

# **Una soluzione in-house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite**

## **Autori**

*Cinzia Castagnaro\*, Martina Lo Conte\*, Stefania Macchia\*\*, Manuela Murgia\*\**

**2002**

### ***Sommario***

Questo documento delinea una strategia per la gestione delle indagini che adottano la tecnica di rilevazione CATI, alternativa rispetto a quella finora realizzata in Istat. Tale strategia consiste nel fatto di avvalersi di una ditta esterna esclusivamente per la realizzazione della rilevazione (la società mette quindi a disposizione le postazioni CATI e il personale selezionato per l'effettuazione delle interviste) e di fornire a quest'ultima tutto il software da utilizzarsi per la rilevazione (comprendente il modulo per la schedulazione delle chiamate, il questionario elettronico e la procedura che produce gli indicatori di monitoraggio), sviluppato in Istat e già testato mediante un apposito pre-test. Il lavoro mette in luce i numerosi vantaggi di questa soluzione e descrive la metodologia di progettazione della fase di rilevazione CATI, nonché il pacchetto software integrato che realizza queste funzioni, già messo a punto ed adottato per l'Indagine Campionaria sulle Nascite; tale pacchetto è generalizzabile per il suo utilizzo per altre indagini. Sono infine accennati i primi risultati ottenuti con questa strategia, nell'ambito della citata indagine.

### ***Abstract***

This document is aimed at describing a strategy for the management of CATI surveys, which is alternative to what Istat has realised until now. This strategy consists in relying on a private company only for interviewing (the company provides CATI workstations and interviewers) and in giving it all the software to be adopted, produced in Istat and already tested through an ad hoc pre-test (the calls scheduling module, the electronic questionnaire and the reports module to monitor interviewing). This work highlights the numerous advantages of this strategy and describes the methodology for planning the data collection phase of a CATI survey as well as the integrated software package - which supports all these functions - already developed and adopted during the Sample Survey on Births. Being a generalised software package it could be used for other CATI surveys. Some words will be finally spent also to describe the first results obtained with this strategy during the mentioned survey.

### **Indice**

#### **Introduzione**

**1 Le motivazioni di una strategia in-house**

**Introduzione**

**1 Le motivazioni di una strategia in-house**

**2 Le fasi di disegno e sviluppo dell'applicazione CATI in un'indagine sulle famiglie**

2.1 Reperimento dei numeri di telefono

2.2 I moduli di una rilevazione CATI

2.2.1 Il questionario elettronico

2.2.1.1 Opportunità della realizzazione di un pre-test e sue finalità

2.2.1.2. Il laboratorio CASIC per la realizzazione dei pre-test

2.2.1.3 Elementi emergenti dal pre-test per l'ottimizzazione del questionario elettronico

2.2.1.4 La progettazione e lo sviluppo del questionario elettronico

2.2.1.5 La codifica dei dati rilevati a testo libero

2.2.3 Il sistema di indicatori di monitoraggio della rilevazione

2.3 La definizione degli aspetti amministrativi per la realizzazione del progetto

2.4 L'effettuazione dell'indagine pilota

2.5 L'indagine a regime

**3 L'indagine campionaria sulle nascite**

3.1 L'acquisizione dei dati e i contenuti informativi

3.2 La lista universo e l'abbinamento dei numeri telefonici

3.3 Il disegno campionario

**4 Descrizione dell'applicazione sviluppata per l'Indagine Campionaria sulle Nascite**

4.1 Il questionario elettronico

4.1.1 Gli standard adottati per la gestione dello schermo

4.1.2 Gli standard adottati per la gestione degli errori

4.1.3 Gli standard adottati per la gestione delle non risposte parziali

4.1.4 La personalizzazione dei testi dei quesiti e dei messaggi d'errore

4.1.5 La scelta della lingua con cui condurre l'intervista e le annotazioni in corso di intervista

4.1.6 La codifica assistita delle variabili testuali

**4.2 Le fasi dell'applicazione CATI**

4.2.1 Controllo e caricamento dei nominativi campione

4.2.2 La schedulazione delle chiamate

4.2.3 Gli indicatori di monitoraggio e la funzione di scarico dei dati

4.2.4 L'"utility" per la riorganizzazione della base dati

4.3 Le fasi precedenti l'indagine a regime

4.3.1 Il pre-test

4.3.1.1 Principali risultati del pre-test

4.3.1.2 Il de-briefing con gli operatori nella fase di pre-test

4.3.2 Il briefing agli intervistatori

4.3.3 La fase pilota

4.4 L'indagine a regime

4.4.1 Alcuni problemi sorti in fase di rilevazione

**5 Principali risultati**

5.2 Analisi degli esiti finali di contatto

5.2.1 Analisi degli esiti per ricezione lettera e ripartizione geografica

5.2.2 Analisi della durata

5.2.3 Analisi dei "non so, non ricordo", "rifiuto a rispondere" e "altro-specificare"

5.3 Confronto con i dati della rilevazione degli iscritti in anagrafe per nascita

5.3 Modifiche di alcuni controlli soft (signal)

**6 Considerazioni conclusive**

**Introduzione**

La modalità di rilevazione CATI per le indagini statistiche si è nel tempo sempre più diffusa presso gli Istituti Nazionali di statistica (Chiaro, 1996), che adottano diverse strategie operative che vanno

La modalità di rilevazione CATI per le indagini statistiche si è nel tempo sempre più diffusa presso gli Istituti Nazionali di statistica (Chiaro, 1996), che adottano diverse strategie operative che vanno dal rendersi completamente autonomi nella realizzazione di tutta la rilevazione, sviluppando direttamente il software ed utilizzando personale interno per l'effettuazione delle interviste, al fatto di affidarsi completamente a ditte esterne.

Finora presso l'Istat è stata perseguita questa seconda strategia che, pur avendo sempre consentito la riuscita delle rilevazioni, ha spesso comportato una serie di problematiche che hanno a volte implicato una non completa ottimizzazione del processo.

Una delle maggiori difficoltà incontrate è derivata dal fatto che, spesso, le società cui è stato affidato il servizio, selezionate a seguito di gare d'appalto secondo un procedimento che deve rispettare apposite normative, avessero maturato le principali esperienze nel settore delle telemarketing, oppure di indagini di mercato, ben lontane come complessità dalle indagini statistiche.

Si trovavano quindi ad affrontare per la prima volta le problematiche inerenti articolati piani di controlli tra le variabili rilevate ed ad attuare strategie finalizzate a ridurre le mancate risposte.

L'impostazione del lavoro costava quindi un grande impegno da parte dei responsabili di indagine che non avevano sempre la garanzia di ottenere il prodotto con le caratteristiche qualitative attese; ci si riferisce, per esempio, alle fasi di sviluppo e test del questionario elettronico, nonché di sviluppo delle procedure finalizzate alla produzione automatica e giornaliera degli indicatori di monitoraggio della rilevazione.

Si è quindi pensato di sperimentare una nuova strategia che rappresenta una soluzione intermedia tra le due citate all'inizio: avvalersi di una ditta esterna esclusivamente per la realizzazione della rilevazione (la società mette quindi a disposizione le postazioni CATI e il personale selezionato per l'effettuazione delle interviste), ma fornire a quest'ultima tutto il software da utilizzarsi per la rilevazione (comprendente il modulo per la schedulazione delle chiamate, il questionario elettronico e la procedura che produce gli indicatori di monitoraggio), sviluppato in Istat e già testato mediante un apposito pre-test.

In questo lavoro, dopo aver citato i principali vantaggi e svantaggi di questa strategia innovativa, vengono delineate le fasi del disegno e dello sviluppo di un'applicazione CATI, con particolare riferimento alle indagini sulle famiglie, viene quindi descritta l'applicazione sviluppata per la nuova 'Indagine Campionaria sulle Nascite' ed infine presentati i principali risultati ottenuti nel corso della prima rilevazione.

## **1 Le motivazioni di una strategia in-house**

Questa soluzione, definita come intermedia tra le due estreme descritte nell'introduzione, adottata per la prima volta nel corso della 'Indagine Campionaria sulle Nascite', comporta una serie di vantaggi, quali:

- ◆ le attività di progettazione, sviluppo e test del questionario elettronico sono realizzate nell'ambito di una continua collaborazione tra il responsabile di indagine e gli esperti della tecnica di rilevazione, in modo da affinare ed ottimizzare le specifiche in corso d'opera in funzione del prodotto man mano realizzato, di cui si può immediatamente verificare la funzionalità;
- ◆ il responsabile di indagine ha il pieno controllo su tutto il processo di rilevazione e garanzia della qualità dei risultati;
- ◆ nel caso di indagini continue nel tempo, lo sforzo di progettazione, sviluppo e test del software viene fatto una unica volta, non presentandosi la necessità di ripeterlo ogni qual volta, a scadenza del contratto, cambi la società che realizza la rilevazione;
- ◆ è facilitata la selezione della ditta che dovrà effettuare la rilevazione, non essendo necessario fare una valutazione approfondita delle esperienze pregresse, ma richiedendo una serie dettagliata di requisiti inerenti l'architettura informatica, di immediato riscontro;
- ◆ si avrebbe una riduzione dei costi inerenti l'incidenza dello sviluppo del software che le società non sono più chiamate a realizzare. A tale proposito, è difficile valutare l'entità di questa riduzione, in quanto finora le ditte hanno realizzato il servizio sulla base di un costo unitario per intervista che comprende lo sviluppo del software, l'utilizzo dei locali, il costo delle telefonate e le ore uomo impegnate sul progetto. Tuttavia si può ipotizzare che l'incidenza di tale costo sia direttamente proporzionale alla complessità del questionario e della reportistica da sviluppare ed inversamente proporzionale al numero di interviste da effettuare.

D'altro canto, non si può ignorare che tale soluzione implichi la necessità di disporre di almeno una risorsa esperta sul software utilizzato per lo sviluppo, che sia sempre disponibile durante la rilevazione per eventuali aggiornamenti migliorativi o correttivi dell'applicazione installata presso

D'altro canto, non si può ignorare che tale soluzione implichi la necessità di disporre di almeno una risorsa esperta sul software utilizzato per lo sviluppo, che sia sempre disponibile durante la rilevazione per eventuali aggiornamenti migliorativi o correttivi dell'applicazione installata presso la società esterna. Dal punto di vista economico, invece, il costo aggiuntivo è esclusivamente quello delle licenze del software CATI utilizzato che, nel caso di quello finora adottato, come si vedrà di seguito in dettaglio, incide in maniera estremamente limitata.

## 2 Le fasi di disegno e sviluppo dell'applicazione CATI in un'indagine sulle famiglie

Le principali fasi del disegno e dello sviluppo di un'applicazione CATI, così come si vedrà di seguito in dettaglio, possono ricondursi allo schema qui riportato.

### 2.1 Reperimento dei numeri di telefono

Premesso che si considera che la decisione sul piano di campionamento sia già stata presa nel momento in cui ci si trova a progettare l'applicazione CATI, le alternative che possono presentarsi sono:

- ◆ l'estrazione dei numeri di telefono dagli elenchi;
- ◆ l'estrazione dei nominativi dall'anagrafe ed il successivo abbinamento degli stessi con i numeri telefonici.

Fino ad ora, nelle indagini Istat, sia l'estrazione dei numeri di telefono dagli elenchi che l'abbinamento dei nominativi anagrafici con i numeri sono stati delegati alle ditte incaricate di effettuare la rilevazione.

Relativamente all'estrazione dei numeri di telefono dagli elenchi, generalmente le società già dispongono di appositi algoritmi; comunque, qualora il responsabile d'indagine ne prediliga in particolare uno di quelli noti in letteratura (Chiaro, 1996), lo può esplicitare nelle specifiche tecniche.

L'abbinamento 'nominativi – numeri di telefono' è un'attività estremamente delicata; è ovvio infatti che a priori non tutti i nominativi siano abbinabili, in quanto non tutti possiedono un telefono fisso e, anche tra coloro che lo possiedono, ci sono le utenze riservate.

Il riuscire ad attribuire il numero di telefono ai nominativi che lo possiedono dipende poi da una serie di fattori, quali:

- ◆ l'aggiornamento degli elenchi telefonici;
- ◆ l'aggiornamento degli elenchi anagrafici;
- ◆ la corrispondenza tra il titolare del contratto telefonico ed il nominativo selezionato, piuttosto che con un altro componente della famiglia;
- ◆ l'assenza di errori di registrazione nell'acquisizione del nominativo da abbinare e del relativo indirizzo.

Generalmente questa attività è realizzata in due fasi: una prima elaborazione in *batch* per velocizzare l'abbinamento di quanti più nominativi possibile, ed un passaggio manuale in cui l'operatore analizza i casi che non è stato possibile risolvere automaticamente, a causa per esempio di errori ortografici non prevedibili a priori.

Per l'ottimizzazione dei risultati è opportuno quindi tenere presente una serie di fattori:

- ◆ prevedere la registrazione non soltanto del nominativo anagrafico selezionato, ma anche di altri componenti della famiglia, in quanto potrebbero essere titolari del contratto telefonico;
- ◆ laddove la registrazione sia di responsabilità dell'Istat, dare specifiche dettagliate, in modo da uniformare le abbreviazioni utilizzate più ricorrentemente;
- ◆ definire a priori i criteri da adottare per l'abbinamento, nel caso non si verifichi un accoppiamento esatto tra le informazioni anagrafiche (nominativo, comune di residenza ed indirizzo) e gli elenchi telefonici;
- ◆ prevedere un riscontro successivo tra nominativi abbinati e criterio adottato per l'abbinamento, in funzione del fatto che il numero sia risultato corretto o meno, in modo da affinare i criteri di abbinamento da adottare successivamente;
- ◆ definire delle percentuali di abbinamento minime per strato, in modo da insistere di più nella ricerca manuale per gli strati maggiormente a rischio.

Sempre in funzione della massimizzazione del numero di nominativi abbinati, è inoltre possibile (ed è già stato effettuato nell'applicazione descritta in questo lavoro) effettuare una seppur grezza normalizzazione degli indirizzi, utilizzando alcune funzioni previste dal software ACTR (*Automatic Coding by Text Recognition*). Tale sistema, infatti,

Sempre in funzione della massimizzazione del numero di nominativi abbinati, è inoltre possibile (ed è già stato effettuato nell'applicazione descritta in questo lavoro) effettuare una seppur grezza normalizzazione degli indirizzi, utilizzando alcune funzioni previste dal software ACTR (*Automatic Coding by Text Recognition*). Tale sistema, infatti, studiato e realizzato da Statistics Canada (Wenzowski, 1988) e finalizzato alla codifica automatica di risposte rilevate a testo libero, prevede un passaggio di standardizzazione dei testi. Questo passaggio è realizzato attraverso alcune funzioni (fino ad un massimo di 14) attraverso le quali l'utente può definire ad esempio, nell'ambito delle stringhe di testo da analizzare, quali siano i delimitatori, le parti di testo ininfluenti, può ricondurre più abbreviazioni ad una parola estesa, definire sinonimi a livello di parola o di coppia di parole, ecc..(ACTRv3 User Guide).

ACTR è già diffusamente adottato in Istat per la codifica di una serie di variabili (Macchia, Murgia, 2002) e, relativamente alla problematica in esame, può essere più limitatamente utilizzato per ricondurre i diversi modi di rappresentare un toponimo ad una singola grafia, aumentando di conseguenza le probabilità di abbinamento degli indirizzi dei nominativi selezionati con quelli degli elenchi telefonici.

In generale, la funzione di abbinamento potrebbe essere tra quelle realizzate all'interno, previo ovviamente l'acquisto degli elenchi telefonici; in tal caso, si potrebbe pensare ad una procedura generalizzata, adottabile per tutte le indagini telefoniche sulle famiglie che estraggano i nominativi dalle liste anagrafiche.

Qualora, invece, sia delegata alla ditta incaricata della rilevazione, è bene esplicitare a priori:

- ♦ una percentuale massima di non abbinati sul totale dei nominativi forniti (stimabile, sulla base delle esperienze effettuate, di circa il 40%) e delle più dettagliate percentuali per strato;
- ♦ una percentuale massima di abbinati risultati successivamente errati in fase di rilevazione (stimabile, sulla base delle esperienze effettuate, di circa il 7%).

## 2.2 I moduli di una rilevazione CATI

Nella seguente figura sono riportati i principali moduli di una rilevazione CATI. Si procederà in primo luogo con l'esposizione dell'aspetto di maggiore rilievo, inerente il questionario elettronico.

### 2.2.1 Il questionario elettronico

Prima di scendere in dettaglio su come realizzare un questionario elettronico in maniera ottimale, è opportuno puntualizzare che è impossibile ottenere un prodotto di qualità se il questionario di indagine non è già stato pensato in funzione della tecnica di rilevazione da adottare.

Il trasporre infatti su supporto elettronico per una rilevazione CATI un questionario pensato per una rilevazione cartacea, per quanto ci si possa avvalere di un software di elevata qualità e di informatici in grado di utilizzarlo al meglio

- genererà molto probabilmente problemi in termini di pesantezza dell'intervista,
- faciliterà la possibilità di rilevare dati affetti da errori non campionari dovuti, per esempio, ad una non adatta formulazione dei quesiti oppure ad una eccessiva elencazione di modalità di risposta,
- potrà incidere negativamente sui tassi di risposta.

#### 2.2.1.1 Opportunità della realizzazione di un pre-test e sue finalità

Fatta questa premessa, si vogliono qui sottolineare i vantaggi derivanti dalla realizzazione di un pre-test della rilevazione, tramite cui il responsabile dell'indagine può focalizzare, ancora prima dell'indagine pilota, i pregi ed i limiti del questionario somministrandolo ad un campione rappresentativo di unità.

Fatta questa premessa, si vogliono qui sottolineare i vantaggi derivanti dalla realizzazione di un pre-test della rilevazione, tramite cui il responsabile dell'indagine può focalizzare, ancora prima dell'indagine pilota, i pregi ed i limiti del questionario somministrandolo ad un seppur contenuto campione di unità.

