famiglie con esito provvisorio su numero di famiglie monitorate per 100

I rilevatori che a fine periodo di rilevazione non avevano ancora completato le interviste (elevato *tasso di sospensione*³) venivano invitati a completare il lavoro entro i termini prestabiliti. Solo in casi del tutto eccezionali, legati a problemi organizzativi del comune o di disponibilità della famiglia, sono state ammesse delle proroghe di pochi giorni.

La quota di famiglie monitorate rimasta senza esito definitivo a fine rilevazione è diminuita nel corso dei quattro trimestri passando dal 3,7% nel primo trimestre al 2,2% nel quarto (Tavola 3). Come già detto, per queste famiglie è stato chiesto al responsabile d'indagine comunale e in subordine al responsabile regionale, di comunicare, a fine trimestre di rilevazione, l'esito definitivo mediante l'apposita funzione del Sito Web.

6.2 La mancata risposta totale

Un ulteriore aspetto monitorato quotidianamente è stato quello della mancata risposta totale che ha degli evidenti effetti sulla qualità dei dati incrementando la variabilità e la distorsione delle stime dei parametri della popolazione.

² Il tasso di copertura del monitoraggio indica quante famiglie assegnate ai rilevatori sono monitorate. Una famiglia si intende monitorata se per essa sono state comunicate le informazioni su almeno un contatto con esito provvisorio o definitivo. Il tasso è calcolato come numero di famiglie monitorate su numero di famiglie assegnate.

³ Il tasso di sospensione fornisce una misura di quanta parte del campione monitorato non ha ancora un esito definitivo. E' calcolato come numero di famiglie con esiti solo provvisori su numero di famiglie monitorate.

In presenza di un elevato *tasso di caduta*⁴ è stato dunque necessario analizzarne il motivo (rifiuto, errore di lista, irripetibilità) ed intervenire in modo adeguato.

I rilevatori con un *tasso di rifiuto*⁵ superiore alla media sono stati oggetto di analisi più approfondite per tentare di comprendere se il comportamento non collaborativo delle famiglie era imputabile all'approccio utilizzato dagli intervistatori oppure ad altri fattori non direttamente riconducibili all'intervistatore stesso come ad esempio la mancata ricezione della lettera di preavviso. E' infatti, ben nota la maggiore collaboratività delle famiglie preavvisate dell'intervista da una lettera ufficiale in cui si esplicitano gli obiettivi dell'indagine, il rispetto della privacy e la disponibilità di un numero verde a cui rivolgersi per eventuali chiarimenti. Ai rilevatori con elevati tassi di rifiuto il responsabile d'indagine suggeriva di motivare le famiglie all'intervista facendo leva soprattutto sugli obiettivi conoscitivi dell'indagine e sul rispetto della privacy. Inoltre il lavoro degli intervistatori problematici è stato seguito con attenzione nel tempo per cogliere eventuali miglioramenti nel tasso di rifiuto.

Tavola 4 Tasso di caduta e tasso di rifiuto per trimestre e settimana di rilevazione (a)

	Prima settimana		Seconda settimana		Terza settimana		Quarta settimana	
	Tasso di caduta (b)	Tasso di rifiuto (c)	Tasso di caduta (b)	Tasso di rifiuto (c)	Tasso di caduta (b)	Tasso di rifiuto (c)	Tasso di caduta (b)	Tasso di rifiuto (c)
Primo trim.	3.6	6.9	4.6	5.2	6.6	4.9	11.6	5.9
Secondo trim.	3.7	4.9	5.8	5.4	9.3	6.3	13.8	7.4
Terzo trim.	3.6	5.9	5.5	5.2	8.4	5.7	12.4	6.2
Quarto trim.	4.5	4.2	6.3	4.7	10.7	5.3	11.7	5.4

(a) I dati della seconda, terza e quarta settimana sono cumulati.

(b) Il tasso di caduta indica la percentuale di famiglie cadute (esito=12-34) sul totale delle famiglie assegnate monitorate.

(c) Il tasso di rifiuto indica la percentuale di famiglie cadute per rifiuto (esito 24-34) sul totale delle famiglie intervistate (esito 11), cadute per interruzione (esito 12) e per rifiuto (esito 24-34).

Un altro motivo di caduta può essere attribuito alla difficoltà di contattare le famiglie perché poco presenti in casa oppure assenti durante l'intero periodo di rilevazione. Questo aspetto è stato monitorato mediante il *tasso di caduta per irripetibilità*⁶. Al fine di limitare questo tipo di caduta è stato indicato agli intervistatori, nella formazione e durante la rilevazione, di contattare le famiglie in diverse fasce orarie e mostrare maggiore disponibilità a condurre le interviste negli orari più comodi alle famiglie.

4 Il tasso di caduta fornisce la misura di quanto il campione teorico viene ridotto indipendentemente dai motivi della caduta. Esso viene calcolato rapportando il numero di famiglie che non hanno ottenuto esito di intervista completa sul numero di famiglie assegnate. Durante la rilevazione, trattandosi di un indicatore di performance degli intervistatori, il tasso è calcolato rispetto agli esiti provenienti dal sistema IVR e al denominatore si ha il numero di famiglie assegnate.

5 Il tasso di rifiuto viene calcolato come numero di famiglie cadute per rifiuto (esito 24-34) sul totale delle famiglie intervistate (esito 11), cadute per interruzione (esito 12) e per rifiuto (esito 24-34). Esso è calcolato come tasso netto di rifiuto ossia rispetto al totale delle persone che hanno avuto la possibilità di rilasciare l'intervista (intervista completa, interruzione definitiva, rifiuto). Inoltre al numeratore sono inclusi tutti i motivi di rifiuto, anche quelli che comunemente negli indicatori SIDI sono considerati in altri motivi di risposta (malattia grave, problemi familiari, stranieri che non parlano italiano, problemi di comprensione).

6 Il tasso di caduta per irripetibilità indica la percentuale di famiglie cadute dopo aver effettuato tutti i contatti previsti senza essere però riusciti a parlare con qualcuno della famiglia (esito=13) oppure tutti gli appuntamenti previsti senza essere riusciti ad effettuare/concludere l'intervista (esito=14) oppure per assenza della famiglia durante l'intero periodo di rilevazione (esito=15) oppure per impossibilità di accedere all'indirizzo (esito=22) sul totale di famiglie assegnate monitorate.

Un ulteriore motivo di caduta può derivare dal mancato aggiornamento della lista anagrafica di estrazione rispetto a decessi o a trasferimenti in convivenze o in altri comuni (italiani o esteri). Il *tasso di caduta per errore di lista*⁷ durante i quattro trimestri di rilevazione è stato di circa il 4%. In presenza di un'incongruenza tra lista anagrafica e situazione di fatto è stato chiesto al rilevatore di chiedere al comune ulteriori elementi per reperire la famiglia. Solo dopo questo accertamento la famiglia può considerarsi caduta. In generale, ai casi di ritardato aggiornamento della lista anagrafica, da parte del Comune, si aggiungono le mancate comunicazioni, da parte dei cittadini, riguardanti soprattutto anziani che si trasferiscono negli istituti o a casa dei figli, o agli stranieri che si spostano in un altro Comune o all'estero.

Tavola 5 Tasso di errori di lista e tasso di irreperibilità per trimestre e settimana di rilevazione (a)

	Prima settimana		Seconda settimana		Terza settimana		Quarta settimana	
	Tasso di caduta per errori di lista (b)	Tasso di caduta per irreperibilità (c)	Tasso di caduta per errori di lista (b)	Tasso di caduta per irreperibilità (c)	Tasso di caduta per errori di lista (b)	Tasso di caduta per irreperibilità (c)	Tasso di caduta per errori di lista (b)	Tasso di caduta per irreperibilità (c)
Primo trim.	0.5	0.3	1.0	0.5	1.6	1.2	2.4	3.6
Secondo trim.	1.0	0.3	1.4	0.7	2.1	2.0	2.8	4.2
Terzo trim.	1.0	0.3	1.4	0.8	1.8	1.9	2.3	4.2
Quarto trim.	1.6	0.6	1.8	1.0	2.4	3.5	2.6	4.0

(a) I dati della seconda, terza e quarta settimana sono cumulati.

(b) Il tasso di caduta per errori di lista indica la percentuale di famiglie cadute con esito=23 o esito=16-21 sul totale delle famiglie assegnate monitorate.

(c) Il tasso di caduta per irreperibilità indica la percentuale di famiglie cadute con esito=22 o con esito=13-15 sul totale delle famiglie assegnate monitorate.

6.3 Il ricorso a modalità di intervista proxy

Monitorare il ricorso alle interviste proxy consente di elevare gli standard di qualità delle rilevazioni. Da precedenti indagini sulle condizioni di salute e ricorso ai servizi sanitari, è stato osservato ad esempio, che le risposte proxy sul consumo di tabacco e sulla presenza di malattie croniche possono generare distorsioni sulle stime.

L'indagine prevede l'uso di due questionari da sottoporre a tutti i componenti la famiglia di fatto: uno per intervista diretta, l'altro per autocompilazione. Per il primo questionario la risposta proxy è accettata mentre per il secondo la compilazione deve essere strettamente personale. Si accettano risposte proxy da un familiare solo per le persone di età inferiore a 14 anni e per i componenti assenti durante il periodo di rilevazione o non in grado di rispondere per gravi problemi di salute.

Passando dal primo trimestre ai successivi si osserva una sensibile diminuzione della quota di compilazione proxy: il *tasso di proxy per le schede individuali*⁸ passa dal 34% al 31%, il *tasso di proxy per i questionari autocompilati*⁹ passa da 28% al 25% (Tavola 6).

7 Il tasso di caduta per errore di lista indica la percentuale di famiglie cadute perchè non eleggibili (esito=20, esito=23, esito=16-18) oppure trasferite in altro comune italiano (esito=19) sul totale delle famiglie assegnate monitorate.

8 Il tasso di proxy per le schede individuali si calcola rapportando il numero di schede individuali compilate proxy al numero di componenti la famiglia di fatto.

9 Tasso di proxy per questionari autocompilati si calcola rapportando il numero di questionari autocompilati proxy al numero totale di questionari autocompilati per 100.

Tavola 6 - Principali indicatori di qualità dell'indagine per tipo di comune campione, trimestre e settimana di rilevazione(a).

	Prima settimana	Seconda settimana	Terza settimana	Quarta settimana
1° trimestre				
Tasso di proxy per le schede individuali (b)	27.6	30.7	32.6	33.8
Tasso di proxy per questionari autocompilati(c)	26.8	27.4	27.6	28.3
2° trimestre				
Tasso di proxy per le schede individuali (b)	30.2	30.5	30.9	31.0
Tasso di proxy per questionari autocompilati(c)	26.9	26.1	26.2	25.7
3° trimestre				
Tasso di proxy per le schede individuali (b)	26.9	29.7	30.5	31.0
Tasso di proxy per questionari autocompilati(c)	21.8	24.0	25.1	25.5
4° trimestre				
Tasso di proxy per le schede individuali (b)	29,1	30.2	30.5	30.8
Tasso di proxy per questionari autocompilati(c)	24.3	25.0	25.1	25.3

(a) I dati della seconda, terza e quarta settimana sono cumulati.

(b) Numero di schede individuali proxy rapportato al numero di componenti la famiglia di fatto per 100.

(c) Numero di questionari autocompilati proxy rapportato al numero totale di questionari autocompilati per 100.

6.4 Gli interventi alla fine di ciascun trimestre di rilevazione

Alla fine di ciascun trimestre è stato possibile avere un quadro completo sull'andamento della rilevazione per tutti i comuni ed i rilevatori impegnati nell'indagine inclusi anche quelli che non avevano partecipato al monitoraggio.

Tra i rilevatori dei comuni che hanno partecipato ai quattro trimestri di rilevazione è stato possibile ripetere la formazione laddove questi avevano incontrato difficoltà nell'utilizzo dell'IVR o nell'effettuare le interviste secondo le indicazioni metodologiche previste. La formazione dei rilevatori sull'uso dell'IVR, sui contenuti dell'indagine e sugli adempimenti da seguire durante la rilevazione è stata curata dai responsabili regionali dell'indagine in modo omogeneo a livello territoriale.

Nel corso della rilevazione si è proceduto anche alla sostituzione dei rilevatori con persistente scorrettezza riguardo soprattutto al fenomeno delle interviste false o effettuate per telefono.

7 - Gli effetti del monitoraggio sulla qualità dell'indagine

7.1 Influenza del monitoraggio sui rilevatori

Dall'analisi degli indicatori di qualità, calcolati per tutti i rilevatori, emerge che quelli che hanno utilizzato il sistema IVR, consentendo quindi di essere monitorati, hanno lavorato seguendo più scrupolosamente le indicazioni metodologiche.

Si evidenzia innanzitutto, che il *tasso di caduta* (14,9%) tra gli intervistatori monitorati è la metà rispetto a quello osservato tra gli intervistatori che non hanno partecipato al monitoraggio (32,1%) (Tavola 7).

Tavola 7 - Tassi di risposta, rifiuto e caduta per partecipazione dei rilevatori al monitoraggio IVR.

Trimestre di rilevazione	Partecipazione dei rilevatori al monitoraggio					
	Sì			No		
	Tasso di risposta (a)	Tasso di rifiuto (b)	Tasso di caduta (c)	Tasso di risposta (a)	Tasso di rifiuto (b)	Tasso di caduta (c)
Primo Trimestre	90.1	6.5	14.8	75.5	14.2	36.6
Secondo Trimestre	88.4	7.9	16.3	81.9	10.3	35.3
Terzo Trimestre	89.3	6.8	15.4	82.2	9.4	27.9
Quarto Trimestre	91.1	5.6	13.2	87.8	7.0	26.3
Totale	89.7	6.7	14.9	81.4	10.4	32.1

(a) Il tasso di risposta è calcolato come numero di famiglie con intervista completa su numero di famiglie eleggibili per 100

(b) Il tasso di rifiuto indica la percentuale di famiglie cadute per rifiuto (esito=12 o esito=24-34) sul totale delle famiglie intervistate (esito=11), cadute per interruzione (esito=12) e per rifiuto (esito=24-34).