Le finalità del pre-test sono fundamentalmente diverse da quelle tipiche di un'indagine pilota e devono limitarsi a:

- l'ottimizzazione del questionario elettronico;
- la stima della durata delle interviste.

Quest'ultimo dato è estremamente rilevante sotto diversi aspetti, quali:

- se la durata risulta essere eccessivamente lunga per un'intervista telefonica, è opportuno non soltanto ottimizzare il questionario elettronico in tal senso, ma soprattutto riflettere sul contenuto informativo che si vuole rilevare, sforzandosi di limitarlo agli aspetti essenziali;
- se il pre-test è svolto con un certo anticipo rispetto all'indagine a regime, permette di individuare meglio i parametri di prezzo e caratteristiche della ditta esterna necessari per la redazione del capitolato tecnico, infatti:
  - la durata incide sul costo della rilevazione (costo unitario per intervista);
  - dato un periodo stabilito per la rilevazione, la durata incide sul numero di postazioni CATI e di intervistatori da richiedere alla ditta che effettuerà la rilevazione.

Relativamente, però, alla durata emergente dal pre-test, bisogna tener conto che questa risulta spesso essere leggermente sovra stimata, in quanto, a regime, gli intervistatori prendono familiarità con il questionario, ne conoscono meglio il contenuto informativo e riescono a ridurre i tempi di intervista.

### 2.2.1.2. Il laboratorio CASIC per la realizzazione dei pre-test

Per la realizzazione del pre-test, è possibile avvalersi in Istat della struttura e delle competenze del laboratorio CASIC (*Computer Assisted Survey Information Collection*), nell'ambito della quale si instaura una collaborazione tra gli statistici e gli esperti di tecniche di rilevazione ai fini dello sviluppo di un prototipo di questionario elettronico, della realizzazione di un contenuto numero di interviste e dell'analisi dei risultati.

Il laboratorio CASIC è funzionalmente suddiviso in due settori: uno inerente le tecniche di rilevazione assistite da computer che non richiedono la figura del rilevatore, quali, per esempio la tecnica *CASI (Computer Assisted Self Interviewing)* ed uno dedicato specificamente alle tecniche CATI, CAPI (*Computer Assisted Personal Interviewing*) e CADI (*Computer Assisted Data Input*). Il primo settore afferisce alla Direzione Centrale dell'informatica e delle telecomunicazioni, mentre il secondo al servizio Metodologie per la Produzione Statistica, in particolare all'unità operativa 'Metodologie per l'acquisizione dei dati assistita da computer e per la codifica automatica'.

Per quanto attiene le rilevazioni telefoniche, la struttura dispone di 5 postazioni CATI (personal computer separati tra di loro da appositi pannelli divisorii e dotati di altrettanti telefoni con linee esterne dirette).

Il software utilizzato per le rilevazioni telefoniche è BLAISE, studiato e commercializzato da Statistics Netherlands e già ampiamente diffuso presso numerosi uffici nazionali di statistica.

Tale sistema, che è operativo su P.C. in ambiente WINDOWS95, WINDOWS NT 4.0 e superiori, realizza l'acquisizione dei dati sia in modalità CADI, CAPI e CATI e non richiede lunghi tempi di apprendimento (BLAISE Developer's Guide, 1999).

Per la realizzazione delle interviste ci si avvale di personale Istat, per la gran parte messo a disposizione dal servizio responsabile dell'indagine e, limitatamente, dall'unità operativa 'Metodologie per l'acquisizione dei dati assistita da computer e per la codifica automatica'; le interviste vengono effettuate in orario d'ufficio.

Al fine di fornire un ordine di grandezza, si segnala che, ai fini del pre-test, vengono generalmente realizzate un numero di interviste che varia da un minimo di 50 ad un massimo di 200-300.

Per informazione, si rende noto che, fino ad oggi, sono stati realizzati nell'ambito del laboratorio CASIC i pre-test delle seguenti indagini:

- prima indagine pilota sulle Forze di Lavoro
- sperimentazione indagine di qualità del Censimento della Popolazione e delle Abitazioni
- Nuova Indagine Campionaria sulle Nascite
- Indagine sulla Violenza alle Donne.

### 2.2.1.3 Elementi emergenti dal pre-test per l'ottimizzazione del questionario elettronico

Relativamente al questionario elettronico sviluppato per il pre-test, è bene puntualizzare che, sotto alcuni aspetti questo può non essere curato, come dovrà esserlo il questionario da utilizzarsi per la rilevazione vera e propria, quali, per esempio:

- l'aspetto grafico;
- la formulazione della messaggistica d'errore;
- la gestione di un solo blocco di domande, quando questo potrebbe ripetersi più volte, ad esempio in caso di interviste da effettuarsi su tutti i componenti della famiglia.

D'altra parte, però, al fine di ottenere tutti gli elementi per la valutazione dei risultati, è indispensabile che:

- contenga tutti i quesiti del questionario d'indagine;
- siano previsti tutti i controlli;
- sia data la possibilità di raccogliere tutte le osservazioni dei rilevatori sull'andamento dell'intervista.

A seguito del pre-test, si dovrebbe disporre di una serie di elementi, quali:

- la fluidità della struttura del questionario in termini di percorsi;
- la necessità di migliorare il wording di alcuni quesiti;
- l'opportunità di consentire risposte spontanee, piuttosto che fornire un'elencazione di modalità di risposta predefinite;

- ♦ la fluidità della struttura del questionario in termini di percorsi;
- ♦ la necessità di migliorare il wording di alcuni quesiti;
- ♦ l'opportunità di consentire risposte spontanee, piuttosto che fornire un'elencazione di modalità di risposta predefinite;
- ♦ indicazioni di massima sull'eshaustività delle modalità di risposta previste;
- ♦ l'opportunità di estendere le modalità di risposta 'non so/non ricordo', 'rifiuto a rispondere' a quesiti per i quali non era stato previsto;
- ♦ elementi per il miglioramento del modulo di presentazione dell'intervista;
- ♦ la verifica sulla bontà della scelta sulla modalità di codifica delle variabili (è perseguibile la codifica assistita in fase di intervista o è preferibile la codifica successiva alla rilevazione, con l'opportunità di adottare la codifica automatica?).

#### 2.2.1.4 La progettazione e lo sviluppo del questionario elettronico

Come già detto, per lo sviluppo del questionario elettronico viene utilizzato il software Blaise, che consente di realizzare una serie di funzioni indispensabili per un prodotto di qualità, sia da un punto di vista puramente tecnico (tempi di risposta del sistema e semplicità di gestione per l'intervistatore) che sotto l'aspetto più generale funzionale alla fluidità dell'intervista. In fase di progettazione devono quindi essere definite una serie di specifiche che consentano di sfruttare al meglio le potenzialità offerte dal sistema.

Scendendo più in dettaglio, è possibile realizzare:

- La personalizzazione del wording dei quesiti in funzione delle caratteristiche note del rispondente (dati anagrafici) e di risposte già fornite in quesiti precedenti; ciò consente che l'intervista assuma un carattere meno distaccato e più colloquiale.
- Il routing automatico nell'ambito del questionario; è questa una funzione essenziale in un questionario elettronico che solleva completamente il rilevatore dalla gestione dei salti condizionati tra un quesito ed i successivi, inibendo la possibilità di generare errori.
- La gestione dei controlli al fine di prevenire errori in fase di acquisizione; i controlli possono essere inerenti sia il rispetto dei *range* delle singole variabili, che la coerenza tra la risposta fornita ad un quesito e quelle fornite in quesiti precedenti.
- Diverse modalità di trattamento degli errori; è possibile infatti inibire radicalmente l'acquisizione di un dato errato, oppure limitarsi alla segnalazione dell'errore che non necessariamente deve essere rimosso. La prima modalità, chiamata tecnicamente '*hard mode*' è automaticamente applicata dal sistema per i controlli di *range* delle variabili, mentre per i controlli di coerenza è applicata secondo quanto deciso da chi sviluppa il questionario, che può anche optare di ricorrere alla seconda modalità, chiamata tecnicamente '*soft mode*'; tendenzialmente la soluzione ottimale dovrebbe essere quella di trovare un equilibrio tra una soluzione completamente rigida, che inibisce qualsiasi errore, ma rischia di compromettere la conclusione dell'intervista ed una soluzione troppo elastica che lascia all'intervistatore la possibilità di oltrepassare le segnalazioni di errore e di acquisire dati affetti da troppe incoerenze. E' bene quindi procedere a definire un set di controlli fondamentali per la qualità dei dati rilevati, da trattare quindi in modalità *hard* e adottare per gli altri la modalità *soft*.
- Una gestione particolare delle modalità di risposta 'Non so/non ricordo' e 'Rifiuto'; infatti il sistema consente innanzi tutto di selezionare quali siano le variabili per le quali ammettere una (o entrambe) le modalità. Come nel punto precedente, questa scelta deve essere fatta nell'ottica di mediare tra un uso indiscriminato di queste modalità ed il fatto che, non ammettendole, si compromette la prosecuzione dell'intervista, tale che il rilevatore, potrebbe sentirsi indotto ad inventare una risposta in modo da non perdere l'intervista. Inoltre è possibile gestire tali modalità in modo che, pur rendendole ammissibili, non vengano visualizzate tra le altre modalità di risposta, cosicché l'intervistatore non sia tentato di farne abuso; nel caso, infatti, in cui se ne voglia far uso, si rende necessario ricorrere ad appositi tasti funzione.
- L'accesso a dati esterni in corso di intervista. E' possibile infatti accedere a dati già disponibili (quali, per esempio, informazioni di carattere anagrafico oppure dati rilevati nel corso di precedenti rilevazioni) ed utilizzarli come termini di riferimento per appositi controlli di coerenza, che, tendenzialmente, vengono adottati in modalità *soft*.
- La codifica assistita. Il sistema Blaise prevede una funzione di codifica assistita, tramite un algoritmo per il trattamento dei testi che si basa sull'analisi di bigrammi e trigrammi e si applica ad un patrimonio informativo (il cosiddetto dizionario) predisposto dall'utente per ciascuna classificazione.
 

Nel momento in cui il 'dizionario' viene caricato nel sistema BLAISE, questo ne organizza una struttura ad albero, basandosi sul sistema di codici della classificazione di riferimento.

E' consentita poi la navigazione nel dizionario secondo tre modalità:

  - secondo l'albero  $\hat{a}$  il codificatore visualizza in un primo tempo la classificazione a livello dei rami gerarchicamente più elevati, quindi seleziona il ramo ed entra nei codici specifici del livello inferiore e così via;
  - secondo la dizione alfabetica  $\hat{a}$  il codificatore digita un testo ed il sistema ricerca ed estrae dal dizionario i testi che hanno in comune con quello digitato uno o più trigrammi (in caso il testo digitato sia identico ad uno di quelli contenuti nel dizionario, viene estratto questo singolo testo ed assegnato automaticamente il codice);
  - secondo una procedura mista  $\hat{a}$  il codificatore seleziona un ramo della classificazione e quindi effettua una ricerca testuale nell'ambito di quel ramo.
- La gestione dello schermo in modo da definire degli standard che facilitino al massimo il lavoro dell'intervistatore e ne riducano quello che in letteratura è noto come 'effetto segmentazione', tipico della gestione dei questionari sullo schermo di un personal computer. Come si vedrà nel capitolo 4, Blaise già prevede degli standard nella gestione dello schermo sia in termini di colori che di porzioni di schermo da dedicare a diverse funzioni (per esempio il wording dei quesiti, la visualizzazione di sequenze di quesiti del questionario, la messaggistica di errore), che possono comunque essere personalizzati in funzione delle peculiarità di ciascuna rilevazione.
- L'interfaccia verso altri software. Il sistema Blaise non soltanto consente di scaricare, in formato ASCII, i dati rilevati e memorizzati nel proprio ambiente applicativo, ma è in grado di generare automaticamente i tracciati record relativi a ciascun questionario nei formati SAS ed SPSS, per i quali crea anche tutte le *label* e le etichette

errore), che possono comunque essere personalizzati in funzione delle peculiarità di ciascuna rilevazione.

- i) **L'interfaccia verso altri software.** Il sistema Blaise non soltanto consente di scaricare, in formato ASCII, i dati rilevati e memorizzati nel proprio ambiente applicativo, ma è in grado di generare automaticamente i tracciati record relativi a ciascun questionario nei formati SAS ed SPSS, per i quali crea anche tutte le *label* e le etichette delle variabili incluse nel questionario.

### 2.2.1.5 La codifica dei dati rilevati a testo libero

Si è accennato nel paragrafo precedente alla possibilità offerta dal sistema Blaise di codificare nel corso dell'intervista. Indubbiamente, tale modalità di codifica presenta vantaggi di non poco conto, quali:

- la possibilità di richiedere ulteriori dettagli all'intervistato, qualora la risposta fornita sia carente ai fini della codifica;
- eliminare la possibilità di errore di registrazione del codice;
- ridurre i tempi che sarebbero necessari per la codifica successiva rispetto alla rilevazione.

Tuttavia bisogna tener presente che tale funzione, se adottata per una variabile cui è associata una classificazione caratterizzata da un elevato grado di complessità, può sviluppare ed appesantire l'intervista. La scelta deve quindi tener conto di entrambi gli aspetti: la complessità della classificazione da adottare.

Nel caso in cui si opti per la codifica in un momento successivo rispetto all'intervista, ci si può comunque avvalere del supporto del computer, adottando diversi software, a seconda della strategia prescelta.

Le soluzioni possibili sono:

- affidarsi ad un funzionario esperto, avvalendosi comunque del modulo di Blaise per la codifica assistita;
- adottare un altro sistema di codifica disponibile in Istat, qualora si opti per la codifica automatica dei dati (in modalità *batch*). In particolare, si tratta del già citato software ACTR (*Automatic Coding by Text Recognition*), già adottato in Istat per la codifica di numerose variabili quali: Professione, Attività economica, Titolo di studio, Provincia/Comune, Stato estero/Cittadinanza e Natura giuridica delle imprese (Macchia, Mastroluca, Reale, 2001). Di questa strategia bisogna comunque tener conto già in fase di rilevazione, in quanto è opportuno dare agli intervistatori specifiche istruzioni sulla registrazione dei dati testuali tali da favorire il successo della codifica ex-post; per esempio, scoraggiare l'utilizzo di abbreviazioni, prestare particolare attenzione agli errori ortografici o, soprattutto per le variabili Professione ed Attività Economica, evitare la declinazione dei verbi a favore dei sostantivi. Questa modalità di codifica, per sua natura, molto improbabilmente riuscirà ad attribuire un codice al 100% dei testi da codificare (la percentuale sarà tanto più alta, quanto più la classificazione di riferimento è semplice e la variabilità del wording delle risposte è limitata) (Macchia, Murgia, 2001), quindi, per i casi non risolti automaticamente, ci si dovrà avvalere del codificatore esperto, eventualmente supportato dall'applicazione Blaise.

### 2.2.2 Il modulo di schedulazione delle chiamate

L'assegnazione automatica dei nominativi ai rilevatori e le funzioni che devono condizionare l'attivazione di un numero piuttosto che di un altro sono elementi fondamentali per qualsiasi sistema CATI, dai quali dipende non soltanto l'ottimizzazione dell'attività di rilevazione, ma soprattutto la possibilità di minimizzare le mancate risposte totali.

Blaise realizza automaticamente l'attribuzione casuale dei numeri telefonici agli operatori e prevede appositi moduli per la schedulazione delle telefonate, che possono essere opportunamente personalizzati tramite l'impostazione di valori diversi da quelli di *default* dei parametri di sistema.

Senza entrare in dettaglio sulle modalità di funzionamento del sistema (si veda a tal proposito il capitolo 4), si fa qui un accenno dei principali parametri da fornire, una volta definito il periodo di rilevazione, quali:

- la quantità di numeri di telefono da contattare nella giornata (il numero di record del campione da caricare nel cosiddetto *daybatch*);
- le fasce orarie durante le quali effettuare le telefonate;
- gli intervalli di tempo che devono passare tra un tentativo di contatto che ha dato esito provvisorio ed il successivo tentativo (esempio: dopo quanto tempo riprovare un numero che è risultato occupato, oppure libero, oppure segreteria/fax);
- quanti tentativi di contatto debbano essere fatti per ciascun numero in un giorno e per quanti giorni si debba continuare;
- se si debba vincolare che un numero che abbia dato come esito 'appuntamento' debba ritornare allo stesso operatore oppure ritorni casualmente a chiunque.

E' importante sottolineare che la determinazione di questi parametri debba essere definita in funzione di una serie di elementi quali:

- esperienze effettuate in precedenti indagini sulla stessa tipologia di rispondente;
- durata dell'intervista;
- numero di intervistatori a disposizione;
- i periodi dell'anno nei quali si effettuano le interviste.

L'ipotesi iniziale deve comunque essere sperimentata nel corso dell'indagine pilota, che senz'altro fornirà elementi utilissimi per un affinamento della configurazione del sistema. Anche a regime, poi, l'analisi degli indicatori di monitoraggio giornaliero potrà implicare un'ulteriore ottimizzazione dei parametri.

E' infine opportuno definire a chi debba spettare la gestione dei parametri per la schedulazione delle chiamate. Dal



utilissimi per un affinamento della configurazione del sistema. Anche a regime, poi, l'analisi degli indicatori di monitoraggio giornaliero potrà implicare un'ulteriore ottimizzazione dei parametri.

E' infine opportuno definire a chi debba spettare la gestione dei parametri per la schedulazione delle chiamate. Dal momento che, secondo la strategia delineata nel paragrafo 1, si richiede alla società esterna di essere soltanto un utente finale del sistema software fornito dall'Istat, si ritiene che la soluzione più coerente sia quella di non delegare la responsabilità dell'impostazione e della modifica dei parametri di sistema, così da evitare eventuali contenziosi con la ditta che potrebbe essere tentata di modificare i parametri, per esempio riducendo o aumentando il numero di tentativi di contatto da effettuarsi, per ridurre i tempi di indagine.

### 2.2.3 Il sistema di indicatori di monitoraggio della rilevazione

Oltre al monitoraggio sul campo, che ovviamente può difficilmente essere effettuato continuamente durante tutto il periodo di rilevazione, soprattutto quando quest'ultima si estende per lunghi periodi di tempo, l'elemento fondamentale per verificare il buon andamento della rilevazione è l'analisi giornaliera degli indicatori di monitoraggio.

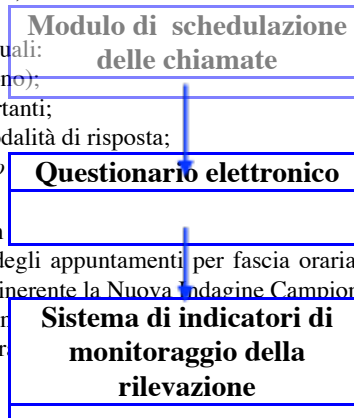
Il set minimo di indicatori deve essere funzionale a tenere sotto controllo, quotidianamente, diversi aspetti, quali:

- l'andamento della rilevazione dal punto di vista quantitativo (numero di interviste concluse positivamente, numero di interviste con esito rifiuto o interruzione definitiva, numero di interviste con esito provvisorio);
- l'andamento della rilevazione dal punto di vista qualitativo (performance dei singoli rilevatori, analisi dei dati rilevati in funzione delle risposte acquisite su particolari quesiti);
- l'adeguatezza dei parametri di schedulazione delle chiamate.

A tal fine si può pensare ad uno schema minimo di indicatori, validi per qualsiasi rilevazione, che comprenda:

- alcuni indicatori più significativi, quali:
  - indicatori per ultimo esito (concluse, sospese, cadute);
  - durata;
  - esiti negativi per motivo;
  - interruzioni per sezione di questionario;
  - livello di riempimento delle quote.
- alcuni indicatori di ulteriore dettaglio, quali:
  - storia degli esiti (per numero di telefono);
  - distribuzione delle variabili più importanti;
  - controllo delle variabili con molte modalità di risposta;
  - concentrazione di 'non sol/non ricordo'.

Il sistema Blaise genera automaticamente un report che riassume essenzialmente gli ultimi esiti cumulati e per intervistatore, nonché la distribuzione degli appuntamenti per fascia oraria e priorità; tuttavia, come ampiamente descritto nel paragrafo 4, per l'applicazione inerente la Nuova Indagine Campionaria sulle Nascite, è stato sviluppato un set di indicatori in Visual Basic, basato su un database Access e Excel, facilmente adattabile ad altre indagini, tramite la modifica di opportuni parametri.



### 2.3 La definizione degli aspetti amministrativi per la realizzazione del progetto

Nel paragrafo precedente ci si è addentrati su aspetti tecnici inerenti i moduli software per la realizzazione di una rilevazione CATI secondo la strategia, denominata *in-house*, in funzione del sistema software in uso presso l'Istat, al fine di focalizzare sugli elementi e le specifiche che devono essere definite in fase di progettazione.

Tuttavia, volendo ricondursi ad uno schema per fasi, una volta che sia stata stabilita la modalità di reperimento dei numeri di telefono, che sia stato progettato il questionario elettronico, sviluppato un prototipo e realizzato un pre-test (avvalendosi della struttura del laboratorio CASIC), non soltanto si può procedere allo sviluppo del questionario da utilizzarsi a regime per l'indagine, ma si dispone di tutti gli elementi da inserire nel capitolato tecnico da utilizzarsi per l'espletamento delle procedure amministrative per la selezione della società che dovrà effettuare la rilevazione.

In particolare, come già accennato, si disporrà di nozioni più precise sulla durata delle interviste che consentirà di definire:

- numero di postazioni da richiedere;
- numero di intervistatori da utilizzare;
- il prezzo unitario per intervista.

Inoltre, per quanto attiene i requisiti tecnici legati all'utilizzo del sistema Blaise, la ditta esterna dovrà disporre di un ambiente applicativo con i seguenti requisiti minimi di sistema:

- rete di Pc in ambiente Windows 95, Windows NT 4.0 o superiori, il cui server dovrà avere:
  - Processore Pentium II o superiori
  - Memoria RAM a 256 Mb (in realtà Blaise richiede un minimo di 64Mb, ma è stato verificato che una RAM così dimensionata garantisce performance nettamente migliori)
- relativamente alla configurazione della rete che collega queste postazioni, poiché l'applicazione *client-server* usa come protocollo di trasmissione TCP/IP su Netbeui, questa dovrà essere dimensionata in modo tale da supportare l'interattività tra tutti i client ed il server; in particolare le stazioni di lavoro dovranno essere collegate ad una velocità di 100 Mbit/s (tecnologia Fast/Ethernet) ed il server collegato ad uno switch di centro stella in modo da evitare problemi di contesa con altre stazioni di lavoro (eventuali collisioni). E' preferibile poi che, durante la fase di rilevazione, tale rete sia dedicata alla rilevazione Istat, oppure che, con accorgimenti tecnici (vedi Virtual Lan),

l'interattività tra tutti i client ed il server; in particolare le stazioni di lavoro dovranno essere collegate ad una velocità di 100 Mbit/s (tecnologia Fast/Ethernet) ed il server collegato ad uno switch di centro stella in modo da evitare problemi di contesa con altre stazioni di lavoro (eventuali collisioni). E' preferibile poi che, durante la fase di rilevazione, tale rete sia dedicata alla rilevazione Istat, oppure che, con accorgimenti tecnici (vedi Virtual Lan), si faccia in modo che il dominio di broadcast comprenda solamente le stazioni client più il server, utilizzati dall'Istat.

Infine, relativamente alle licenze Blaise, si puntualizza che esistono due tipi di licenze: quelle per chi sviluppa (*developer licence*) e quelle da utente finale, differenziate a seconda della tecnica di rilevazione *adottata* (*end-user licences*). Poiché la titolarità di tutte le licenze è dell'Istat, che risponde del fatto che la ditta che effettua la rilevazione le utilizzi soltanto per il progetto in essere, è consigliabile installare presso la ditta esclusivamente le licenze da utente finale che non consentono l'utilizzo del software per altre finalità.

A partire da quest'anno le licenze Blaise sono state acquistate dall'Istat secondo un accordo chiamato 'Corporate licence' che abilita all'utilizzo di un massimo di 75 licenze, laddove le *developer licence* valgono 1 e le *end-user licence* valgono 0,25. Si è ritenuto che tale soluzione sia più vantaggiosa, in quanto, non richiedendo acquisti ad hoc per ogni singola indagine, rende più flessibile l'utilizzo del software da parte di tutti gli utenti.

Tuttavia, per poter pianificare il rinnovo annuale del contratto con il CBS e/o la sua modifica, è opportuno che ciascun servizio espliciti al Dipartimento dell'Integrazione e degli Standard l'utilizzo che intende fare nell'anno di questo software.

#### 2.4 L'effettuazione dell'indagine pilota

Una volta che si abbia a disposizione la ditta esterna che effettuerà la rilevazione, è possibile effettuare l'indagine pilota.

E' bene sottolineare che, secondo questa strategia, l'indagine pilota, oltre a prefiggersi gli obiettivi tipici di tutte le pilota, debba avere la peculiarità di verificare tutto l'aspetto procedurale di gestione dell'applicazione presso la società esterna. A tal fine, deve essere organizzata in modo che:

- possa eventualmente implicare un ristretto numero di giorni di rilevazione, ma utilizzi l'architettura informatica della ditta come sarà a regime a tutti i rilevatori con il massimo numero di postazioni CATI previste per l'indagine;
- oltre al controllo dei consueti indicatori di monitoraggio, debbano essere sottoposte a monitoraggio le performance della rete e di tutta l'architettura hw-sw della ditta; per questa attività, è necessario quindi coinvolgere professionalità specifiche disponibili in Istat;
- debba essere testato il buon funzionamento di tutto il processo di rilevazione, includendo gli aspetti inerenti l'invio giornaliero dei report, il sistema di schedulazione delle chiamate, ecc.;
- ne consegue che, durante l'indagine pilota, il monitoraggio giornaliero presso la società esterna debba essere effettuato non soltanto dagli esperti di indagine che supervisionano i rilevatori e verificano eventuali problematiche del questionario da un punto di vista contenutistico, ma anche dagli esperti Blaise, nonché dagli informatici esperti di connessioni e sistemisti Windows.

Al termine dell'indagine pilota, quindi, dovranno emergere una serie di spunti, quali:

- elementi per un ulteriore perfezionamento del questionario;
- argomenti da approfondire in un ulteriore *briefing* degli intervistatori;
- adattamenti del sistema di schedulazione delle chiamate;
- aggiustamenti dell'architettura informatica della società esterna.

#### 2.5 L'indagine a regime

Si ritiene infine opportuno riportare in una tabella le attività da espletarsi durante la rilevazione in funzione della titolarità della responsabilità di ciascuna di esse nell'ambito della strategia finora considerata, secondo un'organizzazione che appare ottimale. Molte di queste attività sono realizzabili anche a distanza, via rete, tuttavia alcune di esse richiedono un intervento in loco. Per esempio, l'installazione di tutto il software può essere effettuata a distanza, mentre il monitoraggio quotidiano sul suo corretto funzionamento, da intensificarsi durante i primi giorni di rilevazione, deve essere fatto presso la ditta. E' evidente quindi che è auspicabile che la società si trovi nella stessa provincia del servizio Istat responsabile dell'indagine; tuttavia, qualora non sia così, per alcune delle attività citate, la responsabilità potrebbe essere condivisa tra Istat e la ditta esterna, oppure potrebbe comunque restare a carico Istat, delegando personale degli uffici regionali, opportunamente formato.

Tabella.2.1 - attività da espletarsi durante la rilevazione e relativa responsabilità

Istat	Installa presso la ditta <ul style="list-style-type: none"> <li>• L'eseguibile del questionario elettronico</li> <li>• L'elenco dei nominativi da intervistare</li> <li>• Il sistema di schedulazione delle chiamate, già impostato con parametri predefiniti</li> <li>• Il software per la produzione dei report</li> <li>• Alcune <i>utility</i> di sistema per l'ottimizzazione della gestione dello spazio disco del server e per lo scarico dei dati.</li> </ul>
-------	---

	<ul style="list-style-type: none"> <li>• Il software per la produzione dei report</li> <li>• Alcune <i>utility</i> di sistema per l'ottimizzazione della gestione dello spazio disco del server e per lo scarico dei dati.</li> </ul>
	Effettua il monitoraggio in loco (soprattutto durante i primi giorni di rilevazione)
La società	Effettua giornalmente il <i>day-batch</i> Effettua le interviste Manda in esecuzione le <i>utility</i> per lo scarico dei dati e l'ottimizzazione dello spazio disco Manda in esecuzione i report Invia dati e report giornalmente ad Istat
Istat	Interviene in caso di malfunzionamenti Aggiorna, se necessario, il sistema di schedulazione delle chiamate Consulta gli indicatori di monitoraggio ed interviene con eventuali correttivi

### 3 L'indagine campionaria sulle nascite

Nel 2002 l'Istat svolge per la prima volta una indagine campionaria sulle nascite. Tale indagine si inserisce all'interno del nuovo Sistema Integrato di Osservazione sulle Nascite e i Parti, costituito da un insieme di rilevazioni, alcune delle quali già esistenti da numerosi anni, altre di recente istituzione. Il suo impianto si è reso necessario a seguito delle modificazioni normative in tema di dichiarazione di nascita introdotte dalla Legge 127/97, o Bassanini-bis, e del successivo Regolamento di attuazione (D.P.R. 20 Ottobre 1998 n.403). Tali variazioni normative hanno reso impossibile la prosecuzione della tradizionale rilevazione delle nascite di fonte Stato Civile, che l'Istat ha condotto dal 1926.

L'obiettivo concettuale complessivo dell'indagine consiste nel garantire un monitoraggio corrente dell'andamento della fecondità in Italia, possibilmente con un livello di attendibilità accettabile per le regioni, almeno per quanto riguarda le principali caratteristiche strutturali demografiche. Oltre a questi aspetti, l'indagine campionaria, essendo rivolta direttamente alle madri, offre l'opportunità di rilevare le caratteristiche di contesto della nascita, come la partecipazione della donna al mercato del lavoro, l'affidamento e la cura del bambino, la divisione dei ruoli nel lavoro domestico. Di grande interesse è inoltre il confronto di alcune caratteristiche socio-demografiche delle madri con quelle delle donne nel loro complesso desumibili da altre indagini dell'Istituto (Censimenti, Indagini Campionarie sulle Famiglie).

La popolazione oggetto di indagine è costituita dai nati vivi iscritti in Anagrafe riconosciuti almeno dalla madre, relativi - per la prima edizione dell'indagine - alla seconda metà del 2000 e alla prima del 2001. Le interviste vengono rivolte esclusivamente alla madre del bambino, che viene contattata ad una distanza media di un anno e mezzo dalla nascita.

Le motivazioni principali per cui si è deciso di effettuare un'indagine CATI sulle nascite sono essenzialmente due. Innanzi tutto, l'acquisizione telefonica dei dati permette un disegno di campionamento più efficiente perché privo di "stadi": si salta infatti lo stadio "Comune" e si estrae direttamente la madre attraverso la lista dei nominativi che costituiscono l'universo. In secondo luogo, con questa metodologia i dati vengono registrati direttamente, effettuando una serie di controlli e verifiche delle incompatibilità e permettendo di risolvere immediatamente alcune incoerenze insieme alla persona intervistata. Questo garantisce qualità delle informazioni e maggiore tempestività nella produzione dei risultati, che non dovrebbero allontanarsi troppo dal verificarsi della nascita (i tempi di avvio dell'indagine sono già posticipati a causa dei tempi di acquisizione della lista universo, che verrà illustrata nel par. 3.2).

Alcuni inconvenienti intervengono però sia a causa della tecnica utilizzata, sia per la particolarità della lista universo dell'indagine in questione. Come succede per ogni rilevazione telefonica, infatti, si restringe la popolazione oggetto di indagine alle sole unità per le quali viene reperito il numero telefonico e occorrono analisi ad hoc per ovviare alla distorsione campionaria che ne deriva. Inoltre, la predisposizione della lista universo dai dati di un'altra indagine Istat di recente istituzione, come si accennava, comporta tempi prolungati - per l'arrivo dei modelli, la registrazione e i controlli - e soprattutto provoca maggiori difficoltà di abbinamento dei nominativi ai numeri telefonici.

Per testare il questionario e l'intero percorso preliminare all'esecuzione delle interviste, è stato realizzato un pre-test nel laboratorio CASIC dell'Istat nel settembre 2001. Dopo l'effettuazione di una indagine pilota eseguita dalla stessa ditta vincitrice dell'appalto, l'indagine definitiva è stata avviata ad aprile 2002 e prevede tre fasi, ciascuna composta di un quadrimestre di nascite (luglio-ottobre 2000, novembre 2000-febbraio 2001, marzo-giugno 2001). Al momento è stata completata la prima fase di rilevazione e si conta di portare a termine le restanti entro la fine del 2002. I risultati discussi nel corso del presente lavoro si riferiranno quindi solo alle interviste del primo quadrimestre di nascite.

#### 3.1 L'acquisizione dei dati e i contenuti informativi

L'impianto organizzativo dell'indagine prevede un sistema di acquisizione articolato in due questionari: una versione

### 3.1 L'acquisizione dei dati e i contenuti informativi

L'impianto organizzativo dell'indagine prevede un sistema di acquisizione articolato in due questionari: una versione ridotta (*short form*) ed una estesa (*long form*). La prima viene somministrata a tutto il campione e consente di ottenere stime rappresentative a livello regionale per le principali determinanti delle nascite, mentre la seconda viene somministrata solo ad 1/3 del campione, garantendo la rappresentatività ad un livello territoriale più ampio per il primo ciclo di rilevazione e regionale in prospettiva quando potranno essere accorpati i dati di più edizioni di indagine.

Il questionario nella sua versione completa (*long form*) si compone di 4 sezioni:

1. 'Il contesto familiare e il parto', che raccoglie le principali notizie socio demografiche sui genitori, sul nato, e sugli altri componenti familiari, informazioni di carattere socio sanitario sul parto e sulla storia riproduttiva della donna;
2. 'Il lavoro della madre prima e dopo la nascita del figlio', in cui si acquisiscono informazioni specifiche sul lavoro della donna e, in modo particolare, si mira a cogliere eventuali variazioni intercorse nella vita professionale in seguito alla nascita del figlio.
3. 'La cura del bambino e la divisione del lavoro familiare', con domande sulla rete di organizzazione familiare per provvedere alle esigenze di accudimento del bambino, sui soggetti scelti per l'affidamento (nonni, parenti, baby sitter, asilo nido ed altri servizi all'infanzia, ecc.), sulla partecipazione del compagno/marito allo svolgimento dei lavori domestici e all'accudimento del/i bambino/i;
4. 'L'abitazione e il contesto socio-economico', sezione conclusiva che mira a raccogliere indicazioni sul contesto in cui la nascita è avvenuta attraverso variabili relative all'abitazione, ai beni posseduti e al reddito della famiglia.

La versione *short form* del questionario comprende invece tutta la prima sezione, con le notizie strutturali di base sui genitori, il nato, il parto, e l'aggiunta di alcune domande sulle principali caratteristiche dell'abitazione.

Mentre nella *short form* si somministra un numero di quesiti che varia da 29 a 79, nella *long form* possono essere formulate dalle 76 alle 212 domande, per lo più concentrate nella prima sezione (vedi prospetto che segue).