(c) Il tasso di caduta indica la percentuale di famiglie cadute (esito=12-34) sul totale delle famiglie campione.

Nel corso dei trimestri è stata quindi prestata maggiore attenzione al lavoro degli intervistatori che non avevano aderito all'uso dell'IVR. Il controllo accurato della loro attività e la rinnovata formazione hanno consentito di migliorare le loro prestazioni, anche se sono rimaste le notevoli differenze rispetto ai rilevatori che hanno partecipato al monitoraggio: il loro tasso di caduta è sensibilmente diminuito passando dal 36,6% del primo trimestre al 26,3% del quarto (Tavola 7).

Alcune differenze tra i rilevatori si osservano anche riguardo alla compilazione proxy o non proxy del questionario autocompilato: la quota di interviste proxy è 18,5% tra i monitorati contro il 23,7% tra i non monitorati.

Durante e dopo la rilevazione sul campo è stata controllata la regolarità delle interviste contattando per telefono due famiglie per ciascun rilevatore (in caso di irregolarità sono state contattate tutte le famiglie intervistate). Circa il 5% delle interviste controllate sono risultate irregolari perchè false o effettuate per telefono. La quota di intervistatori che ha effettuato almeno un'intervista irregolare è superiore tra quelli che non hanno aderito al monitoraggio (8,4% contro 5,1% tra rilevatori partecipanti al monitoraggio). I rilevatori che hanno violato le regole impartite dall'Istat inventando le interviste non potranno più svolgere l'attività di intervistatore per le successive indagini sociali effettuate dai Comuni.

7.2 Confronto tra indagini

Al fine di apprezzare se l'adozione del sistema IVR genera un guadagno nella qualità della rilevazione, si confrontano due edizioni dell'indagine Aspetti della vita quotidiana: quella del 2006 che ha recepito il sistema IVR per la comunicazione degli esiti dei contatti e quella del 2005 che ha utilizzato solo la scheda di monitoraggio cartacea registrata a fine rilevazione. Passando da un'edizione all'altra si osservano dei miglioramenti, infatti, il *tasso di mancata risposta totale* passa dal 20,5% al 18,7% e il *tasso di risposta* dal 79,5% sale all'81,3% (Tavola 8).

Si osserva che, a parità di utilizzo del sistema IVR, gli indicatori di qualità dell'indagine "Condizione di salute e ricorso ai servizi sanitari 2004-2005" sono migliori rispetto a quelli dell'indagine

“Aspetti della vita quotidiana 2006”. Questo guadagno aggiuntivo di qualità, probabilmente, deriva dalla diversa modalità di esecuzione delle due indagini: una sola rilevazione nell’arco dell’anno per Aspetti della vita quotidiana, quattro rilevazioni a cadenza trimestrale per Condizioni di salute e ricorso ai servizi sanitari. In quest’ultimo caso infatti, si ha la possibilità di intervenire alla fine di ciascun trimestre sui rilevatori dei comuni che ripetono l’indagine nell’anno. Inoltre, non è da escludere il maggiore coinvolgimento dei cittadini verso i temi della salute e dei servizi sanitari rispetto alle tematiche più generiche dell’indagine Aspetti della vita quotidiana.

Tavola 8 - Indicatori di qualità sulla rilevazione previsti nel sistema SIDI per due indagini Multiscopo

	Tasso di errori di lista (a)	Tasso di mancata risposta totale (b)	Tasso di risposta (c)	Tasso di rifiuto (d)	Tasso di Mancato contatto (e)	Tasso di Mancata risposta per altri motivi (e)
Condizioni di salute e ricorso ai servizi sanitari 2004-2005	1.6	16.2	83.8	4.8	7.7	1.5
Aspetti della vita quotidiana 2005 (monitorata con scheda cartacea)	1.2	20.5	79.5	5.3	5.4	5.0
Aspetti della vita quotidiana 2006 (monitorata con IVR)	2.0	18.7	81.3	6.0	7.3	3.2

a) tasso di errori di lista è calcolato come (unità non eleggibili+non contatti dovuti ad errori di lista)/unità risolte per 100.

b) Il Tasso di mancata risposta totale è calcolato come (unità non rispondenti+unità non risolte)/ (unità eleggibili+unità non risolte) per 100.

c) Il Tasso di risposta è calcolato come unità rispondenti/ (unità eleggibili+unità non risolte) per 100.

d) Il Tasso di rifiuto è calcolato come unità che si rifiutano/ unità eleggibili per 100.

e) Il Tasso di mancato contatto è calcolato come unità non contattate/ unità eleggibili per 100.

f) Il Tasso di mancata risposta per altri motivi è calcolato come unità non contattate per motivi diversi da rifiuto e mancato contatto/ unità eleggibili per 100.

8 - Altri indicatori che potrebbero entrare nel sistema SIDI

Si potrebbero affiancare agli indicatori sulla rilevazione, già presenti nel sistema SIDI, quelli per la valutazione delle modalità di rilevazione, come ad esempio il rispetto dei tempi di rilevazione oppure la percentuale di interviste proxy sui rispondenti e quelli sui controlli di qualità, effettuati durante o dopo la rilevazione, volti ad individuare la corretta applicazione delle norme metodologiche per le interviste.

Inoltre si potrebbero inserire degli indicatori da utilizzare per l’analisi delle fasi precedenti alla rilevazione sul campo che pure sono affidate ai Comuni e quindi non direttamente sotto il controllo dell’Istat, come ad esempio la fase di estrazione delle famiglie campione. Con riferimento a quest’ultima, il monitoraggio della tempestività della spedizione del file delle famiglie campione, tramite un indicatore ad Hoc, permette di valutare le cadute dovute ai comuni campione che non trasmettono la lista entro la data indicata dall’Istat.

Un’altra fase precedente alla rilevazione sul campo, che merita di essere monitorata, è la formazione dei rilevatori comunali sui contenuti dell’indagine e sulle modalità di esecuzione delle interviste. Opportuni indicatori sono necessari per monitorare la presenza degli interessati all’intera riunione di formazione in quanto l’acquisizione di tutte le informazioni è propedeutica per una corretta esecuzione dell’indagine.

Appendice – Elenco degli esiti dei contatti

ESITI PROVVISORI	
Occupato / Fax (solo contatto telefonico)	1
Segreteria telefonica (solo contatto telefonico)	2
Non c'è nessuno / non risponde / utente non raggiungibile (sia contatto telefonico che a domicilio)	3
Persona che non può fissare un appuntamento (sia contatto telefonico che a domicilio)	4
Appuntamento prima di iniziare l'intervista (sia contatto telefonico che a domicilio)	5
Appuntamento dopo aver iniziato l'intervista o per ritirare l'autocompilato (sia contatto telefonico che a domicilio)	6
L'Intestatario della Scheda di Famiglia (ISF) e / o il coniuge / convivente non vivono più presso l'abitazione, ma vivono nel Comune e	7
Né l'ISF, né il coniuge / convivente vivono presso l'abitazione e non si sa dove vivono (sia contatto telefonico che a domicilio)	8
Nome dell'ISF inesistente, non si sa nulla della famiglia (sia contatto telefonico che a domicilio)	9
Indirizzo / numero civico inesistente o impossibile da localizzare (solo a domicilio)	10
ESITI DEFINITIVI	
UNITA' NON RISOLTE	
Il Comune non è in grado di fornire elementi ulteriori per reperire la famiglia (soltanto dopo esiti provvisori 8, 9, 10) (solo a domicilio)	21
Motivo sconosciuto (esito fittizio utilizzato in due casi 1) famiglie per le quali non è stato possibile recuperare il motivo della non risposta 2) famiglie di comuni caduti dopo l'estrazione delle famiglie)	37
UNITA' RISOLTE	
UNITA' NON ELEGGIBILI	
Non più esistenti	
Decesso dell'unico o di tutti i componenti (solo a domicilio)	16
Con variazioni di stato	
La famiglia vive all'estero (sia contatto telefonico che a domicilio)	20
Fuori target	
Seconda casa, in cui si trascorre meno della metà dell'anno (solo a domicilio)	17
La famiglia vive permanentemente in istituzione/convivenza (solo a domicilio)	18
All'indirizzo non corrisponde una famiglia, ma un negozio, uno studio o un insieme di persone che non costituiscono famiglia (solo a domicilio)	23
UNITA' ELEGGIBILI	
RISPONDENTI	
Intervista completata (solo a domicilio)	11
NON RISPONDENTI	
Rifiuto	
Interruzione definitiva dell'intervista (solo a domicilio)	12
Rifiuto per mancanza di tempo, momento poco opportuno (solo a domicilio)	26
Rifiuto per diffidenza/paura (solo a domicilio)	29
Rifiuto per sfiducia nello Stato/Istituzioni/ISTAT (solo a domicilio)	30
Rifiuto per opinioni negative nelle indagini statistiche (solo a domicilio)	31
Rifiuto per assenza di interesse per l'argomento (solo a domicilio)	32
Rifiuto per difesa ad oltranza della privacy (solo a domicilio)	33
Rifiuto per indisponibilità assoluta, non dà spiegazioni (solo a domicilio)	34
Mancato contatto	
Errori di lista	
La famiglia vive in altro Comune italiano (solo a domicilio)	19
Altri motivi	
Famiglia poco presente in casa: effettuati tutti i contatti previsti senza essere però riusciti a parlare con qualcuno della famiglia (sia contatto telefonico che a domicilio)	13
Famiglia poco presente in casa: effettuati tutti gli appuntamenti previsti senza essere riusciti ad effettuare/concludere l'intervista (solo a domicilio)	14
Famiglia assente durante l'intero periodo di rilevazione (sia contatto telefonico che a domicilio)	15
Altri motivi di non risposta	
Impossibilità di accedere all'indirizzo (solo a domicilio)	22
Rifiuto per malattia grave (es. disabilità mentale) (solo a domicilio)	24
Rifiuto per problemi familiari/personali (ad es. lutto), esclusa malattia grave (solo a domicilio)	25
Rifiuto per stranieri che non parlano italiano (solo a domicilio)	27
Rifiuto per problemi di comprensione/scarsa cultura/dialetto (solo a domicilio)	28

Bibliografia

BRANCATO G., CARBINI R., PELLEGRINI C., SIGNORE M., SIMEONI G., *Assessing Quality through the Collection and Analysis of Standard Quality Indicators: The ISTAT Experience*, European Conference on Quality in Survey Statistics Q2006

BRANCATO G., PELLEGRINI C., SIGNORE M., SIMEONI G., *Standardising, Evaluating and documenting Quality: the implementation of Istat Information System for Survey Documentation – SIDI*, European Conference on Quality and Methodology in Official Statistics. Q2004 Mainz, 24-26 May 2004. CD ROM

ISTAT 2007, a cura di Luciana Quattrociocchi, *Come si progetta il lavoro del monitoraggio sul campo di una indagine sulle famiglie*, Metodi e Norme, in corso di pubblicazione

ISTAT 2006, a cura di Gianlorenzo Baratta, *Il sistema di indagini multiscopo – Contenuto e metodologia nelle indagini*, Metodi e norme n. 31.

ISTAT 2003, a cura di Gianlorenzo Baratta e Monica Perez, *Metodologia e organizzazione dell'indagine multiscopo sulla domanda turistica "Viaggi e vacanze"*, Metodi e norme, n. 17

ISTAT 2001, *Indagini sociali telefoniche - Metodologia ed esperienze della statistica ufficiale* Metodi e norme, n. 10

MONTECOLLE S. (2007) , *Gli effetti delle innovazioni di processo introdotte nell'indagine aspetti della vita quotidiana sugli indicatori di copertura e mancata risposta totale*. In "Atti del Seminario sulla Qualità: l'esperienza dei referenti del sistema informativo SIDI – I giornata", Istat, Contributi, in corso di pubblicazione. (DA CONTROLLARE)

Statistics Canada quality guidelines, third edition, October 1998

Il sistema SIDI: uno strumento per le analisi di qualità

Giovanna Brancato, Riccardo Carbini, Concetta Pellegrini, Marina Signore, Giorgia Simeoni

Istat, DCMT/PAQ

1. Introduzione

Il presente lavoro illustra, attraverso alcuni esempi, le potenzialità informative di SIDI (Sistema Informativo di Documentazione delle Indagini). Queste sono rappresentate, in primo luogo, da interrogazioni ed analisi disponibili attraverso i due sistemi di interrogazione: SIDI-TOP, per gli utenti interni all'Istituto, e SIQual, per gli utenti del sito web dell'Istat (Pellegrini, 2006). Inoltre, sono possibili analisi *ad hoc* sulla qualità del complesso dei processi produttivi, utili a supportare i processi decisionali e le politiche dell'Istituto.

La possibilità di condurre analisi valide e significative a partire da SIDI dipende fortemente dal livello di qualità e completezza della documentazione presente nel sistema. Gli sforzi che si sono compiuti negli anni passati hanno consentito oggi di poter disporre di una base di metadati e indicatori già sufficiente per esplorare numerosi aspetti della qualità. E' importante che l'impegno dei referenti per la qualità e documentazione, che si occupano dell'aggiornamento dei metadati e degli indicatori, continui nei prossimi anni; così come la costante attività di monitoraggio da parte del gruppo SIDI, che assicura la coerenza e la validità delle informazioni immesse nel sistema. Tutto ciò rappresenta un requisito indispensabile per poter ampliare sempre di più la base informativa sulla qualità e rispondere a future nuove esigenze informative.

2. Indicatori e metadati per le analisi di qualità

Il sistema SIDI gestisce in modo integrato informazioni qualitative (**metadati**) e informazioni quantitative (**indicatori standard di qualità**) per i processi statistici dell'Istituto (rilevazioni ed elaborazioni¹) (Brancato et al., 2004a). Il buon grado di popolamento del sistema, oggi raggiunto, fa sì che possa essere proficuamente utilizzato per supportare le analisi di qualità e le decisioni strategiche dell'Istituto (Brancato et al., 2006b). Ciò può essere fatto utilizzando il patrimonio informativo gestito nel sistema in diversi modi. In primo luogo, è possibile utilizzare le funzionalità standard offerte da SIDI-TOP (sistema di interrogazione e di navigazione di SIDI accessibile dalla Intranet) e da SIQual (sistema informativo sulla qualità sul sito Web dell'Istituto). Il gruppo responsabile di SIDI produce inoltre report periodici e approfondimenti per i vertici dell'Istituto (Brancato *et al.*, 2006a). Infine, per quanto riguarda in particolare le analisi quantitative, è possibile esportare gli indicatori di qualità in excel per effettuare ulteriori analisi.

I metadati gestiti nel sistema, per le rilevazioni, sono tematicamente raggruppati come segue: aspetti generali del processo; disegno di indagine e unità statistiche; operazioni e azioni di controllo del processo; modelli di rilevazione; ed, infine, supporti di diffusione.

¹ Per rilevazione si intende un processo finalizzato alla produzione di informazioni statistiche da parte dell'ente titolare, consistente nella raccolta di dati direttamente presso imprese, istituzioni e persone fisiche oppure attraverso documenti amministrativi e/o fonti organizzate pubbliche e private (registri, archivi, basi di dati) e nel loro successivo trattamento. Per elaborazione si intende un processo finalizzato alla produzione di informazioni statistiche da parte dell'ente titolare, consistente nel trattamento di dati statistici derivanti da precedenti rilevazioni od elaborazioni dello stesso o di altri soggetti, ovvero di dati di cui l'ente dispone in ragione della sua attività istituzionale.

Gli **aspetti generali** contengono:

- dati di sintesi del processo, quali: una descrizione dell'indagine; lo stato del processo (attivo, sospeso, cessato); la direzione, il servizio di appartenenza e il responsabile; la periodicità; la tipologia (diretta, amministrativa, mista); informazioni sull'atto amministrativo (se pertinenti); i riferimenti alla normativa che regola il processo e manuali standard (se presenti); eventuali informazioni aggiuntive.
- i metadati relativi all'area e settore di appartenenza e ai fenomeni osservati.

Nell'area **disegno di indagine** e **unità statistiche** sono rappresentate le informazioni relative a:

- il disegno di campionamento,
- le unità di rilevazione e gli archivi (o liste) da cui sono estratte;
- le diverse tipologie di unità di analisi (osservate individualmente, di sommario o subset);
- gli archivi aggiornati dal processo.

Un'altra area è dedicata alla descrizione del **processo produttivo**, cioè:

- le singole operazioni svolte nelle varie fasi in cui si articola il processo
- le azioni di controllo poste in essere per prevenire, monitorare in corso d'opera e valutare a posteriori l'impatto degli errori associati alle operazioni di indagine.

Nella parte relativa ai **modelli di rilevazione** è possibile descrivere i questionari utilizzati per raccogliere i dati e di associarne le immagini, mentre nell'area **supporti di diffusione** vengono documentate le diverse forme utilizzate per diffondere i dati all'esterno (banche dati, pubblicazioni cartacee, volumi on-line,...).

I metadati gestiti per le elaborazioni si differenziano per la presenza della descrizione delle fonti di input utilizzate per l'elaborazione: nome della fonte, ente di appartenenza, periodicità, unità statistiche utilizzate, tipologia di dati utilizzati, giudizio qualitativo sull'adeguatezza della fonte; e ovviamente per l'assenza dei modelli di rilevazione.

In generale la descrizione qualitativa avviene attraverso la selezione di voci standard da appositi tesauri di voci (Brancato et al., 2004b). Tuttavia la maggior parte delle informazioni sopra menzionate possono essere dettagliate attraverso l'inserimento di descrizioni testuali libere. Inoltre quasi tutte le tipologie di metadati gestiti nel sistema hanno associato un periodo di validità che consente di prendere visione sia della situazione corrente che di quella pregressa.

Gli indicatori misurano la qualità del processo e del prodotto. Sono stati definiti i seguenti gruppi di indicatori standard: **copertura** (errori della lista o dell'archivio di estrazione delle unità di rilevazione); **mancata risposta** (errori di mancata risposta totale e relativi motivi); **codifica** (errori nella codifica di una variabile rilevata in modo testuale e codificata con procedure automatiche o semiautomatiche); **controllo e correzione** (errori rilevati e corretti dalla procedura automatica o semiautomatica di controllo e correzione per il processo); **controllo e correzione per variabile** (impatto delle procedure di controllo e correzione a livello di singola variabile); **politica di revisione** (descrizione della politica secondo la quale sono revisionati i dati finali e misure dell'errore associato); **tempestività** (tempo che intercorre tra il rilascio dei dati all'utente ed il periodo di riferimento dei dati); **confrontabilità nel tempo** (indicatore della lunghezza di serie storiche confrontabili per una determinata variabile); **coerenza tra dati provvisori e definitivi** (errore che si è commesso nel rilascio anticipato dei dati); **coerenza con fonti esterne** (indicatore di processo applicabile ad alcune indagini di tipo amministrativo per le quali si dispone di dati individuali relativi ad eventi e di riepiloghi o conteggi aggregati indipendenti).

Gli indicatori di copertura, mancata risposta, controllo e correzione, controllo e correzione per variabile possono essere calcolati in modo diretto (non ponderato) nel qual caso danno una informazione sulla qualità della fase oppure in modo ponderato utilizzando ad esempio, nel caso di copertura e mancata risposta, i pesi campionari. Gli indicatori ponderati forniscono una valutazione dell'impatto dell'errore riferito alla popolazione osservata (ad esempio un tasso di risposta non ponderato pari al 80% ci dice la

percentuale di campione che ha risposto e potrebbe corrispondere ad un tasso di risposta ponderato superiore al 80% se le unità che hanno risposto hanno un peso maggiore).

In generale per ciascun gruppo è definito un set di indicatori standard che consente di analizzare in dettaglio quel determinato aspetto. Gli indicatori sono calcolati per ciascuna replicazione del processo produttivo. Per ogni gruppo di indicatori viene individuato un insieme di metadati rilevanti, da selezionare tra quelli già documentati, e si possono inserire delle note testuali. I metadati rilevanti consentono di esplicitare le modalità operative dell'operazione a cui si riferiscono gli indicatori (ad esempio l'archivio di estrazione per il gruppo copertura, la tecnica di rilevazione prevalente per il gruppo mancata risposta totale, la periodicità per la tempestività, la variabile per l'indicatore di coerenza tra dati provvisori e definitivi).

Oltre ai metadati e agli indicatori vengono gestiti nell'area **documenti**: i documenti metodologici di tipo generale o relativi ad uno o più processi, i documenti operativi relativi all'esecuzione di una o più fasi del processo (ad esempio le circolari o i manuali di istruzione); i documenti Eurostat relativi alla qualità; i regolamenti europei e i relativi processi che regolamentano; i report sulla qualità prodotti dai processi (ad esempio report per Eurostat); i report su SIDI relativi al popolamento del sistema e ad analisi descrittive sulla qualità; i questionari di indagine per accedere direttamente alle caratteristiche e all'immagine del questionario; ed infine le note ASI cioè le note metodologiche standard dei processi che hanno contribuito all'Annuario Statistico Italiano.

Con riferimento alle funzionalità di analisi, si sottolinea che SIQual è dotato di strumenti di ricerca avanzati per l'analisi dei metadati relativi ai processi produttivi² ed in particolare di un datawarehouse di metadati. Attraverso tali funzionalità è possibile esplorare i metadati delle rilevazioni e delle elaborazioni impostando criteri di ricerca di interesse e ottenendo tavole personalizzate dove nelle celle è riportato il numero di processi che soddisfa le richieste fatte. Ad esempio, è possibile analizzare la qualità delle operazioni svolte dalle diverse indagini dell'Istituto, incrociando opportunamente operazioni e azioni di controllo come presentato nel paragrafo 4.

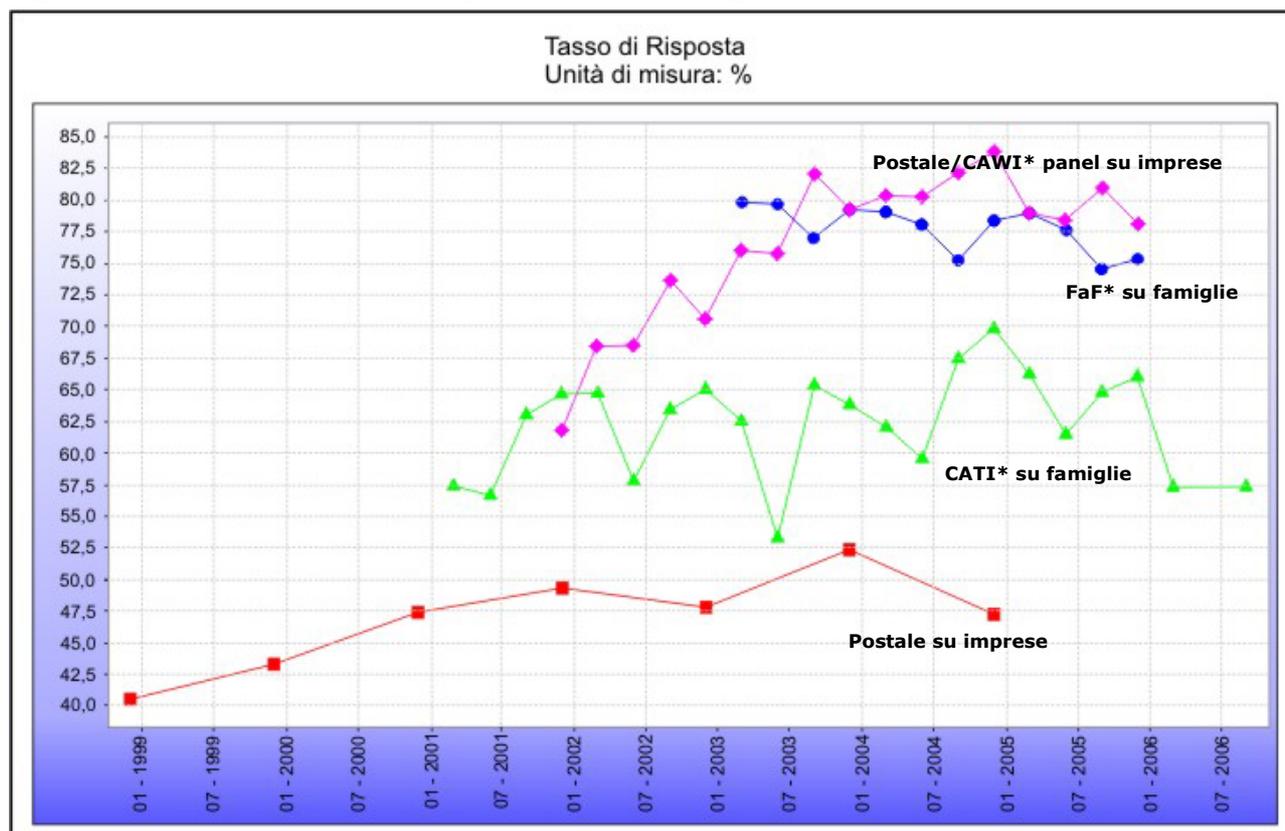
Le funzionalità standard offerte da SIDI-TOP consentono di analizzare congiuntamente i metadati e gli indicatori di qualità, aspetto importante perché consente di contestualizzare gli indicatori stessi, sia quando sono riferiti all'analisi di una singola indagine (ad esempio analisi dell'andamento nel tempo di un set di indicatori), sia quando si confrontano indagini diverse, sia, ancora, quando si vogliono effettuare confronti di una o più indagini con valori medi generali o specifici³. Esempi di analisi che possono essere effettuate a partire dagli indicatori attualmente gestiti nel sistema sono presentate nel paragrafo 5.

Si riporta in figura 2.1 un esempio di analisi standard offerta dal sistema nella quale sono rappresentati e confrontati graficamente i tassi di risposta di quattro diverse indagini selezionate perché adottavano ciascuna una diversa tecnica di raccolta dei dati. Come si vede dal grafico, l'indagine multitecnica presenta tassi di risposta migliori delle altre ed un trend positivo, mentre come atteso, i tassi di risposta più bassi si ottengono per la tecnica postale.

² SIQual attualmente non diffonde indicatori di qualità. L'Istituto sta infatti predisponendo una politica per comunicare la qualità agli utenti esterni a seguito della quale verrà avviata la diffusione di Schede Qualità contenenti indicatori quantitativi agli utenti esterni. Le Schede Qualità saranno diffuse attraverso SIQual.

³ Valori medi generali sono i valori medi di un determinato indicatore calcolati con riferimento a tutte le indagini dell'Istituto che hanno immesso dati nel sistema per quell'indicatore, mentre valori medi specifici si riferiscono a valori medi calcolati su sottogruppi di indagini, tra quelle che hanno valori dell'indicatori, che presentano caratteristiche comuni. Esempi di valori medi specifici sono: tempestività medie delle indagini mensili e tempestività medie delle indagini annuali; tassi medi di risposta delle indagini CATI e tassi medi di risposta delle indagini postali.

Figura 2.1. Confronto tra tassi di risposta di diverse indagini



*FaF: Faccia a faccia; CAWI: Computer Assisted Web Interviewing; CATI: Computer Assisted Telephone Interviewing

3. Strumenti per la reportistica e stato di popolamento del sistema

SIDI è stato progettato come uno strumento di supporto al controllo di qualità dei processi produttivi dell'Istituto e, come tale, consente di soddisfare diversi obiettivi:

- facilitare il responsabile di indagine nel monitoraggio e nella documentazione della qualità;
- confrontare la qualità di indagini diverse tenendo conto di fattori rilevanti comuni;
- fornire ai vertici dell'Istituto informazioni periodiche sulla qualità dei processi dell'Istituto.

Il sistema, nel tempo, ha modificato ed ampliato le sue finalità ed i suoi obiettivi ed è riconosciuto-come un valido e completo strumento per l'analisi ed il confronto dei processi statistici dell'Istituto (Cigas, 2006). Questa maggiore attenzione nella documentazione del sistema è stata raggiunta grazie ad una strategia articolata.