Tabella 3.1 Struttura dei questionari

Questionario	Numero minimo di quesiti	Numero massimo di quesiti
<i>Long Form</i>	76	212
di cui:		
Sezione 1. 'Il contesto familiare e il parto'	27	113
Sezione 2 'Il lavoro della madre prima e dopo la nascita del figlio'	34	62
Sezione 3 'La cura del bambino e la divisione del lavoro familiare'.	6	23
Sezione 4. 'L'abitazione e il contesto socio-economico'	9	14
<i>Short Form</i>	29	79

### 3.2 La lista universo e l'abbinamento dei numeri telefonici

La lista di campionamento è fornita dai dati di un'altra rilevazione Istat, quella sugli Iscritti in Anagrafe per Nascita (mod. Istat P.4). Istituita il 1/1/1999, questa rilevazione raccoglie a livello esaustivo schede individuali su ogni nato che viene iscritto in Anagrafe. Le informazioni raccolte riguardano i genitori (data di nascita, stato civile e cittadinanza), il nato (data e luogo di nascita, cittadinanza) e l'intestatario della scheda di famiglia (indirizzo completo del luogo di residenza). Tali informazioni permettono di reperire i numeri di telefono delle madri che saranno intervistate. L'operazione di abbinamento viene eseguita dalla Società che ha avuto in appalto la gestione delle telefonate, secondo i criteri forniti dall'Istat. Tale ricerca viene effettuata all'inizio di ciascuna fase di rilevazione sull'universo delle nascite, circa 180.000 nominativi ogni volta, da cui viene successivamente estratto il campione da intervistare.

Tabella 3.2. Percentuale di abbinamenti per regione

Regione	Abbinati		Non abbinati		Totale	
	V.A.	%	V.A.	%	V.A.	%
<i>Piemonte</i>	7,352	65.2	3,917	34.8	11,269	100.0
<i>Valle d'Aosta</i>	292	68.9	132	31.1	424	100.0
<i>Lombardia</i>	19,110	67.0	9,426	33.0	28,536	100.0
<i>Veneto</i>	10,492	70.5	4,396	29.5	14,888	100.0
<i>Friuli Venezia Giulia</i>	2,130	66.5	1,071	33.5	3,201	100.0
<i>Liguria</i>	2,341	61.5	1,468	38.5	3,809	100.0
<i>Emilia Romagna</i>	7,725	64.9	4,170	35.1	11,895	100.0
<i>Toscana</i>	6,470	65.6	3,205	31.4	9,675	100.0

<i>Friuli Venezia Giulia</i>	2,150	60.5	1,071	33.5	3,201	100.0
<i>Liguria</i>	2,341	61.5	1,468	38.5	3,809	100.0
<i>Emilia Romagna</i>	7,725	64.9	4,170	35.1	11,895	100.0
<i>Toscana</i>	6,470	65.6	3,395	34.4	9,865	100.0
<i>Umbria</i>	1,561	66.5	787	33.5	2,348	100.0
<i>Marche</i>	2,917	71.8	1,143	28.2	4,060	100.0
<i>Lazio</i>	9,606	59.2	6,627	40.8	16,233	100.0
<i>Abruzzo</i>	2,400	71.2	970	28.8	3,370	100.0
<i>Molise</i>	610	69.6	267	30.4	877	100.0
<i>Campania</i>	13,369	57.1	10,036	42.9	23,405	100.0
<i>Puglia</i>	8,595	56.7	6,556	43.3	15,151	100.0
<i>Basilicata</i>	1,334	67.6	639	32.4	1,973	100.0
<i>Calabria</i>	3,667	52.7	3,297	47.3	6,964	100.0
<i>Sicilia</i>	9,852	52.4	8,939	47.6	18,791	100.0
<i>Sardegna</i>	2,882	60.2	1,902	39.8	4,784	100.0
<i>Bolzano</i>	941	51.4	888	48.6	1,829	100.0
<i>Trento</i>	1,315	74.4	452	25.6	1,767	100.0
<b>Totale</b>	<b>114,961</b>	<b>62.0</b>	<b>70,478</b>	<b>38.0</b>	<b>185,439</b>	<b>100.0</b>

L'operazione di ricerca dei numeri telefonici per la prima fase ha ottenuto il 62% di abbinamenti, variando da un minimo di 51,4% per Bolzano e 52,7 per la Calabria ad un massimo di 74,4% per Trento e 71,8% per le Marche (tabella 3.2).

Durante le fasi di abbinamento dei numeri è stato necessario effettuare un monitoraggio dei nominativi abbinati e non abbinati, attraverso l'analisi della copertura degli strati (regioni e classi d'età) e della percentuale di non abbinati nelle varie province. Dopo una prima elaborazione da parte della ditta, sono stati quindi richiesti ulteriori abbinamenti da ricercarsi manualmente, necessari a garantire la copertura in tutti gli strati per la selezione del campione.

### 3.3 Il disegno campionario

Il disegno campionario prevede un campionamento ad uno stadio stratificato; i domini territoriali di riferimento delle stime sono di due tipi: l'intero territorio nazionale e le regioni geografiche. Sulla base delle informazioni contenute nella lista universo, è possibile stratificare le unità della popolazione per regione di residenza e classe di età della madre. La numerosità campionaria totale è costituita da 50.000 madri (circa il 10% di tutto l'universo) dei nati iscritti in anagrafe tra giugno 2000 e maggio 2001; solo una dimensione così elevata potrà garantire una buona rappresentatività delle stime dei nati per filiazione (legittimi e naturali) e fino al terzo ordine di nascita.

Per raggiungere il numero di interviste complete previste, è stato individuato un numero aggiuntivo di nominativi per sostituire quelli "caduti" per motivi vari (numero telefonico errato, rifiuto, interruzione definitiva, ecc.). A tale scopo è stato utilizzato il criterio del sovracampionamento; tuttavia, trattandosi della prima edizione dell'indagine, per la prima fase si è operato nel modo seguente.

All'inizio dell'indagine, sono stati forniti alla ditta i nominativi del campione base. Dopo i primi 20 giorni di rilevazione è stato analizzato il file con gli esiti dei contatti (definitivi e provvisori) ottenuti fino ad allora. Sono stati quindi calcolati i tassi di caduta per ogni strato e separatamente per *short* e *long*, ed estratti i nominativi da aggiungere al campione base con il metodo del sovracampionamento.

Nella seconda e nella terza fase saranno invece utilizzati i tassi di caduta riscontrati nella prima fase, per consegnare anticipatamente alla ditta i nominativi da intervistare ottenuti con il sovracampionamento.

## 4 Descrizione dell'applicazione sviluppata per l'Indagine Campionaria sulle Nascite

L'applicazione CATI sviluppata per l'Indagine Campionaria sulle Nascite comprende un insieme articolato di operazioni e passaggi. Fulcro dell'applicazione è il questionario elettronico descritto in dettaglio nei successivi paragrafi.

### 4.1 Il questionario elettronico

Per la realizzazione del questionario elettronico è stato utilizzato il software *Blaise* versione 4.53, costituito da un pacchetto per l'acquisizione controllata dei dati, sviluppato dall'Ufficio di Statistica Olandese, scritto in Pascal e già ampiamente diffuso presso altri stati membri della Comunità Europea, in quanto Eurostat stesso ne promuove l'utilizzo.

Il questionario si compone essenzialmente di due moduli:

- il modulo CATI per la gestione delle telefonate;

pacchetto per l'acquisizione controllata dei dati, sviluppato dall'Ufficio di Statistica Olandese, scritto in Pascal e già ampiamente diffuso presso altri stati membri della Comunità Europea, in quanto Eurostat stesso ne promuove l'utilizzo. Il questionario si compone essenzialmente di due moduli:

- il modulo CATI per la gestione delle telefonate;
- il modulo DATAMODEL per l'acquisizione dati, contenente:
  - ⇒ la definizione delle variabili;
  - ⇒ la struttura delle regole per la somministrazione dei quesiti;
  - ⇒ la definizione dei controlli di coerenza tra variabili;
  - ⇒ un sotto-modulo per la codifica assistita delle variabili testuali;
  - ⇒ le istruzioni di collegamento a file esterni.

In particolare nel questionario realizzato sono state definite:

- 677 variabili "definitive", ossia memorizzate sul database, cui vanno aggiunte tutte le variabili ausiliarie utilizzate per passaggi intermedi ma non necessarie per l'elaborazione finale dei dati;
- 195 regole di controllo di cui 101 per la gestione dei controlli *soft* e 94 per i controlli di tipo *hard*;
- un sotto-modulo per la codifica assistita della variabili Comune e Stato estero di nascita e di residenza;
- due programmi per calcolare automaticamente:
  - ⇒ l'ordine di nascita, variabile fondamentale per l'indagine;
  - ⇒ lo strato di appartenenza delle unità campionarie, per verificare eventuali migrazioni fra strati rispetto a quelli definiti a monte nel disegno campionario;
- le regole di confronto con i dati di fonte anagrafica.

Il calcolo dell'ordine di nascita e dello strato di appartenenza costituiscono elementi di fondamentale utilità, infatti:

- l'ordine di nascita viene calcolato da programma sulla base di dati oggettivi raccolti durante la telefonata; il risultato viene sottoposto a verifica chiedendone conferma all'intervistata. In questo modo si garantisce un'elevata precisione e quindi qualità del dato raccolto;
- lo strato di appartenenza viene calcolato da programma attraverso lo stesso algoritmo utilizzato per la definizione degli strati nella fase di campionamento, ma sulla base di dati acquisiti durante l'intervista e non derivanti da fonte anagrafica (utilizzati in fase di disegno campionario). In questo modo è possibile memorizzare lo strato aggiornato, confrontarlo con quello calcolato a priori e controllare *on-line* gli spostamenti fra strati delle unità campionarie.

Da quanto appena descritto, si deduce l'altra funzione contenuta nel questionario elettronico, ossia la possibilità di raccogliere le informazioni di carattere anagrafico (date di nascita, comune di residenza, nome del bambino, stato civile, ecc.) confrontarle con quelle contenute nel mod. Istat P4 e correggere queste ultime in caso di discrepanza. Questo aspetto, come meglio spiegato nel capitolo "Principali risultati" è di notevole importanza in quanto avrà un impatto non indifferente sui metodi di raccolta dati utilizzati nell'indagine amministrativa alla base dell'attuale rilevazione telefonica.

#### 4.1.1 Gli standard adottati per la gestione dello schermo

Nella definizione del questionario elettronico si è cercato di utilizzare standard di visualizzazione che facilitassero il più possibile il lavoro degli intervistatori, riducendone la fatica e aumentandone contemporaneamente la produttività.

Per la visualizzazione a video dei quesiti si è cercato, prima di tutto, di rispettare la struttura del questionario cartaceo: in questo modo è stato semplificato il compito del ricercatore responsabile dell'indagine nella fase di controllo e test del questionario elettronico.

Inoltre, grazie alle potenzialità offerte da *Blaise*, è stato possibile, nel questionario a video:

- visualizzare, su un'unica schermata, intere sequenze di domande su una stessa tematica (non limitandosi quindi ad una schermata per quesito);
- introdurre dei salti pagina al cambiare delle sezioni, ossia di argomento;
- utilizzare colori diversi secondo la funzione svolta dal testo:
  - ⇒ **nero** per le domande da leggere all'intervistata;
  - ⇒ **verde** per i titoli di sezione;
  - ⇒ **rosso** per tutte le istruzioni o i messaggi esplicativi per l'operatore da non leggere al rispondente.

Per la visualizzazione del questionario, *Blaise* realizza automaticamente la suddivisione in orizzontale dello schermo:

- nella parte inferiore è riportato l'elenco dei quesiti appartenenti ad una sezione: per la loro visualizzazione è stato scelto di utilizzare un identificativo sintetico "parlante" (ad esempio per la domanda "*Potrebbe dirmi quando è nata?*" è stato utilizzato l'identificativo sintetico "*data nascita madre*"), piuttosto che il numero di domanda, riportato sul questionario e visualizzato automaticamente da *Blaise*, poco esplicativo per l'operatore telefonico in caso di "ritorno" sulle domande in presenza di un errore;
- nella parte superiore dello schermo compare il testo per esteso della domanda selezionata (ossia su cui è posizionato il cursore nella parte inferiore del video); oltre al testo compaiono

- domande in presenza di un errore;
- nella parte superiore dello schermo compare il testo per esteso della domanda selezionata (ossia su cui è posizionato il cursore nella parte inferiore del video); oltre al testo compaiono ovviamente tutte le eventuali spiegazioni e istruzioni per l'operatore.

Di seguito sono riportati alcuni esempi di schermate del questionario elettronico esemplificative degli standard adottati appena descritti.

**Esempio di schermata Blaise e di colori: in verde il titolo della sezione, in nero i testi da leggere**

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

-NOTIZIE SUL NATO-  
Ora le farò alcune domande su CRISTIAN  
I14. Mi può gentilmente indicare se è:

1. figlio suo e del suo compagno/marito  
 3. figlio suo

Paternità  x1

Nascita  
Sesso   
Data Nasc Bambino   
Dove   
Comune  
Stato estero

Old 6/70 Modified Dirty Navigate Nati

**Esempio di schermata Blaise e di colori: in nero il testo da leggere, in rosso l'avvertenza per l'operatore di digitare il peso in grammi anche se l'intervistata lo esprime in chilogrammi.**

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

I22. Quanto pesava CRISTIAN alla nascita?  
(Specificare il peso in grammi GGGG)

Enter a numeric value between 300 and 7000

Gemellare  SemplicPrematuro   
Vivi maschi Sett. gravidanza  
Vivi femmine Mesi gravidanza  
Morti Presenza padre   
Luogo parto  x1 Anomalia nascita   
Specificare Tipo anomalia  
Tipo parto  x1 Allattato   
Peso bambino  Mesi

Old 7/70 Modified Dirty Navigate Nati

#### 4.1.2 Gli standard adottati per la gestione degli errori

Gli errori gestiti da *Blaise* (in relazione alla tecnica CATI) sono di due tipologie:

- errori di *range*, riguardanti il dominio delle variabili: il rilevatore è obbligato a rimuoverli per poter continuare l'intervista introducendo valori contenuti nell'intervallo definito;

Gli errori gestiti da *Blaise* (in relazione alla tecnica CATI) sono di due tipologie.

- errori di *range*, riguardanti il dominio delle variabili: il rilevatore è obbligato a rimuoverli per poter continuare l'intervista introducendo valori contenuti nell'intervallo definito;
- errori di coerenza, concernenti risposte incompatibili tra due o più domande o riguardanti risposte che fuoriescono da una "casistica tipo". In questo caso sono fornite al rilevatore due possibilità in funzione del tipo di incompatibilità definita in fase di progettazione (si veda par. 2.2.1.4):
  - se l'incompatibilità è di tipo *hard*, il rilevatore è obbligato a rimuoverla per continuare l'intervista;
  - se invece è di tipo *soft*, il rilevatore può correggere o confermare il dato inserito.

Stabilire se un errore è di tipo *hard* o *soft* spetta a chi definisce il questionario. Lo standard adottato in questo caso è stato quello di definire un errore *hard* o *soft* a seconda se contrasti o meno con regole estremamente significative per la qualità dei dati rilevati. In particolare, per questa indagine sono stati definiti:

- *hard* gli eventuali errori di compatibilità tra dati oggettivi, come tra date, tra numero di figli propri della donna e gravidanze dichiarate, tra la presenza/assenza in famiglia del coniuge e lo stato civile della donna;
- *soft* gli eventuali errori derivanti da risposte "particolari", ossia che fuoriescono da una "casistica tipo" e che quindi non è detto che siano necessariamente sbagliate.

Vediamo alcuni esempi.

Sono di tipo *hard* gli errori che contraddicono le seguenti regole:

- ⇒ "la madre non può essere nata prima del figlio";
- ⇒ "il numero di figli avuti da una donna non può superare il numero di gravidanze senza parti gemellari"
- ⇒ "lo stato civile non può essere quello di nubile in caso di coniuge convivente"

Rappresentano casi di errori *soft* i seguenti esempi:

- ⇒ per la variabile "numero di figli per donna", dopo aver definito un *range* di validità piuttosto ampio, è stato considerato errore di tipo *soft* una risposta maggiore o uguale a "6 figli", risposta non errata, ma particolare rispetto alla normalità dei casi che prevede una prole decisamente meno numerosa. Questo controllo *soft* è stato pensato per prevenire eventuali errori di digitazione.
- ⇒ Per la variabile "durata del periodo di aspettativa post-parto" è stato definito un errore *soft* per le durate inferiori ai tre mesi che rappresentano un periodo obbligatorio di assenza dal lavoro. Il controllo è *soft* perché pensato per correggere un eventuale errore di memoria dell'intervistata, in quanto questo tipo di risposte non costituisce di per sé un errore oggettivo.

Cosa succede in caso di inserimento di un dato errato?

*Blaise* apre una "finestra di dialogo" intitolata *Hard Error* o *Active Signal (soft error)*, contenente un messaggio corredato sia da un commento esplicativo dell'errore stesso che dall'elenco delle domande con i valori registrati che hanno generato l'errore: in questo modo il rilevatore è in grado di capire immediatamente il tipo di incongruenza registrata e posizionarsi sulle domande con risposta sbagliata per inserire quella corretta.

Gli standard adottati per la messaggistica di errore riguardano:

- il colore del testo – **rosso** se il messaggio è un avviso per l'operatore che non deve essere letto all'intervistata, **nero** in caso sia una risposta registrata coerentemente con quanto detto dall'intervistata, ma che contraddice informazioni che questa ha dato precedentemente e che quindi le deve essere letto per capire dove si sia verificata l'incongruenza;
- il modo di redigere il testo che deve contenere frasi "parlanti" necessarie non solo quando il messaggio debba essere letto al rispondente, ma anche nel caso in cui sia un avviso per il solo operatore che deve essere messo nelle condizioni di capire immediatamente l'anomalia registrata. Occorre, infatti, ricordare che l'operatore si trova al telefono con un altro individuo e quindi un messaggio di errore contenente frasi brevi e poco esplicative allungherebbe il tempo necessario all'intervistatore per correggere l'errore creando dei lunghi momenti di pausa che potrebbero infastidire il rispondente che non può vedere cosa stia succedendo dall'altro capo del telefono;
- l'attribuzione di un identificativo numerico a ciascun messaggio di errore al fine di facilitare la manutenzione del *software*.

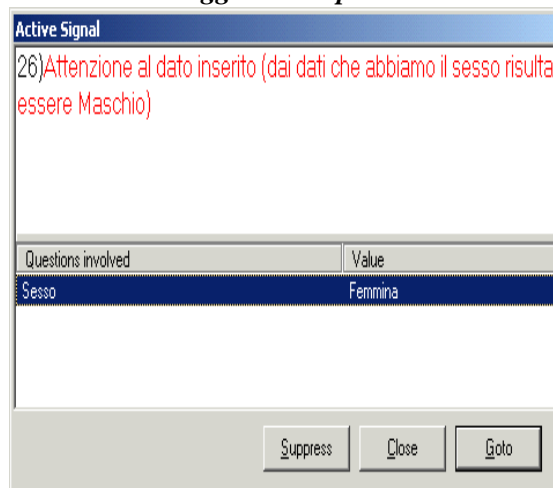
A questi standard si collega quello descritto precedentemente per gli identificativi delle variabili



manutenzione del *software*.

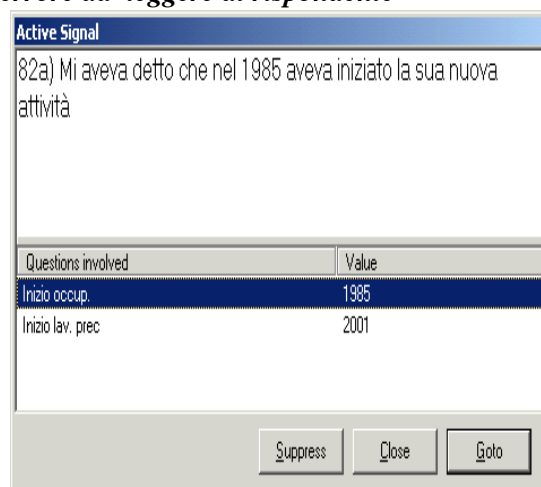
A questi standard si collega quello descritto precedentemente per gli identificativi delle variabili riportati nella parte inferiore dello schermo: l'identificativo che esplicita in modo sintetico, ma chiaro, il contenuto della variabile è di grande utilità per l'operatore che riesce ad individuare immediatamente la domanda che ha causato l'incongruenza tra quelle elencate nella "finestra di dialogo" del messaggio di errore.

#### ***Esempio di messaggio di errore da non leggere al rispondente***



#### ***Domanda coinvolta***

#### ***Esempio di messaggio di errore da leggere al rispondente***



#### ***4.1.3 Gli standard adottati per la gestione delle non risposte parziali***

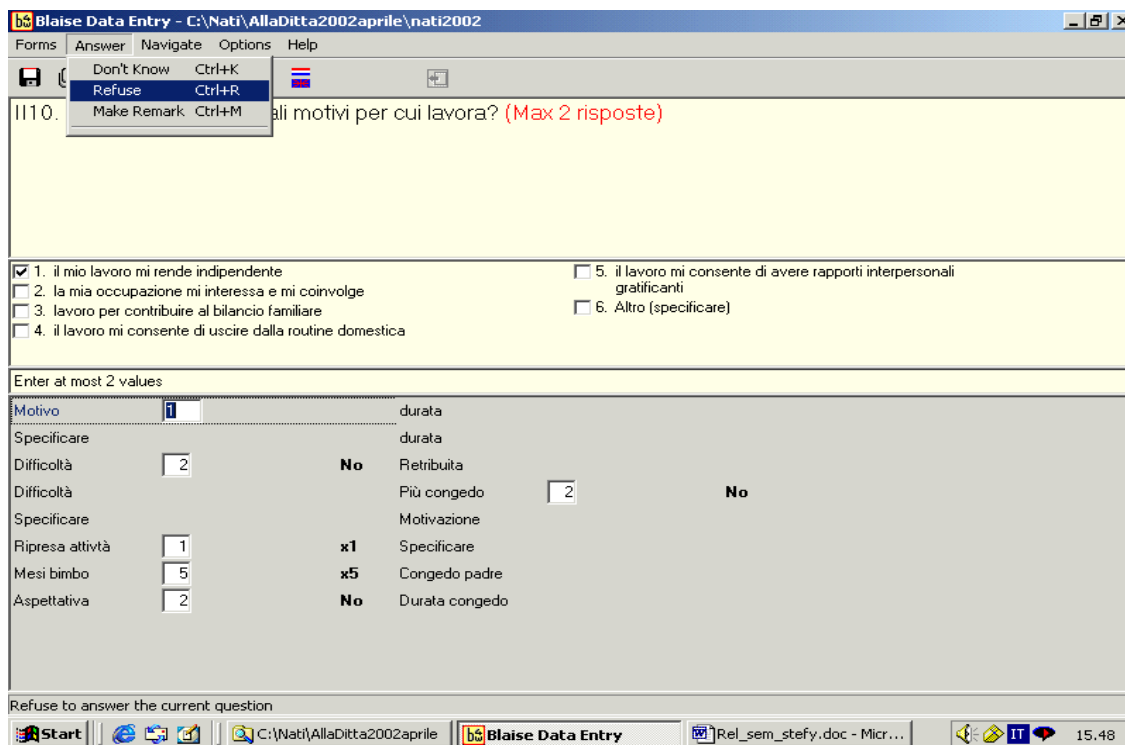
Particolarmente efficace per ottenere una buona qualità dei dati è la gestione offerta da *Blaise* delle modalità di non risposta parziale, ossia il "Rifiuto" e il "Non so/Non ricordo" (cfr. par. 2.2.1.4). E' noto come la presenza esplicita di queste modalità fra le alternative di risposta, possa indurre il rilevatore ad uno scarso impegno nell'ottenere le informazioni richieste. D'altra parte, l'assenza delle due voci impedirebbe il proseguimento dell'intervista qualora il

Particolarmente efficace per ottenere una buona qualità dei dati è la gestione offerta da *Blaise* delle modalità di non risposta parziale, ossia il “Rifiuto” e il “Non so/Non ricordo” (cfr. par. 2.2.1.4). E’ noto come la presenza esplicita di queste modalità fra le alternative di risposta, possa indurre il rilevatore ad uno scarso impegno nell’ottenere le informazioni richieste. D’altra parte, l’assenza delle due voci impedirebbe il proseguimento dell’intervista qualora il rispondente non volesse o non potesse rispondere oppure indurrebbe il rilevatore a dare una risposta fittizia falsando i dati. Grazie a quanto offerto da *Blaise* è stato possibile utilizzare due standard diversi per la gestione delle mancate risposte parziali a seconda se riferiti ai quesiti cruciali per l’indagine o alle altre domande meno determinanti:

- per i primi è stata disattivata la possibilità di inserire il “Rifiuto” e il “Non so/Non ricordo”: la mancata risposta ai quesiti cruciali rappresenta, infatti, l’interruzione definitiva dell’intervista;
- per tutte le altre domande, pur non esplicitando i due *item* fra le alternative di risposta, si è lasciata al rilevatore la possibilità di introdurli premendo una combinazione di tasti funzione (CTRL + M per il “Rifiuto” e CTRL + K per il “Non so/Non ricordo”) o “cliccando” con il mouse sulla voce *Answer* della menù posto sulla barra *Blaise* dei comandi: è evidente come questo espediente limiti il ricorso a queste due modalità di risposta in quanto l’operatore non le vede a video e quindi non è indotto a sceglierle e per inserirle deve abbandonare la “tastierina numerica”, utilizzata normalmente per inserire i dati, e utilizzare entrambe le mani per la loro digitazione (gestione che si complica quando le interviste si svolgono attraverso il normale apparecchio telefonico – ormai sempre più raro - e non tramite le cuffie con auricolare).

Queste due strategie hanno contribuito a limitare le mancate risposte parziali e le interruzioni definitive come meglio spiegato nel successivo capitolo “Principali risultati”.

#### **Esempio della modalità di inserimento delle mancate risposte parziali**



#### **4.1.4 La personalizzazione dei testi dei quesiti e dei messaggi d’errore**

Un altro standard utilizzato nella realizzazione del questionario elettronico riguarda la parametrizzazione dei testi introduttivi delle domande, delle domande stesse e dei messaggi di errore effettuata utilizzando risposte date in corso di intervista a domande precedenti oppure informazioni anagrafiche della famiglia memorizzate nella base dati dei nominativi campione (informazioni del mod. Istat P4).

In questo modo è stato possibile sia personalizzare l’intervista rendendola più scorrevole e più piacevole per il rispondente – ad esempio, tutti i quesiti riguardanti il bambino sono stati personalizzati citandone il nome - che semplificare il lavoro dell’intervistatore offrendogli la possibilità di leggere un messaggio di errore creato *ad-hoc* per l’incongruenza incontrata.

Vediamo in dettaglio alcuni esempi di parametrizzazione effettuati utilizzando le variabili del mod. P4 o le risposte a domande precedenti.

#### **1) Esempio di parametrizzazione di un testo introduttivo:**

*Il testo cambia a seconda che il “Numero di figli conviventi, oltre al bambino campione” - domanda che precede quella esemplificata - sia uguale o maggiore di uno.*

#### **1a) Parametrizzazione in caso di un solo figlio convivente:**

1a) Parametrizzazione in caso di un solo figlio convivente:

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

-I FIGLI CONVIVENTI-

Parliamo ora dell'altro figlio che vive con lei  
I58. Mi può gentilmente indicare se è:

1. figlio suo e del suo compagno/marito  
 2. figlio precedente di suo compagno/marito  
 3. figlio suo  
 4. figlio adottivo

Paternità  x2

Sesso  **Maschio**

Data nascita figlio cc

Dove

Comune

Stato estero

reddito

Old 13/70 Modified Dirty Navigate Nati

1b) Parametrizzazione in caso di più di un figlio convivente:

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

-I FIGLI CONVIVENTI-

Parliamo ora degli altri figli che vivono con lei, iniziando dal più piccolo  
I58. Mi può gentilmente indicare se è:

1. figlio suo e del suo compagno/marito  
 2. figlio precedente di suo compagno/marito  
 3. figlio suo  
 4. figlio adottivo

Paternità  x1

Sesso  **Maschio**

Data nascita figlio cc

Dove

Comune

Stato estero

reddito

Old 13/70 Modified Dirty Navigate Nati

2) Esempio di parametrizzazione doppia di un testo di una domanda

Il testo della domanda contiene due parametri: " il nome del bambino", il cui valore è acquisito dal mod. P4 e "a chi viene affidato il bambino" il cui valore si desume dalla domanda precedente III3. Nei due esempi di intervista riportati vediamo che:

- nel primo caso per il figlio di nome " Cristian" la mamma ha dichiarato di affidarlo ai " nonni"
- nel secondo caso il bambino di nome " Michele" viene affidato alla "baby sitter"

Esempio 2a

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

III4. Per quali motivi, ha scelto di affidare CRISTIAN a i nonni ? (DIGITARE 01 02 ECC.. COME MODALITA' DI RISPOSTA. Se l'intervistata esita leggere le alternative - max 3 risposte)

1. Per approccio educativo  11. Altro? (specificare)

2. per far socializzare il bambino

3. per la convenienza economica

4. per la fiducia

1114. Per quali motivi, ha scelto di affidare CRISTIAN a un/a baby sitter ? (DIGITARE 01 02 ECC.. COME MODALITA' DI RISPOSTA. Se l'intervistata esita leggere le alternative - max 3 risposte)

1. Per approccio educativo  11. Altro? (specificare)

2. per far socializzare il bambino

3. per la convenienza economica

4. per la fiducia

5. per la flessibilità degli orari

6. per comodità

7. per non esporre il bambino al rischio di frequenti malattie

8. per igiene

9. per la qualità delle cure date al bambino

10. non avevo altre alternative

Enter at most 3 values

Custodia  x3 Ore giorno

Specificare Ore settimana

Custodia altri

Frequenza Specificare

Custodia Asilo nido

Specificare Motivo

Ore giorno Specificare

Ore settimana

Old 62/70 Modified Dirty Navigate Nati

### Esempio 2b

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

1114. Per quali motivi, ha scelto di affidare MICHELE a una baby sitter ? (DIGITARE 01 02 ECC.. COME MODALITA' DI RISPOSTA. Se l'intervistata esita leggere le alternative - max 3 risposte)

1. Per approccio educativo  11. Altro? (specificare)

2. per far socializzare il bambino

3. per la convenienza economica

4. per la fiducia

5. per la flessibilità degli orari

6. per comodità

7. per non esporre il bambino al rischio di frequenti malattie

8. per igiene

9. per la qualità delle cure date al bambino

10. non avevo altre alternative

Enter at most 3 values

Custodia  x8 Ore giorno

Specificare Ore settimana

Custodia altri

Frequenza Specificare

Custodia Asilo nido

Specificare Motivo

Ore giorno Specificare

Ore settimana

Old 62/70 Modified Dirty Navigate Nati

### 3) Esempio di parametrizzazione di un messaggio di errore.

La data di nascita del bambino dichiarata dalla madre viene confrontata con quella memorizzata nella base dati (mod. Istat P4)

Active Signal

27) Può ripetere, dai dati che abbiamo la data di nascita risulta essere il 15/09/2000

Questions involved	Value
Data Nasc Bambino	31/05/2000

Suppress Close Goto

#### 4.1.5 La scelta della lingua con cui condurre l'intervista e le annotazioni in corso di intervista

Per poter condurre le interviste anche nell'area di Bolzano è stato previsto il questionario in lingua tedesca. La traduzione ha riguardato non solo i testi delle domande, ma anche dei messaggi di errore da leggere all'intervistata.

Per cambiare la lingua, in questo caso da quella italiana di *default*, a quella tedesca, *Blaise* permette di introdurre un'icona sulla barra del menù: in questo modo con un semplice "click" del mouse si può passare immediatamente da una lingua all'altra, non solo prima di iniziare l'intervista, ma anche durante il suo svolgimento. Questo è stato particolarmente utile per quei rispondenti di Bolzano che dopo aver iniziato a rispondere alle domande in italiano mostravano l'esigenza di

di introdurre un'icona sulla barra dei menu: in questo modo con un semplice click del mouse si può passare immediatamente da una lingua all'altra, non solo prima di iniziare l'intervista, ma anche durante il suo svolgimento. Questo è stato particolarmente utile per quei rispondenti di Bolzano che dopo aver iniziato a rispondere alle domande in italiano mostravano l'esigenza di continuare l'intervista nella lingua a loro più familiare.

Un altro standard adottato durante lo svolgimento dell'intervista è stato quello relativo all'utilizzo dei campi note. Questi campi associati ad ogni variabile consentono all'operatore di introdurre annotazioni, commenti o suggerimenti sia propri sia dell'intervistato che non possono essere raccolti dal questionario finalizzato alla rilevazione delle variabili oggetto di studio. Come vedremo meglio nella descrizione della fase di pre-test, i campi note sono risultati particolarmente utili per modificare i testi o la struttura delle domande in quanto poco chiare o formulate in modo da appesantire troppo la conversazione telefonica. Seguono due esempi di schermate relativi alla scelta della lingua e all'introduzione del campo note.

### ***Esempio di scelta della lingua e inserimento dei campi note***

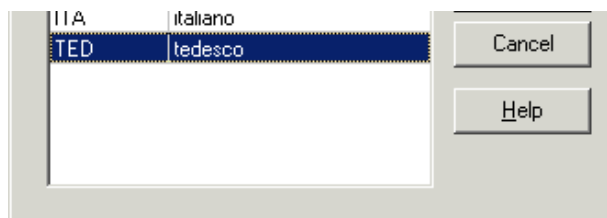
The screenshot shows the 'Blaise Data Entry' application window. The title bar reads 'Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002'. The menu bar includes 'Forms', 'Answer', 'Navigate', 'Options', and 'Help'. The main area displays a questionnaire in German: '-Die Pflege des Kindes- III1. Wer kummert sich am meisten um das Kind wenn Sie arbeiten? (Lasciare la risposta aperta e codificare)'. Below the question is a list of 12 radio button options, with option 8 ('Eine baby-sitterin') selected. At the bottom, there is a data entry form with fields for 'Betreuung' (x8), 'Stunde pro tag', 'Stunde pro woche', 'Grund', 'Kindergarten', and 'Stunde pro woche'.

### ***Finestra per l'inserimento dei campi note***

The screenshot shows a 'Remark' window with a large empty text area for entering notes. At the bottom, there are three buttons: 'Save', 'Cancel', and 'Help'.

The screenshot shows the 'Form Languages' dialog box. It contains a table with two columns: 'Name' and 'Description'. The table has two rows: 'ITA' with 'italiano' and 'TED' with 'tedesco'. The 'TED' row is selected. To the right of the table are three buttons: 'OK', 'Cancel', and 'Help'.

Name	Description
ITA	italiano
TED	tedesco



#### 4.1.6 La codifica assistita delle variabili testuali

Data la tecnica di rilevazione, il questionario adottato per l'indagine è totalmente strutturato. Vi sono comunque dei quesiti a testo libero che riguardano:

- la cittadinanza, il luogo di nascita e di residenza (comune italiano o Stato estero) del padre, della madre e del bambino "campione";
- la descrizione della professione svolta dall'intervistata.

Gli standard adottati per la codifica di queste variabili sono stati quelli di:

- utilizzare il modulo *Blaise* di codifica assistita per il primo gruppo di variabili;
- acquisire i testi per la variabile professione e codificare in *batch*, a fine rilevazione, con il software ACTR v3.