All'inizio vi era il fattore dell'"obbligatorietà": ogni anno alcune caratteristiche del sistema venivano poste in direttiva⁴ per il popolamento e gli obiettivi previsti venivano, ovviamente, raggiunti, anche se spesso con molta fatica da parte del gruppo centralizzato responsabile del sistema. In seguito hanno svolto un ruolo importante il crescente uso interno delle informazioni gestite in SIDI e la diffusione di informazione all'esterno tramite SIQual. Infine, grande importanza hanno avuto gli strumenti di reportistica interna.

Con la realizzazione di SIDI-TOP è stata sviluppata una funzione che permette di ottenere report dinamici sullo stato di popolamento del singolo processo o di gruppi di processi d'interesse al momento dell'esecuzione. Questa funzionalità è stata sviluppata per soddisfare la necessità dei responsabili di

⁴ Le direttive sono indicazioni strategiche emanate dal Consiglio d'Istituto la cui attuazione è monitorata dall'Ufficio Valutazione e Controllo Strategico.

indagine o dei vertici dell'Istituto di conoscere lo stato di popolamento dei processi di interesse. Oltre a ciò, due volte l'anno, vengono realizzati e diffusi sempre attraverso il sistema SIDI-TOP dei report di approfondimento sullo stato di popolamento del sistema.

Questi report mettono in rilievo i cambiamenti avvenuti rispetto alla situazione precedente. Oltre ad un report generale vengono realizzati dei report specifici per ogni Direzione produttiva con informazioni di dettaglio relative ai processi statistici svolti al loro interno. Questa attività di reportistica aiuta a motivare l'aggiornamento ed a migliorare la qualità dell'informazione documentata nel sistema.

Molto importante è all'interno del sistema la possibilità di fornire una documentazione sempre aggiornata e tempestiva sullo stato di popolamento e sugli incrementi del sistema sia riguardo ai metadati che agli indicatori di qualità.

Fino a maggio 2005 i report che venivano diffusi richiedevano delle elaborazioni ad hoc spesso molto complesse per realizzare gli incroci richiesti nelle tavole e non sempre era possibile mettere in evidenza la componente temporale (ad es.: in che data l'informazione veniva aggiornata nel sistema).

A tale proposito è stata realizzata una *stored procedure pl/sql* che permette di memorizzare in apposite tabelle del database (DB), alla data di esecuzione, tutte le principali caratteristiche dei singoli processi, in alcuni casi come frequenze, in altri nel loro dettaglio. Quindi è come se al momento dell'esecuzione della procedura venisse effettuata una "fotografia" delle principali informazioni qualitative e quantitative presenti, ad una specifica data, nel sistema. Questo fa sì che mettendo a confronto le informazioni memorizzate, per le diverse occasioni del monitoraggio è possibile ricostruire, non solo tutti gli aggiornamenti avvenuti nel tempo nel sistema, ma anche gli incrementi nel popolamento.

Qui di seguito vengono mostrati due esempi di tavole realizzate a partire dall'informazione memorizzata nel DB attraverso tale procedura. Nella tabella 3.1 vengono mostrati gli aggiornamenti in merito al calcolo degli indicatori, mentre nella tabella 3.2 vengono mostrati gli aggiornamenti a livello di principali metadati documentati nel sistema.

Tabella 3.1. Numero di occasioni o di variabili per occasione di indagine per le quali sono stati calcolati gruppi di indicatori di qualità

Gruppi di indicatori di qualità	Processi eleggibili attivi	Occasioni o variabili per occasioni calcolate al*		
		06/05/2005	19/12/2005	19/12/2006
Copertura e Mancata risposta	122	81 (25)	143 (33)	295 (45)
Controllo e correzione	105	34 (15)	98 (18)	151 (22)
Tempestività e puntualità	205	49 (20)	775 (149)	1424 (196)
Risorse	253	19 (12)	34 (16)	95 (60)
Coerenza con fonti esterne**	42			8 (6)
Coerenza dati provvisori e definitive**	36			125 (12)
Controllo e correzione per variabile**	105			7 (1)
Politica di revisione dei dati finali**	21			36 (3)
Confrontabilità**	192			143 (49)

*Tra parentesi è riportato il numero di processi che hanno calcolato indicatori per il gruppo

** Le funzioni per la gestione di questi gruppi di indicatori sono state rilasciate a giugno del 2006

In riferimento alla tabella 3.1 va comunque considerato che non tutti i processi documentati nel sistema possono calcolare tutti i gruppi di indicatori, a tal fine è stata riportata l'informazione relativa al numero di processi eleggibili per gruppo. Viene inoltre riportato lo stato di popolamento del sistema, a tre diverse date, espresso come numero di occasioni di indagine per cui sono stati calcolati gli indicatori. Infatti per uno stesso processo può essere stata calcolata una serie storica per gruppo di indicatori per più edizioni di indagine. Il numero dei differenti processi che hanno calcolato gli indicatori è riportato in parentesi.

Tabella 3.2 – Numero dei differenti metadati documentati nel sistema per le diverse occasioni del monitoraggio

Data del monitoraggio	Processi documentati	Numero di elementi per tipologia					
		Unità di analisi	Fenomeni osservati	Disegno di indagine	Operazioni	Azioni di controllo	Questionari
06/05/2005	261	655	823	147	4033	2027	53
19/12/2005	269	714	851	172	4536	2216	127
19/12/2006	287	776	894	184	4873	2314	172

Oggi con la messa in esercizio sul portale d'Istituto del sistema SIQual, sistema realizzato sfruttando il patrimonio informativo contenuto in SIDI, c'è una maggiore attenzione nel fornire i metadati e una maggiore consapevolezza dell'importanza di diffondere all'utenza esterna un'informazione sempre più dettagliata e corretta anche a livello di metadati.

Nuove funzioni sviluppate consentono di inserire nel sistema diverse tipologie di documenti sulla qualità da diffondere sul sito, attraverso il sistema SIQual, a corredo della documentazione di indagine e come informazioni di carattere generale.

Di seguito viene riportata una tabella con lo stato di popolamento delle diverse tipologie di documenti.

Tabella 3.3 – Numero di documenti consultabili dal sistema distinti per tipologia

Tipo documento	Numero
Immagine questionario	172
Regolamenti comunitari	39
Documenti Eurostat	12
Report approfonditi sulla qualità	24
Documenti metodologici	8
Documenti operativi	12
Note metodologiche ASI	337

Come precedentemente detto nel paragrafo 2, per poter standardizzare le informazioni presenti nel sistema ci si è avvalsi dell'utilizzo di tesauri che sono degli elenchi di voci standard adottati per la

documentazione delle diverse caratteristiche del processo statistico. Questo da un lato aveva permesso di documentare in modo standard il sistema ma lo rendeva, in alcuni casi, troppo sintetico mentre sempre più spesso i referenti avevano l'esigenza di dettagliare in modo più preciso l'informazione fornita. Per ovviare a questo inconveniente nel sistema sono state introdotte delle modifiche per permettere di corredare tutti gli elementi presenti nel sistema con informazioni descrittive.

Oggi nel sistema ogni processo statistico ha una sua descrizione, più o meno analitica, che ne descrive gli obiettivi o le caratteristiche, ed anche per le altre tipologie di metadati sono state inserite note e descrizioni che permettono di descrivere con maggior dettaglio il processo stesso, qui di seguito viene riportata una tabella che illustra il livello di popolamento di tali descrizioni.

Tabella 3.4 – Numero di descrizioni o note per tipologia di elemento collegato

Tipologia	Numero
Processo statistico	235
Questionario	151
Fenomeni osservati	36
Unità d'analisi	57
Operazioni	38
Supporti di diffusione	44
Manuali standard	5
Serie storiche	41
Disegno di campionamento	55
Note per gruppo di Indicatori standard di qualità	134

4. Analisi di qualità basata sui metadati presenti nel sistema

La filosofia del sistema SIDI si basa sull'ipotesi che per la corretta interpretazione della qualità di una rilevazione, gli indicatori quantitativi debbano essere affiancati da informazioni di tipo qualitativo sul processo, e cioè da quelle che in SIDI sono rappresentate come *operazioni*.

Spesso l'analisi dei soli indicatori standard, anche se presentati con l'informazione sulle caratteristiche rilevanti del processo, non riflette sufficientemente le numerose attività che vengono svolte quotidianamente presso le unità produttive per assicurare la qualità. Nel sistema SIDI queste attività sono in genere rappresentate dalle *azioni di controllo*, e cioè dall'insieme di attività di prevenzione, controllo in corso d'opera e valutazione a posteriori dell'errore, difficilmente sintetizzabili attraverso indicatori.

Gli esempi di analisi di qualità riportati in questo paragrafo hanno come obiettivo quello di fare emergere e quantificare l'elevata mole di attività che contribuisce a garantire la qualità dell'informazione statistica prodotta dall'Istituto. Si tratta di analisi di tipo esclusivamente descrittivo e che, contrariamente a quelle che verranno presentate nel paragrafo 5, non hanno finalità valutative.

Nelle tabelle presentate, poiché più operazioni dello stesso tipo (per esempio la tecnica di rilevazione) e la maggior parte delle azioni di controllo associate ad una certa operazione, possono essere svolte all'interno di una stessa indagine, la somma delle operazioni (o delle azioni di controllo) per indagine non è rappresentativa del numero di indagini. Inoltre, indagini che adottano più di una tecnica di rilevazione possono comparire più volte nelle tabelle, come per esempio la "Rilevazione sulle forze di lavoro" che è in parte faccia a faccia assistita da computer (CAPI) in parte telefonica (CATI).

Infine, è opportuno ricordare che le analisi presentate in questo paragrafo si basano sulla documentazione presente in SIDI, e quindi possono risentire di eventuali situazioni di documentazione parziale o non accurata. Tuttavia, dati gli elevati sforzi condotti negli ultimi anni per controllare lo stato di completezza e qualità della documentazione immessa nel sistema, si ritiene che le analisi presentate possano riflettere la situazione reale dell'Istituto.

4.1. La qualità della fase di rilevazione: la mancate risposte totali

Nella conduzione delle rilevazioni, l'Istat impegna molte energie con l'obiettivo di garantire un congruo numero di rispondenti. In SIDI, queste attività sono documentate come azioni di controllo relative alla fase di rilevazione. La loro importanza è diversa a seconda della tecnica di rilevazione adottata. La tabella che segue riporta le frequenze assolute e percentuali delle indagini che adottano azioni di controllo per migliorare i tassi di risposta, per tipologia di tecnica e modalità di controllo delle mancate risposte totali.

Tabella 4.1. Qualità della fase di rilevazione in indagini dirette o miste rispetto al controllo della mancata risposta totale (frequenza assoluta e percentuale)

Azioni	Faccia a Faccia	CATI	Postale	Telematica
Prevenzione	16 (94,2)	20 (90,9)	65 (80,2)	29 (87,9)
Sollecito	2 (11,8)	3 (9,1)	54 (66,7)	29 (87,9)
Correzione	10 (58,8)	10 (45,5)	24 (29,6)	14 (42,4)
Stima	2 (11,8)	2 (9,1)	7 (8,6)	2 (6,1)
Totale rilevazioni	17	22	81	33

Nelle attività di "Prevenzione" vi sono quelle azioni di controllo consistenti in invio di lettere di preavviso sulla rilevazione, illustrazione sugli obiettivi dell'indagine, garanzie sul rispetto del segreto statistico, etc.; la categoria "Correzione" include diversi metodi correttivi (pesi, duplicazioni di record, etc.) messi in atto per ridurre gli effetti delle mancate risposte totali; le azioni di "Stima" sono costituite da indagini su campioni di non rispondenti o analisi *ad hoc* per stimare la variabilità e/o la distorsione dovuta ai non rispondenti. E' evidente come per tutte le modalità di raccolta dei dati l'attività di prevenzione dell'errore sia molto elevata. Queste attività, insieme alle azioni di sollecito (postali, telefonici, ecc.) sono di riconosciuta importanza ed efficacia, particolarmente in indagini postali, che le adottano estensivamente. Dalla tabella 4.1. si evince che attività più onerose sul fronte dei costi, quelle che consentono di stimare l'impatto della non risposta, sono ovviamente meno frequenti.

Dalla documentazione presente nel sistema emerge che la maggior parte delle 85 rilevazioni di fonte amministrativa o a carattere misto (in parte di fonte amministrativa in parte diretta) adotta una trasmissione dei dati di tipo cartaceo, ma anche che un buon numero utilizza almeno la modalità telematica (Tabella 4.2.). Anche per questa tipologia di indagini emerge una generale elevata attitudine ad effettuare attività di prevenzione e di sollecito, indipendentemente dalle modalità di trasmissione.

Tabella 4.2. Qualità della fase di rilevazione in indagini di fonte amministrativa o miste per modalità di trasmissione del dato (frequenza assoluta e percentuale)

Azioni	Trasmissione su supporto cartaceo	Trasmissione su supporto informatico	Trasmissione telematica
Prevenzione	18 (40,0)	11 (50,0)	15 (41,7)
Sollecito	23 (51,1)	8 (36,4)	19 (52,8)
Correzione	3 (6,6)	4 (18,2)	3 (8,8)
Stima	0 (0,0)	1 (4,5)	0 (0,0)
Totale rilevazioni	45	22	36

Per le 199 rilevazioni (dirette, di fonte amministrativa o miste), la frequenza delle modalità di sollecito dichiarate è riportata nella tabella 4.3.

Tabella 4.3. Modalità di sollecito per tipologia di rilevazione (frequenza assoluta e percentuale)

	Dirette	Fonte amministrativa	Miste	Totale
Telefonici	48 (42,1)	27 (42,8)	6 (27,3)	81
Postale o con Postel	38 (33,3)	10 (15,9)	1 (4,5)	49
Fax, fax-server o postfax	19 (16,7)	10 (15,9)	0 (0,0)	29
Posta elettronica	7 (6,1)	4 (6,3)	0 (0,0)	11
Totale rilevazioni	114	63	22	199

In più del 40 % delle indagini di tipo diretto, è effettuato almeno un sollecito di tipo telefonico o postale. Per le indagini di fonte amministrativa o miste è di gran lunga predominante il peso del sollecito telefonico rispetto al totale delle modalità di sollecito utilizzate, come è da attendersi, visto che in genere fanno riferimento ad un numero di unità/enti di rilevazione più limitato.

4.2. La qualità della fase di rilevazione: gli intervistatori

Un altro aspetto particolarmente critico per la qualità dei dati finali di una indagine è rappresentato dall'attività degli intervistatori. Infatti, laddove la tecnica di indagine preveda la presenza dell'intervistatore, questo è il tramite attraverso cui viene somministrato il questionario e vengono rilevati i dati e può, con il suo atteggiamento, la sua preparazione e le sue capacità, influenzare fortemente la qualità del dato rilevato. Per questo motivo è usuale all'Istat condurre numerose attività per la formazione, monitoraggio e valutazione del lavoro degli intervistatori.

Dai dati presenti in SIDI (Tabella 4.4.) risulta che, per le tre tecniche con rilevatore prese in considerazione, l'attività di prevenzione dell'errore consistente nella effettuazione di corsi di formazione e nella predisposizione accurata di libretti di istruzione è estensivamente condotta (più del 73% delle indagini che adottano il CATI fino al 100% delle indagini che adottano il CAPI). Anche l'attività di controllo in corso d'opera è effettuata con regolarità. Questa va dalle riunioni di *debriefing*, alla supervisione in sala, fino alla verifica della qualità e completezza dei questionari attraverso reinterviste e al monitoraggio attraverso indicatori sui tassi di risposta specifici per rilevatore.

Tabella 4.4. Qualità della fase di rilevazione rispetto al controllo degli intervistatori (frequenza assoluta e percentuale)*

Azioni	Interviste faccia a faccia con questionario cartaceo	Interviste faccia a faccia assistite da computer (CAPI)	Interviste telefoniche assistite da computer (CATI)
Prevenzione	14 (93,3)	3 (100,0)	19 (86,4)
Controllo in corso d'opera	13 (86,7)	3 (100,0)	17 (77,3)
Stima	3 (20,0)	1 (33,3)	4 (18,2)
Totale rilevazioni	15	3	22

* Si noti che le rilevazioni faccia a faccia con questionario cartaceo e quelle assistite da computer sommano a 18 e non a 17 come nella tabella 4.1, in quanto una rilevazione adotta entrambe queste modalità.

4.3. La fase di trattamento dei dati: le procedure di controllo e correzione

In SIDI le procedure di controllo e correzione sono documentate in relazione alla tipologia: procedure completamente deterministiche, procedure completamente probabilistiche, procedure miste, con il dettaglio delle diverse combinazioni. Tipicamente una procedura di controllo e correzione può essere costituita da diversi passi in sequenza e prevedere quindi più tipi di tecnica di controllo e correzione.

Sulle 199 rilevazioni documentate nel sistema SIDI, il 66% ha documentato almeno una operazione di controllo e correzione. Questo dato, probabilmente basso, potrebbe dipendere da una non completa documentazione relativamente a queste attività nel sistema SIDI. Volendo andare ad esaminare queste per tecnica di rilevazione, nella tabella che segue, la tecnica CAPI è stata raggruppata insieme alla CATI in quanto entrambe utilizzano questionari elettronici somministrati da un intervistatore.

Tabella 4.5. La fase di trattamento dei dati: le procedure di controllo e correzione per tecnica di rilevazione (frequenza assoluta e percentuale)

	Faccia a Faccia (cartaceo)	CATI e/o CAPI*	Postale	Telematica
Procedure completamente deterministiche	11 (73,3)	6 (33,3)	23 (28,4)	8 (24,2)
Procedure completamente probabilistiche	8 (53,3)	2 (11,1)	0 (0,0)	0 (0,0)
Procedure completamente interattive	3 (20,0)	8 (36,4)	14 (17,3)	4 (12,1)
Procedure miste	9 (60,0)	14 (63,6)	24 (29,6)	15 (45,5)
Totale rilevazioni	15	22	81	33

* Le 3 indagini CAPI che vengono sottratte dalla colonna relativa alle rilevazioni Faccia a faccia sono anche CATI e quindi non si sommano a questa colonna, che quindi rimane invariata rispetto al totale riportato nelle precedenti tabelle.

Più del 70 % delle rilevazioni faccia a faccia con questionario cartaceo effettua almeno un piano deterministico per il controllo e la correzione dei dati. Anche per le indagini postali questo è il metodo più utilizzato. Le procedure completamente probabilistiche, in genere utilizzate per le indagini sugli individui o sulle famiglie, sono completamente assenti per le indagini postali e telematiche, tecniche usualmente adottate sulle imprese.

Le procedure interattive sono più frequenti nelle indagini di tipo CATI e/o CAPI dove ovviamente sono ridotti i sistemi completamente deterministici potendo effettuare parte dei controlli durante l'intervista. Relativamente alle procedure miste, bisogna sottolineare che la maggior parte delle volte si tratta di procedure con localizzazione di tipo deterministico e imputazione da parte di operatore, ad eccezione del caso delle CATI e/o CAPI dove c'è una certa proporzione di procedure con imputazione probabilistica.

4.4. La documentazione sulla qualità per le elaborazioni

Rispetto alle rilevazioni, le elaborazioni sono state documentate nel sistema SIDI molto più di recente, a partire dal 2003, e con la principale finalità di produrre le note metodologiche dell'Annuario Statistico Italiano. Pertanto il livello di approfondimento disponibile per la documentazione delle elaborazioni è inferiore a quello delle rilevazioni. L'analisi presentata in questo paragrafo quindi è solo una valutazione iniziale di alcuni aspetti di processo relativi alle elaborazioni.

In totale, le elaborazioni documentate nel sistema sono 88. Nella tabella 4.6. si presentano le frequenze delle operazioni di "Pre-elaborazione delle fonti". Più di un quarto delle elaborazioni effettua una analisi preventiva della idoneità delle fonti utilizzate come fonti di input (per molte elaborazioni bisogna considerare che non vi sono alternative all'utilizzo di determinate fonti). Circa il 30% delle elaborazioni effettua delle trasformazioni per conciliare le definizioni o le classificazioni e quasi il 30% effettua attività di controllo e correzione di dati anomali e/o mancanti.

Tabella 4.6. Operazioni di pre-elaborazione sulle fonti per le elaborazioni (frequenza assoluta e percentuale)

	Frequenze
Analisi di idoneità delle fonti	23 (26,1)
Trasformazione dei dati per conciliare le definizioni/classificazioni	87 (98,9)
Integrazione o unione di micro o macro dati	22 (25,0)
Controllo e correzione di dati anomali e/o mancanti	38 (28,8)
Totale elaborazioni	88

Infine, è attività corrente per i settori di produzione quella di effettuare un vero e proprio processo di validazione prima di rilasciare provvisoriamente o definitivamente il dato. Nella tabella 4.7. si riportano le frequenze delle diverse operazioni di validazione documentate per le elaborazioni.

Tabella 4.7. Operazioni di validazione per le elaborazioni (frequenza assoluta e percentuale)

	Frequenze
Confronto con dati di altre fonti indipendenti	46 (52,3)
Analisi delle serie storiche	79 (89,8)
Analisi di indicatori di qualità a corredo dell'informazione	14 (15,9)
Verifica della completezza delle stime rispetto alle definizioni (esaustività)	7 (7,9)
Studi o sperimentazioni supplementari sulla qualità dei dati utilizzati	8 (9,0)
Totale elaborazioni	88

L'analisi delle serie storiche di indicatori e rapporti caratteristici è una attività di validazione estensivamente utilizzata per le elaborazioni (quasi il 90%). Anche il confronto con fonti indipendenti è frequentemente effettuato prima di rilasciare i dati.

5. Analisi degli indicatori di qualità

Come già menzionato nel paragrafo 2, gli indicatori presenti in SIDI consentono di effettuare delle analisi di sintesi sui livelli e l'andamento della qualità dei diversi processi dell'Istituto. In questo paragrafo verranno presentati alcuni esempi di possibili analisi. Sebbene già sottolineato, è opportuno ricordare che la validità dei risultati ottenuti dipende dal livello di popolamento del sistema, per cui, soprattutto per quegli indicatori che sono presenti solo per un sottoinsieme dei processi in essere, le analisi riportate vanno considerate più a livello di esempio che come conclusioni generalizzabili all'intero Istituto.

5.1 Tempestività

In questo paragrafo vengono riportati alcune analisi relative alla tempestività. L'indicatore è definito come la differenza in giorni tra la data in cui i risultati definitivi di un processo statistico vengono diffusi agli utenti esterni e la data a cui essi si riferiscono; in particolare, per dati che si riferiscono ad un periodo e non ad un momento specifico, viene considerato l'ultimo giorno di tale periodo.

Nella tabella 5.1 viene riportato il valore medio dell'indicatore di tempestività per anno di riferimento dei dati e periodicità delle rilevazioni per cui esso è presente in SIDI (dati aggiornati al 22 dicembre 2006). La copertura in termini di numero di rilevazioni di questo indicatore è pressoché totale, in quanto le direttive dell'Istituto nel 2005 ne hanno reso obbligatorio il calcolo e nel 2006 hanno previsto la prosecuzione della serie storica, però il periodo di riferimento per cui essi sono stati inseriti varia da processo a processo. Si sottolinea che nell'analisi non sono state comprese le rilevazioni con periodicità pluriennale e infrannuale diversa da mensile e trimestrale (ad esempio le rilevazioni semestrali); inoltre non sono state considerate alcune rilevazioni con periodicità di rilascio dei dati diverse per diversi tipi di utenti che comportavano problemi di confrontabilità (ad esempio alcune rilevazioni mensili che forniscono mensilmente i dati a Eurostat ma rilasciano dati agli utenti esterni solo annualmente).

Tabella 5.1. Rilevazioni: tempestività media (in giorni) per anno di riferimento dei dati e periodicità

	mensile	trimestrale	annuale
1995	-	80 (1)	366 (1)
1996	-	78 (1)	593 (4)
1997	-	637 (2)	517 (4)
1998	-	783 (4)	494 (4)
1999	-	591 (4)	586 (4)
2000	-	572 (4)	723 (20)
2001	62 (5)	424 (5)	690 (37)
2002	63 (5)	431 (6)	603 (42)
2003	134 (10)	422 (8)	540 (61)
2004	143 (19)	247 (13)	471 (53)
2005	116 (21)	216 (11)	310 (16)
2006	62 (14)	145 (8)	-
Media generale	112 (22)	375 (16)	558 (74)

Il numero dei processi è riportato in parentesi

La periodicità è un fattore fondamentale per la corretta interpretazione della tempestività, perché periodicità diverse corrispondono in genere a rilevazioni di tipo diverso sia per quanto riguarda la tipologia di dati prodotti (es.: rilevazioni a cadenza mensile in genere portano alla produzione di dati congiunturali, mentre le annuali corrispondono a informazioni strutturali) sia per il livello di complessità del loro processo produttivo. E' quindi fisiologico che i livelli di tempestività siano molto diversi tra loro.

Dall'esame della tabella, osservando i valori della tempestività media per le indagini annuali dal 2000 (primo anno con un numero considerevole di valori) al 2005, si osserva un incoraggiante trend decrescente nel valore dell'indicatore, che indica un miglioramento della tempestività. Ovviamente va considerato che per il 2004 e, soprattutto, per il 2005 tale valore tenderà a salire nei prossimi anni perché è possibile che le rilevazioni più lente a fine 2006 non abbiano ancora diffuso i dati definitivi, ma è anche da sottolineare che la tendenza decrescente è già evidente tra il 2000 e il 2003 e che il valore del 2004, con 53 rilevazioni che già contribuiscono all'indicatore, non dovrebbe incrementare eccessivamente.

Per quanto riguarda le rilevazioni trimestrali, i valori possono sembrare piuttosto elevati, ma bisogna sottolineare in primo luogo che l'indicatore si riferisce alla diffusione di dati definitivi, mentre molto spesso per le rilevazioni di questo tipo è il dato provvisorio quello di maggior interesse per gli utenti. Per avere un'idea della differenza tra dati definitivi e dati provvisori, calcolando la media generale della tempestività dei dati provvisori per le indagini trimestrali si ottiene un valore di 140 giorni rispetto a 375 giorni dei dati definitivi.

Anche per le rilevazioni mensili i valori dell'indicatore risultano piuttosto elevati. In tabella 5.2 viene effettuato un approfondimento dell'analisi. Si può infatti esaminare la distribuzione del valore medio della tempestività per anno di riferimento dei dati e tipologia delle indagini mensili. In questo caso le rilevazioni mensili sono state classificate alternativamente in "Congiunturali" o "Amministrative o strutturali" sulla base della direzione produttiva a cui tale rilevazione afferisce.

Tabella 5.2. Rilevazioni mensili: tempestività media (in giorni) per anno di riferimento dei dati e tipo di processo

	Amministrative o strutturali	Congiunturali	Totale mensili
2001	181 (3)	34 (2)	62 (5)
2002	181 (3)	36 (2)	63 (5)
2003	212 (7)	41 (3)	134 (10)
2004	197 (10)	71 (9)	143 (19)
2005	208 (10)	51 (11)	116 (21)
2006	124 (4)	48 (10)	62 (14)
Media generale	197 (11)	52 (11)	112 (22)

Il numero dei processi è riportato in parentesi

Si osserva che effettivamente le rilevazioni di natura congiunturale presentano livelli medi di tempestività molto buoni, con una media generale pari a 52 giorni. Ciò è dovuto anche, naturalmente, al fatto che buona parte di esse (8 su 11) sono soggette a regolamenti europei. Le rilevazioni a carattere amministrativo o strutturale, invece, hanno tempestività peggiori, dovute, probabilmente, in parte al dover attendere la disponibilità del dato amministrativo (si ricorda che la tempestività fa riferimento al periodo di riferimento del dato, quindi ad esempio al periodo di riferimento dell'evento, che deve essere prima registrato per via amministrativa per poi giungere all'Istat) e in parte alle politiche di diffusione attuate (ad es.: alcune indagini diffondono sul Bollettino mensile di statistica che presenta un ritardo di alcuni mesi), sulle quali si potrebbe intervenire per anticipare la diffusione dei risultati di molte indagini.