L'adozione dei due sistemi di codifica è dipesa sostanzialmente dalla diversa struttura della classificazione delle variabili: estremamente semplice per il comune, lo Stato estero e la cittadinanza, ossia non gerarchica ad un solo livello, complicata, ossia gerarchica e a più livelli, per la professione, per la quale si temeva un eccessivo allungamento dei tempi di intervista a seguito dell'uso della codifica *on-line*. Il modulo di codifica assistita di *Blaise* consente di realizzare il *matching* tra la risposta testuale e il suo *database* di riferimento a seguito di un'analisi dei bigrammi e trigrammi contenuti nei testi. Ciò facilita l'attribuzione del codice anche in caso di non perfetta coincidenza tra il modo di esprimersi del rispondente e i testi del *database*. Inoltre, il modulo di codifica assistita di *Blaise* permette di codificare una risposta a testo libero in tre modi:

- digitando anche soltanto le prime lettere del testo;
- digitando il testo da codificare dopo aver "navigato" nella classificazione ad albero ed aver scelto il "ramo" più basso possibile della classificazione dove trovare un codice corrispondente al testo digitato;
- navigando ad albero fino a raggiungere il codice corrispondente al testo da codificare.

Con il primo modo di operare, adatto per classificazioni semplici, l'operatore digita il testo da codificare, il software si collega alla tabella esterna dove sono memorizzate, in questo caso, le dizioni dei comuni e degli Stati con i rispettivi codici, propone il testo o i testi corrispondenti a quello digitato e riporta sul campo della variabile di interesse il codice selezionato dall'operatore. Si tratta quindi di un sistema di codifica molto veloce che offre un'elevata qualità dei risultati essendo il margine di errore di codifica molto basso; per i comuni, ad esempio, l'errore è limitabile ai casi di omonimia di comuni situati in province diverse, anche se la presenza della provincia, accanto alla dizione del comune, limita moltissimo la possibilità di errore.

Confrontando infatti la distribuzione del comune di residenza della madre di fonte P4 con quella dei comuni rilevati e codificati durante l'intervista non si osserva alcun rilevante cambiamento: le distribuzioni dei valori assoluti si discostano di pochissime unità per i comuni più grandi, indicando possibili cambi di residenza dal momento dell'iscrizione in anagrafe a quello della rilevazione telefonica.

Per questa indagine, le tabelle alla base della codifica sono state costruite partendo dai dizionari utilizzati per la codifica in *batch* con ACTR delle medesime variabili rilevate nel "Censimento della Popolazione 2001". Senza entrare eccessivamente nel dettaglio, le tabelle contengono, per i comuni, non solo le 8101 denominazioni ufficiali, ma anche i nomi di comuni soppressi, di vecchie denominazioni ufficiali, i modi di chiamare un comune rilevati precedentemente in altre indagini e i nomi delle principali località. Per lo Stato estero, oltre ai nomi in italiano dei singoli Stati, le tabelle utilizzate sono state arricchite con le dizioni nelle principali lingue straniere, insieme alle abbreviazioni o sigle degli Stati esteri nonché con i nomi delle principali città straniere.

Il secondo e il terzo metodo per codificare *on-line* una variabile si adattano a classificazioni con

delle principali città straniere.

Il secondo e il terzo metodo per codificare *on-line* una variabile si adattano a classificazioni con strutture più complesse. Richiedono ovviamente una buona conoscenza della classificazione da parte dell'operatore che deve sapere quale ramo della classificazione scegliere prima di digitare il testo. Per questa prima fase di indagine, nonostante le intervistatrici fossero state formate sulla classificazione della professione, si è preferito non adottare questa tecnica e codificare successivamente in *batch* tutti i testi acquisiti utilizzando il software ACTR.

Questa scelta, fatta a priori e come già detto legata ai tempi di intervista, sarà confermata nelle prossime fasi di rilevazione in quanto la percentuale di testi relativi alla variabile professione codificati univocamente in *batch*, senza alcun l'intervento da parte del codificatore manuale, si è attestata ad un buon livello vale a dire al 60%-61%. Tale percentuale è destinata sicuramente ad aumentare, sia perché verranno impartite alle operatrici istruzioni più dettagliate che consentiranno di eliminare errori ricorrenti nel modo di acquisire la professione, sia perché sarà possibile effettuare un *training* del dizionario sulla base della codifica e quindi di aggiornarlo rispetto alle dizioni incontrate, come sempre avviene in un processo di codifica assistita.

### Esempio di codifica assistita della variabile Comune di nascita

The screenshot shows the 'Blaise Data Entry' software window. The title bar reads 'Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002'. The menu bar includes 'Forms', 'Answer', 'Navigate', 'Options', and 'Help'. The main area contains a question: 'I5.1. In quale comune è nata?'. Below the question is a text entry field with the instruction 'Enter a text of at most 30 characters'. The form is divided into several sections with labels and input fields:

- Data nascita madre:** A dropdown menu with a question mark icon.
- Nata Italia:** A dropdown menu with '1' selected and a 'Si' label.
- Comune:** A dropdown menu.
- Stato estero:** A dropdown menu.
- Vive Italia:** A dropdown menu with a question mark icon.
- Comune:** A dropdown menu.
- Stato estero:** A dropdown menu.
- Italiana:** A dropdown menu with a question mark icon.
- Cittadinanza:** A dropdown menu.
- Anni in Italia:** A dropdown menu.
- Periodo soggiorno:** A dropdown menu.
- Durata soggiorno:** A dropdown menu.
- Stato civile:** A dropdown menu with '2' selected and an 'x2' label.
- Data matrimonio:** A dropdown menu.
- Nascita:** A dropdown menu.
- Stato Civile prec.:** A dropdown menu.

At the bottom of the window, there is a status bar with the following information: 'Old', '4/70', 'Modified', 'Dirty', 'Insert', 'Nati'.

*Si apre la tabella dei comuni (digitando il testo Roma per esteso verrebbe visualizzato un solo record), l'operatore sceglie con un invio il comune di interesse e il software restituisce il codice provincia + comune.*

The screenshot shows the 'Database Browser' window with the title 'c:\tabelle\Comuni - Database Browser'. It displays a table with three columns: 'Descr', 'Cod', and 'provincia'. The table contains the following data:

Descr	Cod	provincia
ROMA	058091	ROMA
ROMAGNANO	065110	SALERNO
ROMAGNANO	022205	TRENTO
ROMAGNANO	003130	NOVARA
ROMAGNANO AL MONTE	065110	SALERNO
ROMAGNANO SESIA	003130	NOVARA
ROMAGNESE	018128	PAVIA
ROMALLO	022154	TRENTO
ROMANA	090061	SASSARI
ROMANENGO	019086	CREMONA
ROMANO	024086	VICENZA
ROMANO	016183	BERGAMO
ROMANO	001223	TORINO
ROMANO CANAVESE	001223	TORINO
ROMANO D'EZZELINO	024086	VICENZA
ROMANO D'ORCHIBIA	016183	BERGAMO

Below the table, there is a search field with the text 'rom' entered. To the right of the search field, there are radio buttons for 'Key type' with 'AlphaDesc' selected and 'ItrigrDesc' unselected. At the bottom right, there are 'Select' and 'Cancel' buttons. The status bar at the bottom left shows the number '13332'.

13332

Blaise Data Entry - C:\Nati\AllaDitta2002aprile\nati2002

Forms Answer Navigate Options Help

I5.1. In quale comune è nata?

Enter a text of at most 30 characters

Data nascita madre	<input type="text" value="?"/>	Cittadinanza	
Nata Italia	<input type="text" value="1"/> <b>Si</b>	Anni in Italia	
Comune	<input type="text" value="058091"/>	Periodo soggiorno	
Stato estero		Durata soggiorno	
Vive Italia	<input type="text" value="?"/>	Stato civile	<input type="text" value="2"/> <b>x2</b>
Comune		Data matrimonio	
Stato estero		Nascita	
Italiana	<input type="text" value="?"/>	Stato Civile prec.	

Old 4/70 Modified Dirty Insert Nati

## 4.2 Le fasi dell'applicazione CATI

Per la realizzazione dell'intera indagine CATI è stato necessario sviluppare un insieme articolato di operazioni da compiersi sia prima che dopo l'acquisizione dati mediante il questionario elettronico. I passi dell'applicazione sono schematizzabili nel seguente diagramma che ne riporta la sequenza operativa.

### 4.2.1 Controllo e caricamento dei nominativi campione

Questa fase consente di mettere i nominativi da intervistare in linea con il questionario elettronico. E' preceduta da una fase di controllo finalizzata a individuare i problemi che renderebbero un nominativo non intervistabile a priori. In particolare la fase di controllo verifica:

- la presenza e la correttezza formale dei campi contenenti le informazioni fondamentali per la definizione a priori e a posteriori degli strati campionari;



È preceduta da una fase di controllo finalizzata a individuare i problemi che renderebbero un nominativo non intervistabile a priori. In particolare la fase di controllo verifica:

- la presenza e la correttezza formale dei campi contenenti le informazioni fondamentali per la definizione a priori e a posteriori degli strati campionari;
- l'esistenza e l'univocità del numero telefonico.

Il numero di telefono costituisce la “chiave primaria” per la schedulazione delle chiamate: è quindi di fondamentale importanza che sia formalmente corretto e univoco. Il controllo accerta che il recapito telefonico, abbinato al nominativo durante la fase di ricerca numeri effettuata dalla ditta esterna, non sia un potenziale numero errato, verificando che il prefisso non sia più corto di due *digit* e il numero più corto di quattro. A monte di questo controllo è stato fornito alla ditta l'elenco dei prefissi o distretti telefonici esistenti per poter separare il prefisso dal resto del numero (suddivisione non più presente sugli archivi di Telecom Italia utilizzati per la fase di abbinamento). I numeri che non superano la fase di controllo vengono restituiti alla ditta per un'ulteriore ricerca.

Al termine della fase di controllo, i nominativi abbinati a numeri di telefono formalmente corretti vengono messi in linea con il questionario elettronico. A questo scopo è stato scritto un programma in *Manipula*, uno dei software contenuti nel pacchetto *Blaise* che, durante la fase di caricamento, effettua un controllo sul campo definito come chiave primaria (il telefono in questo caso) scartando automaticamente tutti i record con chiave primaria doppia. Anche in questa fase i nominativi scartati sono stati inviati alla ditta esterna e sottoposti ad una successiva fase di abbinamento.

Le fasi di controllo e di caricamento sono state ripetute più volte al fine di ottenere numeri di telefono univoci e formalmente corretti per tutti i nominativi di indagine. Occorre precisare che il controllo appena descritto è stato ovviamente effettuato sull'intero universo degli abbinati al fine di avere un “archivio” corretto da cui estrarre i nominativi campione da mettere in linea con il questionario elettronico per la realizzazione delle interviste. Inoltre, in questa prima fase di indagine il caricamento è stato effettuato due volte: la prima volta sono state caricate le unità campionarie “base”, la seconda volta le “riserve” che sono state “accodate” ai nominativi precedenti, senza alterare la struttura dei dati raccolti. Questo espediente ha consentito di superare il problema del non poter determinare a priori un tasso di sovra campionamento non essendo noto il tasso di caduta dell'indagine.

Inoltre è necessario fare un breve cenno alla fase di abbinamento dei numeri telefonici con l'archivio anagrafico effettuata a monte della procedura di controllo e caricamento dei nominativi ed eseguita per questa indagine dalla ditta esterna. Al fine di incrementare la percentuale di abbinati l'Istat ha implementato, analizzando un piccolo gruppo di nominativi, una procedura di standardizzazione degli indirizzi sfruttando il software di codifica automatica ACTR v3. La normalizzazione utilizzata consente di rendere omogenei tutti i toponimi degli indirizzi, ossia permette di riportare ad un'unica grafia tutti i modi di scrivere o abbreviare le parole via, viale, strada, circonvallazione, ecc. presenti nel file anagrafico e quindi di evitare il mancato abbinamento a seguito di indirizzi non conformi a quanto riportato nell'archivio telefonico. L'esito della standardizzazione è stato un incremento del 5% della percentuale di numeri abbinati. Nelle fasi successive di indagine si procederà ad applicare la standardizzazione lavorando sul totale dell'archivio anagrafico al fine di tener conto di un maggior numero di modi in cui i toponimi possono essere scritti e di capire se sia possibile normalizzare anche parole o abbreviazioni dell'indirizzo stesso (ad esempio i numeri, le date, i mesi, ecc). Se questo passaggio garantirà un consistente incremento di abbinati allora la fase di standardizzazione entrerà in modo strutturale e definitivo nell'intero processo di indagine del quale costituirà la fase iniziale.

#### 4.2.2 La schedulazione delle chiamate

Importantissima per un'indagine CATI è la gestione dei parametri che regolano la schedulazione delle chiamate: le scelte effettuate dal responsabile dell'indagine in accordo con quello della tecnica di rilevazione influiscono infatti sulla produttività giornaliera degli intervistatori e quindi sui tempi necessari per il completamento dell'indagine.

Conseguentemente è importante capire quali siano i valori “giusti” da attribuire ai parametri, sapendo che questi possono variare in funzione di diverse variabili come la lunghezza dell'intervista, il numero di rilevatori, le fasce orarie di chiamata, il periodo dell'anno in cui si effettuano le interviste e di tutti quegli eventi (festività, “ponti” vacanzieri) che incidono sulla presenza dei rispondenti presso le proprie abitazioni.

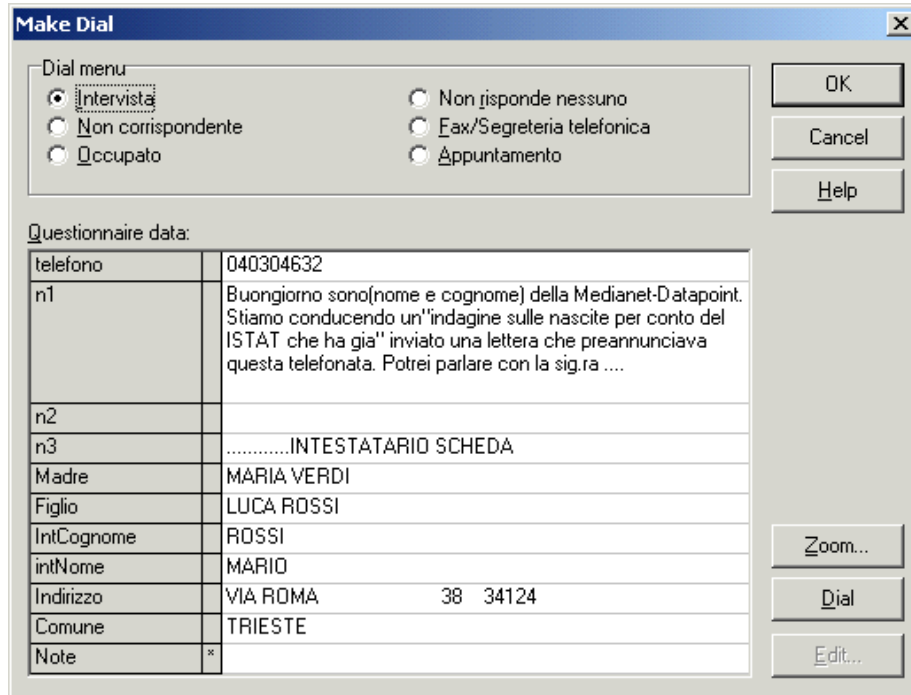
Il modulo *Blaise* “*CATI specification*” consente di introdurre i valori di tutti i parametri influenti sulla schedulazione delle chiamate: nonostante il numero dei parametri da considerare sia elevato

creando le interviste e di tutti quegli eventi (festività, ponti, vacanze) che incidono sulla presenza dei rispondenti presso le proprie abitazioni.

Il modulo *Blaise* “*CATI specification*” consente di introdurre i valori di tutti i parametri influenti sulla schedulazione delle chiamate: nonostante il numero dei parametri da considerare sia elevato come elevato è il grado di interazione fra di essi, il modulo *Blaise*, pensato per l’utente finale, ne permette una gestione abbastanza agile e semplice. Più ardua è ovviamente la fase a monte dove viene stabilito il set di valori; in questa prima fase di rilevazione si è rivelata a tal fine molto utile l’indagine pilota, che ha consentito di modificare i valori dei parametri che erano stati stabiliti a priori sulla base di ipotesi che, in alcuni casi, una volta effettuate le interviste, si sono verificate non adeguate.

Vediamo in dettaglio il tipo di parametri gestibili da *Blaise* e i valori ad essi assegnati, descrivendo prima di tutto come vengono attribuiti gli esiti di chiamata ai singoli nominativi. Prima di entrare nel “vivo” dell’intervista appare la schermata di inizio contatto con i nominativi campione, esemplificata di seguito.

### Schermata di inizio contatto



Nel riquadro in alto sono riportati i possibili esiti della contatto telefonico tra i quali l’operatore sceglie quello corrispondente all’esito del numero digitato. Si è scelto di mettere vicini gli esiti definitivi, quali “Intervista” e “Numero non corrispondente”, e quelli non definitivi, quali “Occupato”, “Non risponde nessuno (libero)”, “Fax/segreteria telefonica” e “Appuntamento”. Tra gli esiti definitivi non compare il “Rifiuto a collaborare”: è stato deciso infatti di poter introdurre questo esito solo dopo aver posto la domanda sulla “Ricezione della lettera” al fine di raccogliere questa importante informazione anche sulle mancate risposte totali e costruire indicatori importanti per l’indagine.

Nel riquadro in basso è riportato il testo introduttivo, da leggere una volta in contatto con il rispondente, insieme ad altre informazioni sulla famiglia, di fonte mod. Istat P4, che potevano essere di ausilio all’operatore telefonico per personalizzare l’intervista; queste sono: il nome e cognome della madre, ossia della persona da intervistare; il nome e cognome del bambino “campione”; il nome e cognome dell’intestatario della scheda familiare; l’indirizzo della residenza.