Un ulteriore esempio di analisi della tempestività è costituito dallo studio delle variazioni della tempestività nel tempo. Le variazioni relative percentuali rappresentano una misura delle variazioni di tempestività che si verificano tra repliche consecutive di un processo; la loro media, calcolata per anno e per diverse periodicità, fornisce una sintesi dell'andamento nel tempo della tempestività dei processi dell'Istituto. In particolare le variazioni non risentono del fatto che le indagini che contribuiscono alle medie possono non essere le stesse in anni diversi, e quindi possono fornire una valida indicazione dell'andamento della tempestività nel tempo. La variazione relativa percentuale viene definita come:

$$v_{ij} = \frac{t_{i(j-1)} - t_{ij}}{t_{ij}} \cdot 100$$

dove t è l'indicatore di tempestività misurato in giorni, i indica la rilevazione mentre j indica la replicazione.

Tabella 5.3. Variazioni relative percentuali medie per replicazione, distribuite per anno di riferimento dei dati e per diverse periodicità dell'indagine (%)

	mensile*	trimestrale	annuale
2001	8,63 (1)	11,98 (4)	0,62 (21)
2002	1,03 (5)	8,57 (6)	5,65 (31)
2003	3,89 (10)	5,19 (7)	12,46 (41)
2004	0,52 (18)	6,33 (13)	12,18 (49)
2005	0,66 (21)	0,93 (10)	29,1 (14)
2006	1,96 (14)	-1,16 (8)	
Media generale	1,38 (22)	5,08 (15)	10,92 (63)

Il numero dei processi è riportato in parentesi

Le medie delle variazioni sono calcolate per ogni anno di riferimento dei dati e per diverse periodicità (tabella 5.3.). I valori positivi indicano un miglioramento della tempestività ovvero che la tempestività di una replicazione j è inferiore alla tempestività della precedente ($j-1$).

I dati mostrano un generale miglioramento della tempestività riferito alla singola replicazione, pari al +10,92% per le rilevazioni annuali, al +1,38% per le mensili ed al +5,08% per le trimestrali. Il valore negativo relativo alle trimestrali del 2006 (-1,16%), è da considerarsi un dato parziale che potrà essere soggetto a variazioni sia in senso positivo che negativo una volta che sarà completato il calcolo degli indicatori per l'anno 2006.

Nel corso del 2006 per la prima volta anche le elaborazioni hanno avuto la possibilità di inserire alcuni indicatori di qualità in SIDI, tra cui l'indicatore di tempestività. Il calcolo di tali indicatori non era obbligatorio, per cui il livello di popolamento deve essere considerato solo "embrionale", tanto più che in SIDI sono presenti solo le elaborazioni che contribuiscono all'Annuario Statistico Italiano. I risultati in termini di medie per anno di riferimento dei dati e periodicità sono comunque riportati in tabella 5.4. I livelli medi di tempestività risultano leggermente più elevati rispetto alle rilevazioni, com'era prevedibile considerando che le elaborazioni si innestano in genere a valle di uno o più processi di rilevazione o elaborazione e quindi risentono della tempestività delle fonti utilizzate.

Tabella 5.4. Elaborazioni: tempestività media (in giorni) per anno di riferimento dei dati e periodicità

	mensile	trimestrale	Annuale
1996	91 (1)		
1997	91 (1)		669 (1)
1998	91 (1)		670 (1)
1999	89 (1)		601 (2)
2000	108 (2)		410 (4)
2001	108 (2)		476 (6)
2002	119 (3)		663 (8)
2003	120 (3)		501 (9)
2004	143 (5)	446 (1)	497 (9)
2005	236 (7)	407 (1)	306 (5)
2006	152 (1)		
Media generale	118 (9)	426 (2)	508 (18)

Il numero dei processi è riportato in parentesi

In conclusione per quanto riguarda la tempestività a livello di Istituto si segnala che negli ultimi anni si sono apprezzati miglioramenti ragguardevoli, ma ci sono sicuramente ancora margini di miglioramento, almeno per alcune tipologie di processo: ad esempio per le rilevazioni annuali il range della tempestività per il 2003 (media 540 gg. calcolata su 61 processi) va da 118 a 1095 giorni. Sebbene i livelli di complessità delle rilevazioni possano essere molto diversi, quelle più lente potrebbero prendere spunto dalle più rapide per introdurre delle innovazioni nel processo che migliorino la tempestività. Dall'analisi degli indicatori emerge che per le indagini congiunturali, per le quali la tempestività è sicuramente più importante che per altre tipologie di processo, i livelli raggiunti sono già ottimali e difficilmente possono essere migliorabili. In tali casi probabilmente conviene investire nel miglioramento dei livelli di accuratezza a parità di tempestività.

5.2 Accuratezza

Si passa ora ad esaminare alcuni esempi di indicatori indiretti sull'accuratezza. In particolare si analizzeranno i valori medi del tasso di errori di lista e del tasso di risposta. Tale analisi permetterà di trarre alcune conclusioni sui livelli e sull'andamento degli errori di copertura e di mancata risposta per le rilevazioni dell'Istituto che hanno inserito i relativi indicatori in SIDI. E' infatti necessario sottolineare che a differenza della tempestività, per gli indicatori di copertura e mancata risposta le direttive hanno richiesto solo che fosse proseguita la serie storica degli indicatori per le rilevazioni che già li avevano immessi o che alcuni processi ampliassero il set degli indicatori. In conseguenza dell'assenza di vincoli di obbligatorietà e del maggiore onere richiesto per il loro calcolo, le rilevazioni che avevano calcolato tali indicatori nel 2005 sono 32 e nel 2006 sono 45 (su 122 eleggibili).

Ovviamente tali processi non possono essere considerati rappresentativi di tutti i processi, ma possono comunque fornire qualche indicazione di massima. Particolare cautela va prestata anche all'analisi dei valori degli indicatori nel tempo, perché le rilevazioni che ogni anno contribuiscono alle medie riportate possono essere diverse.

Analisi degli errori di lista

In tabella 5.5 è riportato il valore medio del tasso di errori di lista per anno di riferimento dei dati e tecnica di rilevazione prevalente. Il tasso di errori di lista è definito come segue⁵:

$$\text{Tasso di errori di lista} = \frac{\text{Numero di unità non eleggibili} + \text{Numero di unità non contattate per errori di lista}}{\text{Numero di unità risolte}} \cdot 100$$

Esso fornisce una valutazione complessiva, sebbene indiretta, della bontà dell'archivio di estrazione delle unità di rilevazione in termini di sovracopertura ed errori nelle variabili di localizzazione, che pregiudicano il contatto con le unità comportando un incremento della mancata risposta.

Tabella 5.5. Media del tasso di errori di lista per anno di riferimento dei dati e tecnica di rilevazione (%)

	Amministrative	CATI	Faccia a faccia	Postali e/o Telematiche	Media generale
1997			0,78 (1)		0,78 (1)
1998				14,91 (2)	14,91 (2)
1999				12,75 (3)	12,75 (3)
2000				11,48 (6)	11,48 (6)
2001		22,12 (3)	0,88 (1)	13,12 (7)	14,46 (11)
2002		7,73 (4)	1,02 (1)	9,02 (7)	7,93 (12)
2003	0 (3)	2,92 (2)	2,65 (4)	8,16 (11)	5,31 (20)
2004	0,62 (4)	9,37 (4)	5,92 (2)	7,64 (15)	6,66 (25)
2005	0 (2)	7,94 (4)	3,66 (4)	5,9 (14)	5,37 (24)
2006	0 (2)	1,36 (1)	3,44 (2)	1,68 (5)	1,67 (10)
Media generale	0,49 (5)	11,63 (7)	2,63 (7)	7,53 (26)	6,63 (45)
Media generale calcolata a fine 2005	0,62 (4)	13,11 (6)	2,85 (3)	8,56 (19)	7,88 (32)

Il numero dei processi è riportato in parentesi

Confrontando in primo luogo i valori medi generali dell'indice calcolato sia con i dati disponibili in SIDI a fine 2005 sia con i dati presenti attualmente si può notare un generale miglioramento del tasso di errori di lista per tutte le diverse tecniche di rilevazione. Essendo le medie calcolate a fine 2006 più robuste di quelle precedenti, poiché calcolate su un maggior numero di processi e di edizioni per processo, il dato va considerato in termini positivi, sebbene non generalizzabile alla totalità delle rilevazioni dell'Istat, per quanto detto in precedenza. Analizzando più in dettaglio l'indicatore rispetto alle varie tecniche di rilevazione, risulta evidente che gli errori di lista sono pressoché nulli per le indagini di natura amministrativa, poiché le unità di rilevazione sono in genere enti locali per i quali sono disponibili liste molto affidabili. Anche per le indagini che utilizzano la tecnica faccia a faccia, il livello medio del tasso risulta sempre molto buono. Va menzionato che tutte le rilevazioni comprese in questa categoria utilizzano come archivio l'anagrafe comunale, che risulta quindi piuttosto affidabile.

⁵ Si definiscono:

- "Unità Non Eleggibili" le unità che non appartengono alla popolazione oggetto di indagine pur essendo presente nell'archivio o lista di estrazione;
- "Unità Non Contattate per Errori di Lista" le unità eleggibili che non hanno partecipato all'indagine perché non si è riusciti a contattarle per imprecisioni o informazioni insufficienti nella lista;
- "Unità Risolte" le unità per cui è stato possibile accertare se erano o meno eleggibili, cioè se appartenevano alla popolazione oggetto di indagine (Hidiroglou *et al.*, 1993)

Le rilevazioni postali e/o telematiche, in cui ricadono in maggioranza le indagini dirette sulle imprese, presentano dei valori mediamente più elevati, sebbene con un trend decrescente nel corso degli anni. Il livello medio più elevato è comprensibile considerando la dinamicità della demografia delle imprese in Italia, per cui è difficile avere a disposizione un archivio perfettamente aggiornato al momento delle diverse rilevazioni, mentre il trend decrescente testimonia gli sforzi messi in atto dall'Istituto per migliorare tali liste, in particolare l'archivio ASIA, che rappresenta la lista di riferimento per il 73% di queste rilevazioni.

Per quanto riguarda le rilevazioni effettuate con tecnica CATI, risulta evidente l'elevata variabilità dei valori medi dell'indicatore nei diversi anni. Ciò è da attribuirsi alle diverse caratteristiche delle rilevazioni che di anno in anno contribuiscono al calcolo delle medie. Il livello medio generale piuttosto elevato dell'indicatore (11,63%) per questo tipo di rilevazioni è indicativo delle problematiche, tipiche delle indagini telefoniche, connesse con la disponibilità di una lista di riferimento consistente con la popolazione obiettivo dell'indagine e aggiornata con le informazioni (i numeri di telefono) che consentano il contatto con essa.

Analisi del tasso di risposta

Un altro esempio di analisi degli indicatori presenti in Sidi è costituito dallo studio del tasso di risposta. In tabella 5.6 è riportato il valore medio del tasso di risposta per anno di riferimento dei dati e tecnica di rilevazione prevalente. Il tasso di risposta è definito come segue⁶:

$$\text{Tasso di risposta} = \frac{\text{Numero di unità rispondenti } i}{\text{Numero di unità eleggibili} + \text{Numero di unità non risolte}} \cdot 100$$

Il tasso di risposta costituisce un importante indicatore indiretto dell'accuratezza, in quanto è associato sia alla distorsione che alla variabilità delle stime.

Dall'esame delle medie generali della tabella 5.6 si osserva che il tasso di risposta delle indagini amministrative presenta i valori più elevati, pari ad oltre il 95%, in quanto i rispondenti sono costituiti in genere da enti locali, che spesso collaborano in base a specifiche norme di legge. Il tasso di risposta per le indagini CATI assume un valore medio pari al 67,05%, naturalmente inferiore alle tecniche faccia a faccia per le maggiori difficoltà di contatto telefonico rispetto al contatto diretto. Infatti le indagini che hanno adottato la tecnica faccia a faccia presentano un tasso di risposta medio pari all' 83,19%. Per le indagini CATI si osserva un progressivo miglioramento del tasso di risposta, che passa dal 56,29% al 73,51%, ad eccezione del 2006, il cui valore non è da considerarsi definitivo.

⁶ Si definiscono:

- "Unità Rispondenti" le unità di rilevazione per le quali è stato possibile rilevare l'informazione desiderata;
- "Unità Eleggibili" le unità che appartengono alla popolazione oggetto di indagine;
- "Unità Non Risolte" le unità per quali non è stato possibile accertare se erano o meno eleggibili. (Hidirolou *et al.*, 1993)

Tabella 5.6. Media del tasso di risposta per anno di riferimento dei dati e tecnica di rilevazione prevalente (%)

	Amministrative	CATI	Faccia a faccia	Postali e/o Telematiche*	
				Monotecnica	Multitecnica
1997			96,43 (1)		
1998				45,35 (2)	
1999				58,79 (2)	51,37 (1)
2000				62,33 (5)	92,48 (1)
2001		56,29 (3)	80,53 (1)	67,64 (6)	86 (1)
2002		62,5 (4)	84,6 (1)	67,73 (5)	87,51 (2)
2003	93,69 (3)	58,68 (2)	81,59 (4)	73,05 (9)	85,94 (2)
2004	94,79 (4)	69,12 (4)	79,62 (2)	59,6 (7)	84,47 (8)
2005	99,23 (2)	73,51 (4)	80,16 (4)	70,29 (6)	74,58 (6)
2006	100 (2)	57,47 (1)	80,43 (2)	71,9 (2)	67,62 (2)
Media generale	95,78 (5)	67,05 (7)	83,19 (7)	67,34 (16)	75,35 (13)
Variazione media relativa percentuale (%)	0,06 (4)	7,67 (5)	-0,32 (3)	1,84 (13)	-0,79 (11)

Il numero dei processi è riportato in parentesi

* Il totale delle rilevazioni postali e/o telematiche diventa 29 (rispetto a 26 indicato nella precedente tavola) perché 3 indagini monotecnica diventano multitecnica nel tempo

L'approccio multitecnica, che nella maggior parte dei casi consiste nell'abbinamento alla tradizionale indagine postale di tecniche web di acquisizione dati, risulta essere sempre più diffuso tra le indagini dell'istituto. L'uso di più tecniche risulta essere efficace nell'ambito delle 29 indagini postali e/o telematiche indicate in tabella (di cui 16 adottano l'approccio monotecnica e 13 l'approccio multitecnica). Infatti per tutti gli anni fino al 2005 il tasso di risposta è maggiore per le indagini che hanno adottato l'approccio multitecnica e in media si ha un vantaggio nel tasso di risposta pari all'8,01%. Per quanto riguarda il dato relativo al 2006, che sembra essere in controtendenza, va tenuto conto del fatto che soltanto 4 indagini contribuiscono al calcolo delle relative medie.

E' stata inoltre calcolata la variazione relativa percentuale per il tasso di risposta definita come

$$vtr_{ij} = \frac{tr_{ij} - tr_{i(j-1)}}{tr_{ij}} \cdot 100$$

dove tr è tasso di risposta, i indica la rilevazione mentre j indica la replicazione. I valori positivi indicano che il tasso di risposta è migliorato tra una replicazione e la successiva del medesimo processo.

I valori riportati nella tabella 5.6 indicano che per le indagini prevalentemente CATI vi è una variazione media positiva nel corso degli anni pari a 7,67%. Anche le indagini monotecnica presentano un miglioramento anche se più moderato, pari a + 1,84%. Sembra, inoltre, che vi sia un leggerissimo peggioramento nel tempo tra le indagini multitecnica (variazione pari a -0,79%). Quest'ultimo dato, comparato con il lieve miglioramento osservato nelle monotecnica e tenendo conto che negli anni la differenza nel tasso di risposta tra mono e multitecnica tende ad assottigliarsi, porta a concludere che nel tempo il vantaggio dell'approccio multitecnica rispetto all'approccio monotecnica si vada progressivamente riducendo. Questa indicazione andrà naturalmente rivalutata in futuro, quando si potrà disporre di serie di dati più lunghe ed un maggior numero di indagini avrà calcolato gli indicatori di mancata risposta.

Considerazioni finali

Le analisi presentate in questo lavoro mostrano come SIDI fornisca un ricco quadro informativo sulla qualità dei processi, rilevazioni ed elaborazioni, dell'Istituto e come questo patrimonio informativo sia utilizzabile per numerose finalità.

In primo luogo, SIDI garantisce una funzione documentale nel senso che fotografa la situazione sulla qualità esistente all'Istituto in un dato momento. Le analisi svolte, in particolare quelle qualitative, testimoniano il notevole investimento che l'Istituto correttamente fa in prevenzione e monitoraggio degli errori non campionari.

In secondo luogo, SIDI consente una analisi dell'evoluzione temporale del "sistema indagini", sia da un punto di vista metodologico-procedurale che da un punto di vista quantitativo. Le funzionalità standard di analisi (ad esempio confronto di indicatori medi nel tempo) e quelle aggiuntive riportate nel lavoro, consentono di prendere atto di quali fattori siano migliorati o, viceversa, peggiorati nel tempo. Ad esempio si riscontrano miglioramenti nella tempestività delle rilevazioni annuali (che potrebbe probabilmente migliorare ancora), miglioramenti nei tassi di errori di lista (in particolare per le indagini CATT), miglioramenti nei tassi di risposta delle rilevazioni postali multitecnica (rispetto alle rilevazioni solo postali) anche se i miglioramenti maggiori sembrano essersi ottenuti nel periodo 2001 - 2004 e l'efficacia sembra attenuarsi negli ultimi anni (aspetto che andrebbe approfondito dato l'investimento crescente dell'Istituto verso l'uso del web).

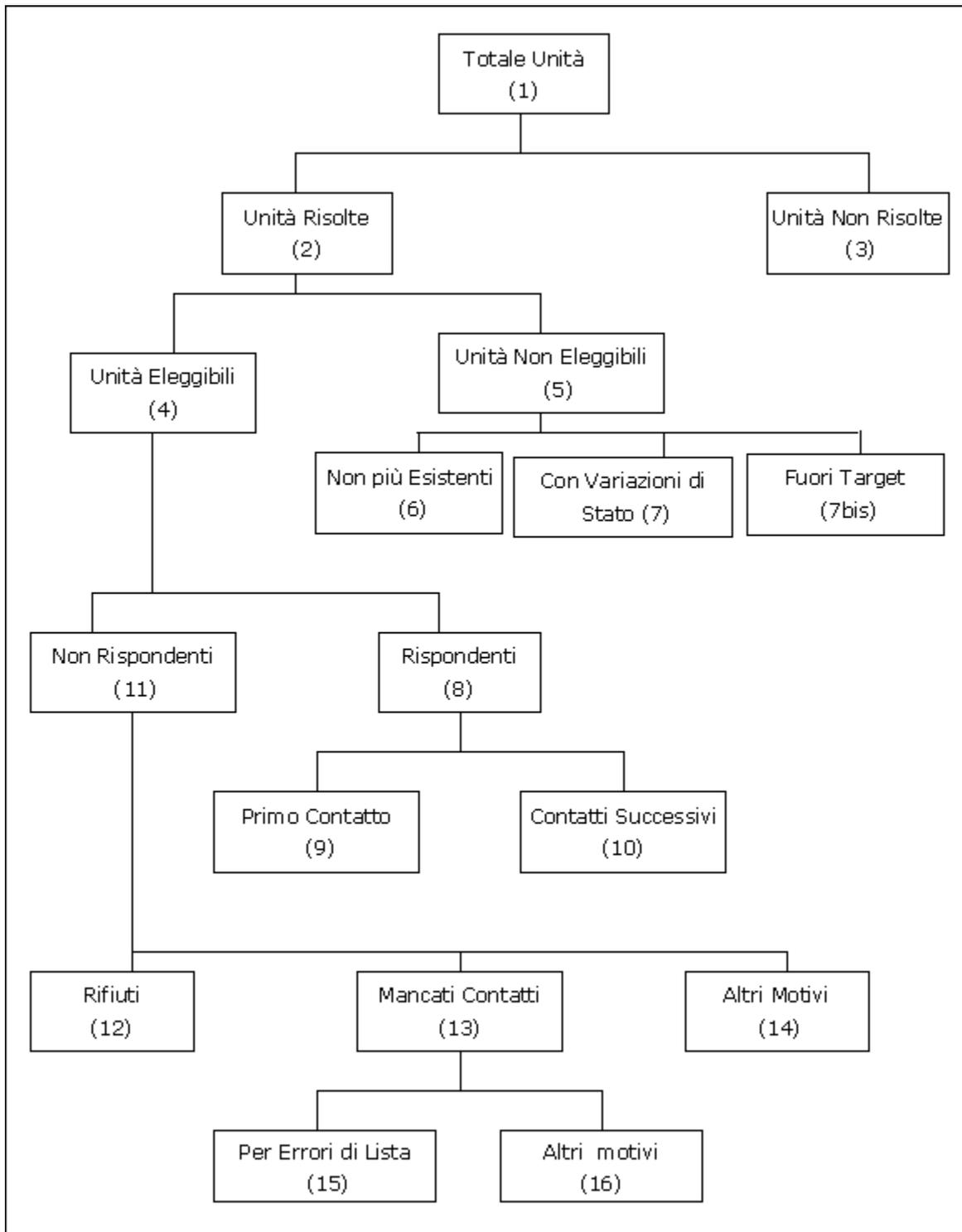
Infine, le informazioni gestite in SIDI supportano un processo di miglioramento della qualità. Ciò può essere attuato a diversi livelli (responsabile di indagine, struttura centralizzata sulla qualità, vertici dell'Istituto), ciascuno dei quali ha finalità specifiche. In generale azioni di miglioramento dovrebbero essere prese, per le rispettive sfere di competenza e con adeguate modalità, da tutti e tre i profili sopra indicati. L'aspetto importante da richiamare è che il sistema consente di verificare, qualitativamente e quantitativamente, l'efficacia delle azioni di miglioramento intraprese. In sintesi, SIDI costituisce una base informativa per l'individuazione dei punti (a livello di processo, di gruppo di indagini o di Istituto) dove intervenire e per la valutazione a posteriori dei miglioramenti di qualità conseguiti.

In questo lavoro si è voluto fornire una panoramica delle potenzialità del sistema a supporto della politica per la qualità dell'Istituto: le attività future dovranno riguardare il consolidamento delle analisi di qualità e l'uso sistematico di informazioni qualitative e quantitative per il miglioramento continuo dei processi.

Riferimenti bibliografici

- Brancato G., Carbini R., Pellegrini C., Signore M., Simeoni G. (2006a) “Assessing quality through the collection and analysis of standard quality indicators: the Istat experience” Proceedings of the European Conference on Quality in Survey Statistics (Q2006), Cardiff, UK
- Brancato G., Pellegrini C, Simeoni G. e Signore M. (2006b) “La documentazione per la qualità: finalità ed usi del sistema SIDI” Lavoro presentato alla I giornata del Seminario sulla qualità: l’esperienza dei referenti del sistema informativo SIDI, 30 ottobre 2006, Roma
- Brancato G, Pellegrini C, Simeoni G. e Signore M. (2004a) “Standardising, Evaluating and Documenting quality: the implementation of Istat information system for survey documentation – SIDI”. Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004). Mainz, 24-26 May 2004. CD ROM.
- Brancato G., Simeoni G. (2004b) “Tesauri del Sistema Informativo di Documentazione delle Indagini (SIDI)”. Istat, Documenti n.1, Anno 2004.
- Cigas M., (2006) “The comparative study on the selected quality issues – content, purpose, usage and main problems” Proceedings of the European Conference on Quality in Survey Statistics (Q2006), Cardiff, UK
- Pellegrini C. (2006) “Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l’esperienza SIDI”. Documenti Istat n.4, 2006
- SIQual <http://siqua.istat.it>

Figura. 3.1 - Classificazione delle unità secondo gli aggregati utili per il calcolo degli indicatori di copertura e mancata risposta totale. Indicazioni generali SIDI



Contributi ISTAT(*)

- 1/2002 - Francesca Biancani, Andrea Carone, Rita Pistacchio e Giuseppina Ruocco - *Analisi delle imprese individuali*
- 2/2002 - Massimiliano Borgese - *Proposte metodologiche per un progetto d'indagine sul trasporto aereo alla luce della recente normativa comunitaria sul settore*
- 3/2002 - Nadia Di Veroli e Roberta Rizzi - *Proposta di classificazione dei rapporti di lavoro subordinato e delle attività di lavoro autonomo: analisi del quadro normativo*
- 4/2002 - Roberto Gismondi - *Uno stimatore ottimale in presenza di non risposte*
- 5/2002 - Maria Anna Pennucci - *Le strategie europee per l'occupazione dal Libro bianco di Delors al Consiglio Europeo di Cardiff*
- 1/2003 - Giovanni Maria Merola - *Safety Rules in Statistical Disclosure Control for Tabular Data*
- 2/2003 - Fabio Bacchini, Pietro Gennari e Roberto Iannaccone - *A new index of production for the construction sector based on input data*
- 3/2003 - Fulvia Ceroni e Enrica Morganti - *La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati*
- 4/2003 - Sara Mastrovita e Isabella Siciliani - *Effetti dei trasferimenti sociali sulla distribuzione del reddito nei Paesi dell'Unione europea: un'analisi dal Panel europeo sulle famiglie*
- 5/2003 - Patrizia Cella, Giuseppe Garofalo, Adriano Paggiaro, Nicola Torelli e Caterina Viviano - *Demografia d'impresa: l'utilizzo di tecniche di abbinamento per l'analisi della continuità*
- 6/2003 - Enrico Grande e Orietta Luzi - *Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat*
- 7/2003 - Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino - *Indagine sperimentale sui posti di lavoro vacanti*
- 8/2003 - Mario Adua - *L'agricoltura di montagna: le aziende delle donne, caratteristiche agricole e socio-rurali*
- 9/2003 - Franco Mostacci e Roberto Sabbatini - *L'euro ha creato inflazione? Changeover e arrotondamenti dei prezzi al consumo in Italia nel 2002*
- 10/2003 - Leonello Tronti - *Problemi e prospettive di riforma del sistema pensionistico*
- 11/2003 - Roberto Gismondi - *Tecniche di stima e condizioni di coerenza per indagini infraannuali ripetute nel tempo*
- 12/2003 - Antonio Frenda - *Analisi delle legislazioni e delle prassi contabili relative ai gruppi di imprese nei paesi dell'Unione Europea*
- 1/2004 - Marcello D'Orazio, Marco Di Zio e Mauro Scanu - *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*
- 2/2004 - Giovanna Brancato - *Metodologie e stime dell'errore di risposta. Una sperimentazione di reintervista telefonica*
- 3/2004 - Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese - *Gli indici dei prezzi al consumo per sub popolazioni*
- 4/2004 - Leonello Tronti - *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*
- 5/2004 - Ugo Guarnera - *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il software Quis*
- 6/2004 - Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca - *La nuova funzione di analisi dei modelli implementata in Genesees v. 3.0*
- 7/2004 - Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca - *MAUSS (Multivariate Allocation of Units in Sampling Surveys): un software generalizzato per risolvere il problema dell'allocazione campionaria nelle indagini Istat*
- 8/2004 - Ennio Fortunato e Liana Verzicco - *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell'Istat*
- 9/2004 - Claudio Pauselli e Claudia Rinaldelli - *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*
- 10/2004 - Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli - *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un'analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*
- 11/2004 - Enrico Grande e Orietta Luzi - *Regression trees in the context of imputation of item non-response: an experimental application on business data*
- 12/2004 - Luisa Frova e Marilena Pappagallo - *Procedura di now-cast dei dati di mortalità per causa*
- 13/2004 - Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni - *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*
- 14/2004 - Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo - *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*
- 15/2004 - Elisa Berntsen - *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l'Archivio delle Unità Locali di Asia*
- 16/2004 - Salvatore F. Allegra e Alessandro La Rocca - *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*
- 17/2004 - Francesca R. Pogelli - *Un'applicazione del modello "Country Product Dummy" per un'analisi territoriale dei prezzi*
- 18/2004 - Antonia Manzari - *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*
- 19/2004 - Claudio Pauselli - *Intensità di povertà relativa: stima dell'errore di campionamento e sua valutazione temporale*
- 20/2004 - Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi - *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*
- 21/2004 - Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale - *Un modello di ottimizzazione per l'imputazione delle mancate risposte statistiche nell'indagine sui trasporti marittimi dell'Istat*