A seconda degli esiti di contatto scelti, il software si comporta in modo diverso:

- scegliendo l’esito intervista, entra nel questionario per l’inizio delle domande;
- scegliendo uno degli esiti “Occupato”, “Non risponde nessuno” o “Fax/segreteria telefonica” il software propone un nuovo nominativo;
- scegliendo “Numero non corrispondente” si apre una finestra di dialogo dove si chiede all’operatore di specificare il motivo di questo esito, vale a dire se il numero digitato è effettivamente errato o se è un numero inesistente oppure se il numero è corretto per quell’indirizzo, ma la famiglia cercata si è trasferita;
- scegliendo “Appuntamento” si apre un’altra finestra di dialogo dove è possibile inserire tutti i dati dell’appuntamento (data, periodo, ora).

Le due finestre appena descritte sono rappresentate di seguito.

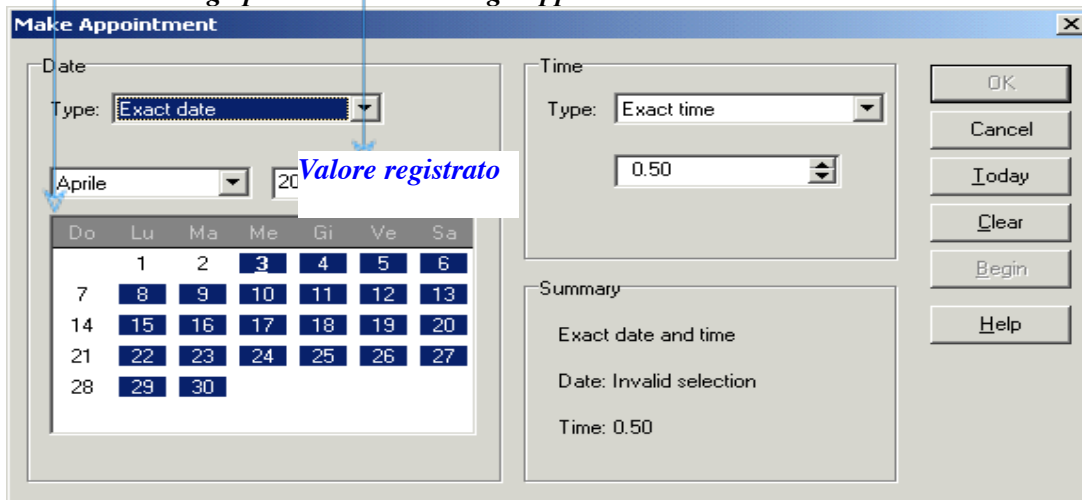
- scegliendo "Appuntamento" si apre un'altra finestra di dialogo dove è possibile inserire tutti i dati dell'appuntamento (data, periodo, ora).

Le due finestre appena descritte sono rappresentate di seguito.

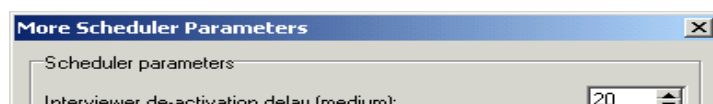
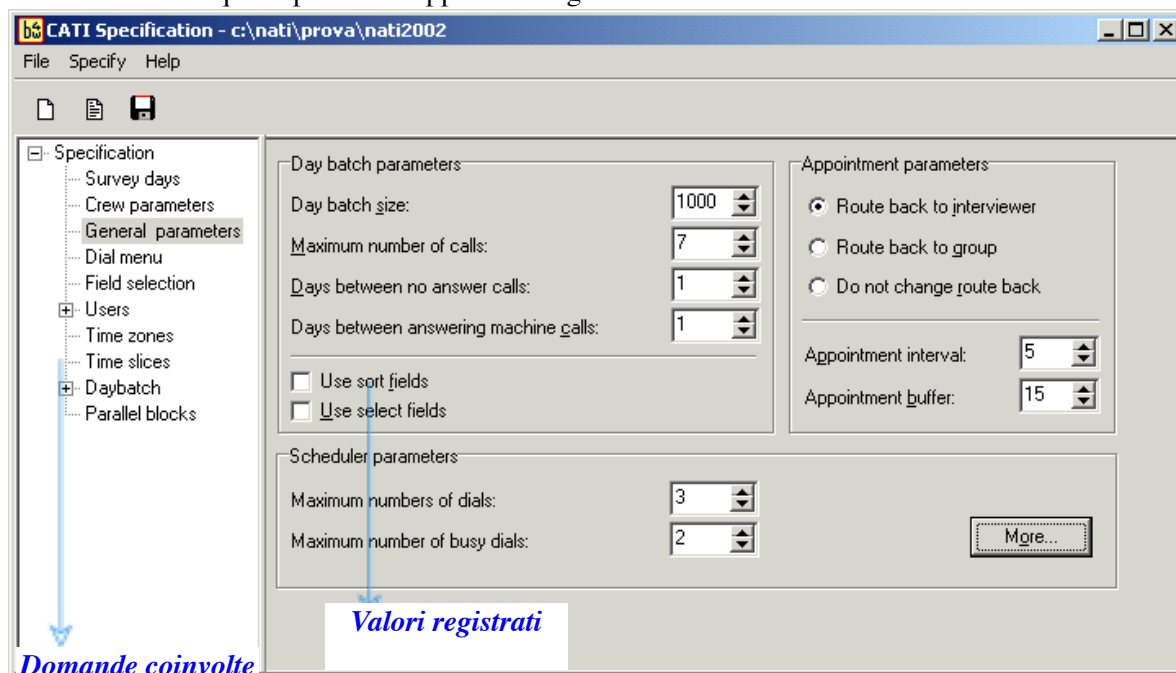
### Finestra di dialogo per l'inserimento del tipo di numero telefonico non corrispondente

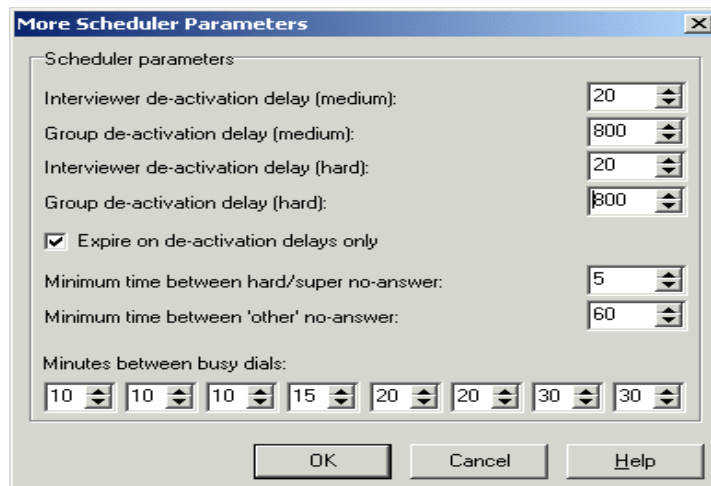


### Finestra di dialogo per l'inserimento degli appuntamenti



La schedulazione delle chiamate dipende dagli esiti delle telefonate e dai parametri di contatto definiti tramite il modulo *Blaise* "CATI specification". La finestra di dialogo di *Blaise* per la valorizzazione di questi parametri appare nel seguente modo.





Nel riquadro a sinistra sono indicati i sottogruppi di parametri che possono essere inseriti; verranno descritti in dettaglio quelli utilizzati per questa indagine.

- *Survey days*: vengono introdotti i giorni di calendario in cui è attiva la rilevazione. Per questa fase di indagine sono state disattivate tutte le domeniche e i giorni relativi alle festività (Pasqua, 25 Aprile e 1° Maggio) ;
- *General parameters*: rappresenta il sottomodulo più ricco, esemplificato nelle due precedenti schermate. Qui vengono infatti definiti i seguenti parametri:
  - ⇒ *Day batch size*: indica il numero di nominativi contattabili per singolo giorno di rilevazione. Questo numero deve essere tale da garantire, da un lato, un tasso di produttività sostanzialmente costante nei vari giorni di rilevazione e, dall'altro, impedire la formazione di code di attesa di numeri telefonici così lunghe da non poter essere evase nell'arco della giornata. Questo parametro è stato valorizzato a 1000 nominativi per la *long form* e 2500 per la *short form*: la diversa entità è legata alla lunghezza del questionario che essendo più breve per la *short form* permette di contattare giornalmente un numero più elevato di individui. La fase pilota ha consentito di migliorare la grandezza del *day batch*, in prima istanza sottodimensionato per una stima più alta della lunghezza effettiva delle interviste. Importante osservare che il primo giorno di rilevazione il *day batch* carica il numero di nominativi specificato prendendoli nell'ordine in cui sono memorizzati nell'archivio dei dati (in questo caso i primi 1000 per la *long form* e i primi 2500 per la *short form*). Dal secondo giorno di rilevazione in poi, carica sempre la stessa quantità di nominativi, ma li ordina in base alle priorità di chiamata che stabilisce in funzione degli esiti dei contatti telefonici del giorno precedente. Ad esempio, il secondo giorno di rilevazione propone per primi gli appuntamenti che scadono in quel preciso giorno e in quella precisa ora, per poi proporre i nominativi che, nella giornata precedente, sono risultati liberi o occupati e, infine, propone i nominativi "nuovi" ossia mai contatti precedentemente.
  - ⇒ *Maximum number of calls*: Indica il numero dei giorni in cui un numero è attivo, ossia contattabile, prima di "cadere" e quindi non essere più chiamato per raggiunto limite di contatti. Il parametro è stato impostato a 7 per fare in modo che ogni numero venisse potenzialmente provato ogni giorno della settimana prima di "cadere". L'esito di questa scelta verrà descritto nel capitolo successivo dei "Principali risultati"
  - ⇒ *Days between no-answer calls/answering machines*: specifica dopo quanti giorni un numero, risultato alla chiamata libero o con segreteria telefonica, deve essere incluso di nuovo nel *day batch*. In questo caso è stato scelto il minimo possibile, ossia 1 giorno di attesa, sempre per fare in modo che ciascun numero venisse potenzialmente provato ciascun giorno della settimana .
  - ⇒ *Maximum number of dials/busy dials*: indica quante volte un numero libero o occupato viene provato nell'arco della giornata. In questo caso, i numeri con esito "occupato" sono stati provati un massimo numero di 6 volte al giorno (3x2=6 tentativi , che costituiscono una *call*), mentre per gli altri esiti non definitivi – libero o segreteria telefonica – i possibili contatti giornalieri sono stati fissati a 3. Anche per questi parametri è stata importante la fase pilota che ci ha consentito di abbassare il numero di tentativi che, scelto inizialmente troppo alto, aveva causato code di attesa troppe lunghe, facendo sì che i numeri non venissero provati tutti i giorni della settimana.
  - ⇒ *Appointment parameters*: permette di gestire gli appuntamenti; per questa indagine è stata scelta l'opzione "*Route back to interviewer*" che consente il ritorno degli appuntamenti all'operatore che li ha presi. Questa scelta è stata effettuata per due ragioni:
    - 1) per una maggiore personalizzazione dell'intervista: il rispondente si trova ad avere un unico riferimento, ossia parla sempre con la stessa persona, ed ha quindi l'impressione di

sceglie l'opzione *Route back to interviewer* che consente il ritorno degli appuntamenti all'operatore che li ha presi. Questa scelta è stata effettuata per due ragioni:

- 1) per una maggiore personalizzazione dell'intervista: il rispondente si trova ad avere un unico riferimento, ossia parla sempre con la stessa persona, ed ha quindi l'impressione di continuità della "conversazione telefonica" soprattutto nel caso in cui l'appuntamento venga preso ad intervista già iniziata;
- 2) per facilitare la ditta esterna nei pagamenti degli operatori (un'intervista conclusa in due momenti da due operatori diversi dovrebbe essere pagata in parti proporzionali alla quantità di domande fatte da ciascun operatore).

Altri parametri sugli appuntamenti sono definiti nel riquadro che si ottiene scegliendo il tasto "*More*". Questi parametri indicano i tempi che un appuntamento resta legato ad un operatore, trascorsi i quali viene "ceduto", in successione, dal rilevatore al suo gruppo di appartenenza, se esiste, e dal gruppo al resto della squadra di rilevazione. In questo caso le scelte fatte sui tempi di attesa e sulla definizione dei gruppi di rilevazione sono dipese dalla lingua in cui veniva condotta l'intervista. In particolare:

- a) non sono stati costituiti gruppi di rilevazione per le interviste in italiano che potevano essere svolte potenzialmente da tutti i 60 operatori;
- b) è stato, invece, definito un gruppo di rilevazione per le interviste in lingua tedesca che potevano essere svolte solo dai tre operatori definiti nel gruppo.

Di conseguenza i tempi di attesa si sono diversificati nel seguente modo:

- a.1) per le interviste in lingua italiana è stato scelto di attendere un tempo prossimo (un po' più alto) alla durata dell'intervista, ossia 20 minuti per la *long form* e 10 minuti per la *short form*. Trascorso questo tempo l'appuntamento viene riproposto ad uno qualunque degli operatori presenti;
- b.1) per le interviste in tedesco, restano validi i 20 e i 10 minuti di attesa prima che un appuntamento passi da un rilevatore ad un altro, questa volta, del gruppo di appartenenza, ed è stato scelto un tempo altissimo di 800 minuti, circa 13 ore, prima che un appuntamento esca dal gruppo: in questo modo si è evitato che un'intervista in tedesco venisse proposta ad un operatore non in grado di sostenerla, creando imbarazzo nell'intervistatore stesso e un'immagine di inefficienza per il rispondente.

⇒ *Minutes between busy dials*: indica i tempi che trascorrono tra esiti telefonici occupati. La scelta effettuata è stata quella di far trascorrere poco tempo, riprovando il numero a distanza ravvicinata: un esito occupato indica infatti che il rispondente è potenzialmente in casa e non bisogna quindi "perdere" l'occasione di contattarlo.

- *Time zones*: indica le fasce orarie di chiamata. In questo caso ne sono state definite due distinguendo i giorni infrasettimanali dal sabato. Per i primi, le interviste potevano effettuarsi dalle 17:30 alle 21:30, mentre per il sabato era attiva la fascia di mattina dalle 9:30 alle 14:30. E' importante osservare che *Blaise* consente di gestire in modo elastico le fasce di chiamata rendendole non obbligatorie per gli appuntamenti; in questo modo è possibile venire incontro alle necessità dei rispondenti stabilendo appuntamenti in orari più consoni alle loro necessità. Tutto questo ha come ovvia conseguenza un aumento, in generale, del tasso di risposta. Inoltre attraverso il modulo *Treat Form* (trattamento dei record) il supervisore di sala può prescindere dai vincoli sugli appuntamenti descritti in precedenza. Questo è particolarmente utile per la gestione degli appuntamenti presi per la mattina (ossia fuori dalla fascia oraria obbligatoria) per i quali è possibile costituire una piccola squadra di rilevatori, anche diversi da quelli che hanno preso gli appuntamenti (altrimenti servirebbero tanti operatori quanti sono gli appuntamenti), che evadono tutte le chiamate.
- *Parallel blocks*: consente di inserire esiti di contatto in corso di intervista. In questo caso i "blocchi paralleli" richiamabili dal questionario permettono di registrare gli esiti di "Rifiuto a collaborare", "Interruzione definitiva" e "Appuntamento".

#### 4.2.3 Gli indicatori di monitoraggio e la funzione di scarico dei dati

Per tenere sotto controllo l'andamento della rilevazione è stato costruito, in accordo con il responsabile di indagine, un insieme molto ricco di indicatori che venivano prodotti giornalmente dalla ditta esterna. Per la produzione degli indicatori di monitoraggio e per il salvataggio giornaliero dei dati si è cercato di facilitare al massimo il lavoro del supervisore di sala; a questo fine è stata installata presso la ditta esterna una funzione che tramite la semplice pressione di un tasto permette di:

- a) salvare i dati delle interviste insieme ad altre informazioni utili (come il file storico dei contatti);
- b) produrre gli indicatori di monitoraggio.

Da un punto di vista tecnico, per la costruzione delle tavole si è cercato, da un lato, di tener conto

- a) salvare i dati delle interviste insieme ad altre informazioni utili (come il numero storico dei contatti),
- b) produrre gli indicatori di monitoraggio.

Da un punto di vista tecnico, per la costruzione delle tavole si è cercato, da un lato, di tener conto delle esigenze specifiche dell'indagine e, dall'altro, di generalizzare il più possibile sia il contenuto delle stesse che il processo di costruzione. In particolare per la loro realizzazione sono stati utilizzati i programmi *Visual Basic 6.0*, *Access* ed *Excel* secondo il seguente flusso: il file ASCII dei dati (file delle interviste o dei contatti a seconda delle tavole da costruire) scaricato direttamente da *Blaise* è stato trasformato in un database *Access* sul quale sono state fatte elaborazioni e *query* mediante *Visual Basic* che ha permesso di esportare i risultati in *Excel*. e di utilizzare quindi anche le potenzialità del foglio elettronico per effettuare ulteriori elaborazioni come calcoli numeri o confronti logici.

Oltre a facilitare il lavoro del supervisore si è cercato di agevolare "l'utente" finale nella lettura dei report: a questo scopo tutti gli indicatori di monitoraggio sono stati organizzati sotto forma di tavole *excel* in modo che potessero essere facilmente consultate a video (senza bisogno di utilizzare necessariamente il supporto cartaceo) e che fosse possibile rielaborare i dati in esse contenuti utilizzando tutte le funzioni offerte dal foglio elettronico. In maggior dettaglio, gli indicatori prodotti sono stati organizzati in sedici tavole *excel* finalizzate a tenere sotto controllo tre macro aspetti dell'indagine:

1. la produttività degli operatori;
2. l'andamento degli esiti di contatto;
3. l'entità numerica dei nominativi disponibili e gli spostamenti fra strato.