- 22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*
- 23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*
- 24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*
- 25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d'indagine*
- 26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*
- 27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell'occupazione regionale: 8° Censimento dell'industria e dei servizi 2001 7° Censimento dell'industria e dei servizi 1991*
- 28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*
- 29/2004 – Paolo Consolini – *L'indagine sperimentale sull'archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*
- 1/2005 – Fabrizio M. Arosio – *La stampa periodica e l'informazione on-line: risultati dell'indagine pilota sui quotidiani on-line*
- 2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*
- 3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*
- 4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L'imputazione delle mancate risposte in presenza di dati longitudinali: un'applicazione ai permessi di costruzione*
- 5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*
- 6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*
- 7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*
- 8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*
- 9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*
- 10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*
- 11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*
- 12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*
- 13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*
- 14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*
- 15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*
- 16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrociochi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*
- 17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*
- 18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*
- 19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*
- 20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*
- 21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*
- 22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*
- 1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*
- 2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*
- 3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*
- 4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*
- 5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*
- 6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*
- 7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*
- 8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*
- 9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*

- 10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*
- 11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*
- 12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcaro e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*
- 13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*
- 14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*
- 15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*
- 16/2006 – Carlo De Gregorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*
- 1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*
- 2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*
- 3/2007 – Carlo Lucarelli e Giampiero Ricci – *Working times and working schedules: the framework emerging from the new Italian lfs in a gender perspective*
- 4/2007 – Monica Scannapieco, Diego Zardetto e Giulio Barcaroli – *La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESSEES/SAS*
- 5/2007 – Giulio Barcaroli e Tiziana Pellicciotti – *Strumenti per la documentazione e diffusione dei microdati d'indagine: il Microdata Management Toolkit*
- 6/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 1^a giornata*
- 7/2007 – Raffaella Cianchetta, Carlo De Gregorio, Giovanni Seri e Giulio Barcaroli – *Rilevazione sulle Pubblicazioni Scientifiche Istat*
- 8/2007 – Emilia Arcaleni, e Barbara Baldazzi – *Vivere non insieme: approcci conoscitivi al Living Apart Together*
- 9/2007 – Corrado Peperoni e Francesca Tuzi – *Trattamenti monetari non pensionistici metodologia sperimentale per la stima degli assegni al nucleo familiare*
- 10/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 2^a giornata*

Documenti ISTAT(*)

- 1/2002 – Paolo Consolini e Rita De Carli - *Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati*
- 2/2002 – Stefania Macchia - *Sperimentazione, implementazione e gestione dell'ambiente di codifica automatica della classificazione delle Attività economiche*
- 3/2002 – Maria De Lucia - *Applicabilità della disciplina in materia di festività nel pubblico impiego*
- 4/2002 – Roberto Gismondi, Massimo Marciani e Mauro Giorgetti - *The italian contribution towards the implementation of an european transport information system: main results of the MESUDEMO project*
- 5/2002 – Olimpio Cianfarani e Sauro Angeletti - *Misure di risultato e indicatori di processo: l'esperienza progettuale dell'Istat*
- 6/2002 – Riccardo Carbini e Valerio De Santis – *Programma statistico nazionale: specifiche e note metodologiche per la compilazione delle schede identificative dei progetti*
- 7/2002 – Maria De Lucia – *Il CCNL del personale dirigente dell'area 1 e la valutazione delle prestazioni dei dirigenti*
- 8/2002 – Giuseppe Garofalo e Enrica Morganti – *Gruppo di lavoro per la progettazione di un archivio statistico sui gruppi d'impresa*
- 1/2003 – Francesca Ceccato, Massimiliano Tancioni e Donatella Tuzi – *MODSIM-P: Il nuovo modello dinamico di previsione della spesa pensionistica*
- 2/2003 – Anna Pia Mirto – *Definizioni e classificazioni delle strutture ricettive nelle rilevazioni statistiche ufficiali sull'offerta turistica*
- 3/2003 – Simona Spirito – *Le prestazioni assistenziali monetarie non pensionistiche*
- 4/2003 – Maria De Lucia – *Approfondimenti di alcune tematiche inerenti la gestione del personale*
- 5/2003 – Rosalia Coniglio, Marialuisa Cugno, Maria Filmeno e Alberto Vitalini – *Mappatura della criminalità nel distretto di Milano*
- 6/2003 – Maria Letizia D'Autilia – *I provvedimenti di riforma della pubblica amministrazione per l'identificazione delle "Amministrazioni pubbliche" secondo il Sec95: analisi istituzionale e organizzativa per l'anno 2000*
- 7/2003 – Francesca Gallo, Pierpaolo Massoli, Sara Mastrovita, Roberto Merluzzi, Claudio Pauselli, Isabella Siciliani e Alessandra Sorrentino – *La procedura di controllo e correzione dei dati Panel Europeo sulle famiglie*
- 8/2003 – Cinzia Castagnaro, Martina Lo Conte, Stefania Macchia e Manuela Murgia – *Una soluzione in-house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite*
- 9/2003 – Anna Pia Maria Mirto e Norina Salamone – *La classificazione delle strutture ricettive turistiche nella normativa delle regioni italiane*
- 10/2003 – Roberto Gismondi e Anna Pia Maria Mirto – *Le fonti statistiche per l'analisi della congiuntura turistica: il mosaico italiano*
- 11/2003 – Loredana Di Consiglio e Stefano Falorsi – *Alcuni aspetti metodologici relativi al disegno dell'indagine di copertura del Censimento Generale della Popolazione 2001*
- 12/2003 – Roberto Gismondi e Anna Rita Giorgi – *Struttura e dinamica evolutiva del comparto commerciale al dettaglio: le tendenze recenti e gli effetti della riforma "Bersani"*
- 13/2003 – Donatella Cangialosi e Rosario Milazzo – *Fabbisogni formativi degli Uffici comunali di statistica: indagine rapida in Sicilia*
- 14/2003 – Agostino Buratti e Giovanni Salzano – *Il sistema automatizzato integrato per la gestione delle rilevazioni dei documenti di bilancio degli enti locali*
- 1/2004 – Giovanna Brancato e Giorgia Simeoni – *Tesauri del Sistema Informativo di Documentazione delle Indagini (SIDI)*
- 2/2004 – Corrado Peperoni – *Indagine sui bilanci consuntivi degli Enti previdenziali: rilevazione, gestione e procedure di controllo dei dati*
- 3/2004 – Marzia Angelucci, Giovanna Brancato, Dario Camol, Alessio Cardacino, Sandra Maresca e Concetta Pellegrini – *Il sistema ASIMET per la gestione delle Note Metodologiche dell'Annuario Statistico Italiano*
- 4/2004 – Francesca Gallo, Sara Mastrovita, Isabella Siciliani e Giovanni Battista Arcieri – *Il processo di produzione dell'Indagine ECHP*
- 5/2004 – Natale Renato Fazio e Carmela Pascucci – *Gli operatori non identificati nelle statistiche del commercio con l'estero: metodologia di identificazione nelle spedizioni "groupage" e miglioramento nella qualità dei dati*
- 6/2004 – Diego Moretti e Claudia Rinaldelli – *Una valutazione dettagliata dell'errore campionario della spesa media mensile familiare*
- 7/2004 – Franco Mostacci – *Aspetti Teorico-pratici per la Costruzione di Indici dei Prezzi al Consumo*
- 8/2004 – Maria Frustaci – *Glossario economico-statistico multilingua*
- 9/2004 – Giovanni Seri e Maurizio Lucarelli – *"Il Laboratorio per l'analisi dei dati elementari (ADELE): monitoraggio dell'attività dal 1999 al 2004"*
- 10/2004 – Alessandra Nuccitelli, Francesco Bosio e Luciano Fioriti – *L'applicazione RECLINK per il record linkage: metodologia implementata e linee guida per la sua utilizzazione*
- 1/2005 – Francesco Cuccia, Simone De Angelis, Antonio Laureti Palma, Stefania Macchia, Simona Mastroluca e Domenico Perrone – *La codifica delle variabili testuali nel 14° Censimento Generale della Popolazione*
- 2/2005 – Marina Peci – *La statistica per i Comuni: sviluppo e prospettive del progetto Sisco.T (Servizio Informativo Statistico Comunale. Tavole)*
- 3/2005 – Massimiliano Renzetti e Annamaria Urbano – *Sistema Informativo sulla Giustizia: strumenti di gestione e manutenzione*
- 4/2005 – Marco Broccoli, Roberto Di Giuseppe e Daniela Pagliuca – *Progettazione di una procedura informatica generalizzata per la sperimentazione del metodo Microstrat di coordinamento della selezione delle imprese soggette a rilevazioni nella realtà Istat*
- 5/2005 – Mauro Albani e Francesca Pagliara – *La ristrutturazione della rilevazione Istat sulla criminalità minorile*
- 6/2005 – Francesco Altarocca e Gaetano Sberno – *Progettazione e sviluppo di un "Catalogo dei File Grezzi con meta-dati di base" (CFG) in tecnologia Web*

- 7/2005 – Salvatore F. Allegra e Barbara Baldazzi – *Data editing and quality of daily diaries in the Italian Time Use Survey*
- 8/2005 – Alessandra Capobianchi – *Alcune esperienze in ambito internazionale per l'accesso ai dati elementari*
- 9/2005 – Francesco Rizzo, Laura Vignola, Dario Camol e Mauro Bianchi – *Il progetto "banca dati della diffusione congiunturale"*
- 10/2005 – Ennio Fortunato e Nadia Mignolli – *I sistemi informativi Istat per la diffusione via web*
- 11/2005 – Ennio Fortunato e Nadia Mignolli – *Sistemi di indicatori per l'attività di governo: l'offerta informativa dell'Istat*
- 12/2005 – Carlo De Gregorio e Stefania Fatello – *L'indice dei prezzi al consumo dei testi scolastici nel 2004*
- 13/2005 – Francesco Rizzo e Laura Vignola – *RSS: uno standard per diffondere informazioni*
- 14/2005 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino – *Launching and implementing the job vacancy statistics*
- 15/2005 – Stefano De Francisci, Massimiliano Renzetti, Giuseppe Sindoni e Leonardo Tininini – *La modellazione dei processi nel Sistema Informativo Generalizzato di Diffusione dell'ISTAT*
- 16/2005 – Ennio Fortunato e Nadia Mignolli – *Verso il Sistema di Indicatori Territoriali: rilevazione e analisi della produzione Istat*
- 17/2005 – Raffaella Cianchetta e Daniela Pagliuca – *Soluzioni Open Source per il software generalizzato in Istat: il caso di PHPSurveyor*
- 18/2005 – Gianluca Giuliani e Barbara Boschetto – *Gli indicatori di qualità dell'Indagine continua sulle Forze di Lavoro dell'Istat*
- 19/2005 – Rossana Balestrino, Franco Garritano, Carlo Cipriano e Luciano Fanfoni – *Metodi e aspetti tecnologici di raccolta dei dati sulle imprese*
- 1/2006 – Roberta Roncati – www.istat.it (versione 3.0) *Il nuovo piano di navigazione*
- 2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*
- 3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*
- 4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*
- 5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*
- 6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*
- 7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*
- 8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*
- 9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*
- 10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*
- 11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*
- 12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*
- 13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*
- 14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*
- 15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*
- 16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*
- 17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*
- 18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*
- 19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*
- 20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*
- 1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*
- 2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*
- 3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*
- 4/2007 – Vincenzo Spinelli – *Processo di Acquisizione e Trattamento Informativo degli Archivi relativi al Modello di Dichiarazione 770*
- 5/2007 – Anna Di Carlo, Maria Picci, Laura Posta, Michaela Raffone, Giuseppe Stassi e Fiorella Tortora – *La progettazione dei Censimenti generali 2010-2011: 1 - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Antonia Manzari, Marilena Pappagallo e Alessandra Reale – *Indagine sulle Cause di Morte: Nuova procedura automatica per il controllo e la correzione delle variabili demo-sociali*
- 7/2007 – Maura Giacommo, Carlo Vaccari e Monica Scannapieco – *Indagine sulle Scelte Tecnologiche degli Istituti Nazionali di Statistica*
- 8/2007 – Lamberto Pizzicannella – *Sviluppo del processo di acquisizione e trattamento informatico degli archivi relativi al modello di dichiarazione 770. Anni 2004 – 2005*
- 9/2007 – Damiano Abbadini, Lorenzo Cassata, Fabrizio Martire, Alessandra Reale, Giuseppina Ruocco e Donatella Zindato – *La progettazione dei Censimenti generali 2010-2011 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*

- 10/2007 – Marco Fortini, Gerardo Gallo, Evelina Paluzzi, Alessandra Reale e Angela Silvestrini – *La progettazione dei censimenti generali 2010–2011 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*
- 11/2007 – Domenico Adamo, Damiana Cardoni, Valeria Greco, Silvia Montecolle, Sante Orsini, Alessandro Ortensi e Miria Savioli – *Strategie di correzione del questionario sulla qualità della vita dell'infanzia e dell'adolescenza. Indagine multiscopo sulle famiglie. Aspetti della vita quotidiana 2005*
- 12/2007 – Carlo Nappi – *Manuale per la preparazione di originali "ready to print"*