#### La produttività giornaliera degli operatori

Tabella.4.1 – Esiti dei contatti telefonici cumulati e giornalieri

Fino al 08/06/02	Totale Cumulato		Lunedì 03/06/02		.....		Sabato 08/06/02	
	v.a.	%	v.a.	%	v.a.	%	v.a.	%
<b>1) INTERVISTA COMPLETA</b>								
2) Numero di telefono inesistente								
2) Numero di telefono errato								
2) Famiglia trasferita								
3) Rifiuto a rispondere								
4) Intervista interrotta definitivamente								
5) Appuntamento								
6) Occupato								
7) Libero								
8) Fax/Segreteria telefonica								
<b>TOTALE CONTATTI</b>								
<i>Numero operatori</i>								
<i>Interviste/uomo/giorno</i>								
<i>Totale ore lavorate effettive</i>								
<i>Interviste/uomo/ora</i>								
<i>Contatti/uomo/ora</i>								

Questa prima tavola consente di osservare quotidianamente la produttività degli operatori telefonici per i quali viene infatti calcolato il numero di interviste giornaliere e orarie (sulla base delle ore di effettivo lavoro) e il numero dei contatti telefonici effettuati in un'ora, che permette di capire l'entità dei "tempi non produttivi" di ciascun operatore. Questa tavola è stata inoltre sviluppata in altre due che consentono di monitorare "l'andamento dei contatti giornalieri per operatore" e "l'andamento dei contatti cumulati per operatore" al fine di verificare se ci sono degli intervistatori che hanno un comportamento molto diverso dalla media. Ad esempio, un numero di interviste orarie troppo alto rispetto alla media indicherebbe che l'operatore o è particolarmente efficiente oppure si comporta in modo poco attinente alle specifiche dettate.

Corredano le informazioni di questa tavola altre due tabelle che consentono di capire i motivi che inducono i rispondenti a non collaborare, rifiutando da subito o interrompendo l'intervista già iniziata.

Corredano le informazioni di questa tavola altre due tabelle che consentono di capire i motivi che inducono i rispondenti a non collaborare, rifiutando da subito o interrompendo l'intervista già iniziata.

Tabella.4.2 – Motivi del rifiuto a collaborare: totale e per operatore

	TOTALE		ISTATN12		ISTATN1		.....
	v.a.	%	v.a.	%	v.a.	.....	
01) Assenza di interesse circa il tema							
02) Dialetto/Scarsa cultura							
03) Disturbati/momento poco opportuno							
04) Malattia/Sordità							
05) Mancanza di tempo							
06) Opinioni negative e sfiducia su indagini							
07) Problemi familiari/Lutto							
08) Bambino deceduto							
09) Bambino fuori target							
10) Sfiducia/Diffidenza/Paura							
11) Stranieri che non parlano italiano							
12) Altro (specificare)							
<b>TOTALE RIFIUTI</b>							

Testo  
introduttivo  
parametrizzato

Tabella.4.3 – Motivi dell'interruzione definitiva: totale e per operatore

	TOTALE		ISTATN12		ISTATN1		.....
	v.a.	%	v.a.	%	v.a.	.....	
1) Assenza di interesse circa il tema							
2) Disturbati/momento poco opportuno							
3) Questionario troppo lungo							
4) Quesito delicato							
5) Quesito difficile da comprendere							
6) Bambino deceduto							
7) Stranieri che non parlano/parlano poco italiano							
8) Altro (specificare)							
<b>TOTALE INTERRUZIONI DEFINITIVE</b>							

Inoltre, per le interruzioni definitive, è stata costruita la seguente tabella che, indicando in quale punto del questionario si sono verificate, permette di capire se esistono particolari sezioni o quesiti critici che inducono i rispondenti a non proseguire la conversazione telefonica.

Tabella.4.4 – Interruzioni definitive per sezione di questionario

	v.a.	% su totale interruzioni
1° SEZIONE: Contesto familiare (moduli)		
2° SEZIONE: Lavoro della madre		
3° SEZIONE: Cura bambino e divisione lavoro familiare		
4° SEZIONE: Abitazione e Contesto socio-economico		
<b>TOTALE INTERVISTE INTERROTTE</b>		

La tabella seguente permette di monitorare l'andamento della durata delle interviste sia per singolo giorno di rilevazione che cumulativamente, riportando la durata media per ciascun operatore, rispetto alle interviste fatte, e i differenziali rispetto alla media generale (rosso se in negativo, nero se in positivo).

giorno di rilevazione che cumulativamente, riportando la durata media per ciascun operatore, rispetto alle interviste fatte, e i differenziali rispetto alla media generale (**rosso** se in negativo , **nero** se in positivo).

Tabella.4.5 – Durata delle interviste per intervistatore

Fino al 06/06/02	numero interviste	Durata media Cumulata	□ rispetto a durata media cumulata totale	numero interviste	Durata media giornaliera	□ rispetto a durata media giornaliera
<b>TOTALE</b>						
ISTATN1						
ISTATN10						
ISTATN11						
ISTATN12						
ISTATN13						
ISTATN14						
.....						

Concludono il primo gruppo di tabelle le seguenti due tavole che riportano i principali tassi di risposta e di non risposta calcolati sul totale e per singolo operatore:

Tabella.4.6 – Tassi cumulati di risposta e non risposta

TASSI	VALORI
Tasso di risposta	
Tasso di interruzione definitivo	
Tasso di rifiuto	
Tasso di caduta	

← *Testo introduttivo parametrizzato*

I tassi indicati sono stati calcolati nel seguente modo:

- Tasso di risposta: interviste complete / (interviste complete + interruzioni definitive + rifiuti);
- Tasso di interruzione definitivo: interruzioni definitive / (interviste complete + rifiuti + interruzioni definitive);
- Tasso di rifiuto: rifiuti / (interviste complete + rifiuti + interruzioni definitive);
- Tasso di caduta: (rifiuti + interruzioni definitive + numeri errati + caduti per raggiunto nr. contatti) /

(interviste complete + rifiuti + interruzioni definitive + numeri errati);

Con questa tabella si è voluto monitorare l'andamento dei precedenti tassi per ciascun operatore riportando anche i differenziali rispetto alla media generale.

Tabella 4.7 – Tassi cumulati di risposta e non risposta per intervistatore

	Tasso di risposta		Tasso di interruzione		Tasso di rifiuto		Tasso di NON risposta	
	Tassi	Delta	Tassi	Delta	Tassi	Delta	Tassi	Delta
<b>TOTALE</b>								
ISTATN1								
ISTATN10								
ISTATN11								
ISTATN12								
ISTATN13								
ISTATN14								
.....	.....	.....	.....	.....	.....	.....	.....	.....

Andamento degli esiti di contatto

Al secondo gruppo di indicatori appartengono le due seguenti tabelle che permettono di analizzare la distribuzione degli esiti per fascia oraria (intervalli di 15 minuti) e per giorno di rilevazione al



### Andamento degli esiti di contatto

Al secondo gruppo di indicatori appartengono le due seguenti tabelle che permettono di analizzare la distribuzione degli esiti per fascia oraria (intervalli di 15 minuti) e per giorno di rilevazione al fine di verificare la presenza di particolari concentrazioni. Come descritto nel capitolo sui “Principali Risultati” queste tavole sono state utili non solo per monitorare il lavoro degli intervistatori, ma anche per capire l’adeguatezza delle scelte fatte sui parametri influenti sulla schedulazione delle chiamate.

*Tabella.4.8– Distribuzione ultimi esiti per tipologia e fascia oraria (15 minuti)*

FASCE ORARIE		Intervista	Rifiuto	Interruzione e definitiva	Occupato	Appuntamento	Libero o Segreteria	Numero errato	TOTALE
9.00.01	9.15.00								
9.15.01	9.30.00								
....	.....								
21.15.01	21.30.00								
TOTALE CONTATTI									

*Tabella.4.9– Distribuzione ultimi esiti per tipologia e giorno di rilevazione*

Settimana che inizia il 03/06/2002	Intervista	Rifiuto	Interruzione e definitiva	Occupato	Appuntamento	Libero o Segreteria	Numero errato	TOTALE
Lunedì								
Martedì								
.....								
Sabato								
TOTALE CONTATTI								

### Entità numerica dei nominativi disponibili e gli spostamenti fra strato.

L’ultimo gruppo di indicatori consente di tenere sotto controllo “l’utilizzo di nominativi” per capire quanti ne restano disponibili per le interviste e di analizzare gli spostamenti tra strato delle unità campionarie a seguito di variazioni dei dati anagrafici. Questo gruppo di tabelle è stato particolarmente utile in questa prima fase di indagine dove, come già osservato, la non conoscenza a priori del tasso di caduta ha comportato la necessità di definire le “riserve” solo dopo aver contattato tutti i nominativi “base” del campione.

*Tabella.4.10 – Ultimi esiti dei contatti telefonici*

	v.a	%
1) INTERVISTA COMPLETA		
2) Numero di telefono inesistente		
2) Numero di telefono errato		
2) Famiglia trasferita		
3) Rifiuto a rispondere		
4) Intervista interrotta definitivamente		
5) Numeri caduti per raggiunto limite dei contatti (occupato, libero, segreteria)		
6) Numeri sospesi totali di cui:		
a) appuntamento		
b) occupato		
c) libero		
d) fax/segreteria telefonica		
TOTALE		

*Tabella.4.11– Sintesi situazione nominativi disponibili e non disponibili*

TOTALE NOMINATIVI FORNITI	v.a.
Numeri telefonici non più disponibili(1+2+3+4+5)	
Numeri telefonici ancora disponibili(6+..)	

<b>TOTALE NOMINATIVI FORNITI</b>	v.a.
Numeri telefonici non più disponibili(1+2+3+4+5)	
Numeri telefonici ancora disponibili(6+..)	
-di cui già contattati	
-di cui mai contattati	

Tabella.4.12– Nominativi caduti e sospesi per l'individuazione delle riserve per singolo strato

Strato	Interviste Programmate	Interviste Concluse	Interviste Mancanti	Nominativi da Sostituire			Mai Contattati
				Totale	Caduti	Sospesi	

Tabella.4.13 – Matrice degli spostamenti fra strati delle interviste

Strato a posteriori	1	2	3	...	.....
Strato a Priori					
1					
2					
3					

Le tavole di monitoraggio appena descritte rappresentano un ampliamento delle statistiche offerte da *Blaise*, il quale permette di osservare *on-line* l'andamento della rilevazione offrendo un insieme standard di indicatori, rappresentati di seguito.

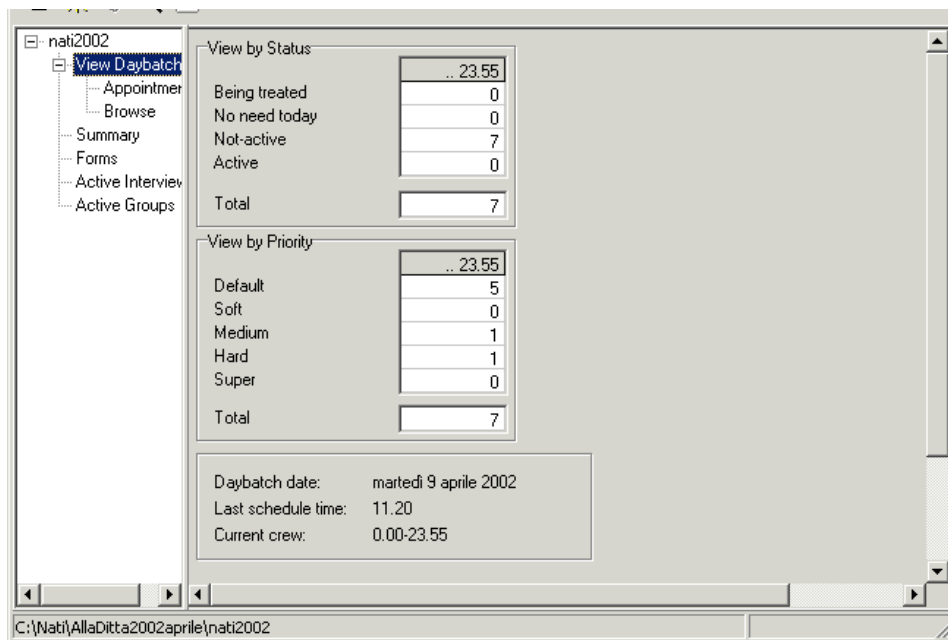
The screenshot shows the 'Cati Management' application window. The left sidebar contains a tree view with 'nati2002' expanded, showing options like 'View Daybatch', 'Appointment', 'Browse', 'Summary', 'Forms', 'Active Interview', and 'Active Groups'. The main area is divided into two sections: 'Not yet done' and 'Done'. Each section contains a table of call counts categorized by status (e.g., No call, No answer, Busy, Appointment, Answering machine) and the number of calls (0, 1, 2, 3, 4, 5, >5). A blue arrow points to the 'Done' section with the text 'Data inserita'.

Questa prima schermata mostra la situazione giornaliera dei nominativi, suddivisi, rispetto agli ultimi esiti di contatto, in due riquadri: nel primo, *Not yet done*, sono indicati i nominativi ancora lavorabili, con l'ultimo esito assunto fino a quel momento; nel secondo *Done* sono riportati i nominativi non più contattabili, con l'esito definitivo di contatto.

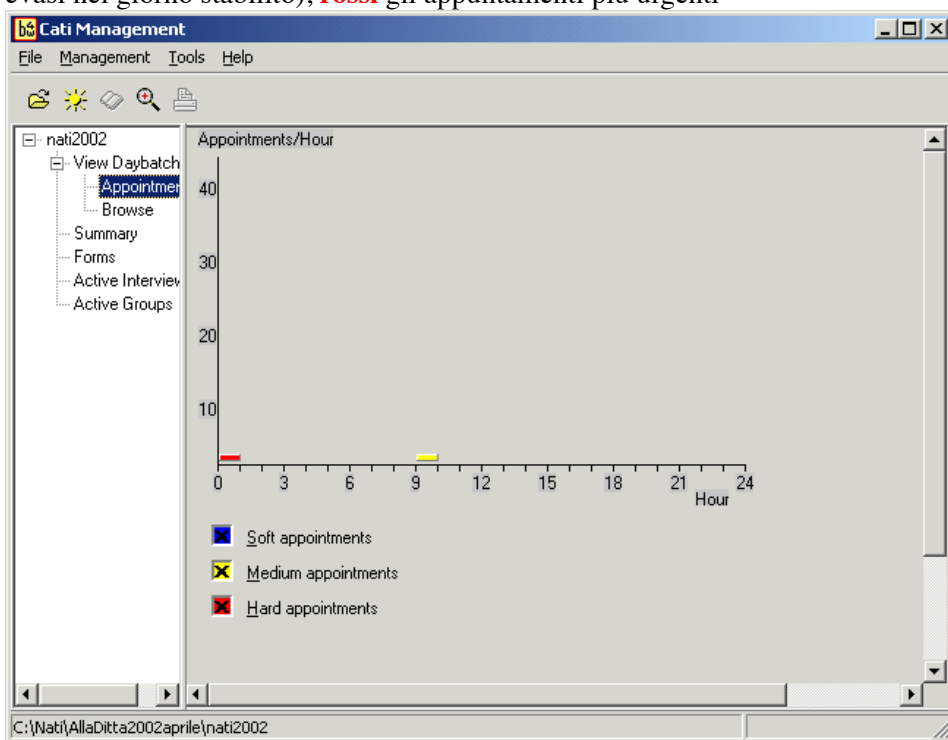
La successiva schermata indica quanti dei nominativi disponibili saranno proposti dallo schedatore in un determinato giorno di rilevazione (*View by status*) e con quale priorità (*View by priority*)

The screenshot shows the 'Cati Management' application window with the 'View by Status' option selected. The main area displays a table with the following data:

Being treated	.. 23.55
No need today	0



Questa finestra mostra la rappresentazione grafica degli appuntamenti per giorno di rilevazione e fascia oraria: il colore è utilizzato per distinguere gli appuntamenti in funzione della priorità con cui verranno proposti dallo schedatore: **blu** e **gialli** con priorità più bassa (appuntamenti presi e non evasi nel giorno stabilito), **rossi** gli appuntamenti più urgenti



#### 4.2.4 L'utility per la riorganizzazione della base dati

Oltre alle funzioni di carico-scarico dati e produzione della reportistica giornaliera è stata installata presso la ditta esterna una utility di Blaise che consente di riorganizzare la base dati. Attraverso le utility interne di Blaise - "Hospital" e "Blaise-to-Blaise" è stato possibile infatti:

- verificare se il database era danneggiato e in tal caso procedere con il recovery delle informazioni;
- "riorganizzare" le informazioni del database delle interviste in modo da ottimizzare l'accesso ai nominativi da parte dello schedatore.

### 4.3 Le fasi precedenti l'indagine a regime

In questo capitolo sono descritte le fasi che sono state propedeutiche all'indagine a regime, vale a dire:

- la fase di pre-test;
- il briefing teorico-tecnico agli intervistatori telefonici;
- la fase pilota.

#### 4.3.1 Il pre-test

Il pre-test dell'indagine è stato svolto internamente all'Istat, utilizzando la struttura CATI (cinque postazioni) del laboratorio CASIC, gestite dal servizio MPS/C.

I principali obiettivi di questa fase di sperimentazione sono stati quelli di:

Il pre-test dell'indagine è stato svolto internamente all'Istat, utilizzando la struttura CATI (cinque postazioni) del laboratorio CASIC, gestite dal servizio MPS/C.

I principali obiettivi di questa fase di sperimentazione sono stati quelli di:

- testare il questionario elettronico, verificandone l'esattezza dei numerosi svincoli e l'adeguatezza dei controlli sia di coerenza interna che con le informazioni contenute nell'archivio dati (mod. Istat P4);
- stimare la durata media dell'intervista che, soprattutto nella versione long, si temeva ne potesse pregiudicare l'esito positivo.

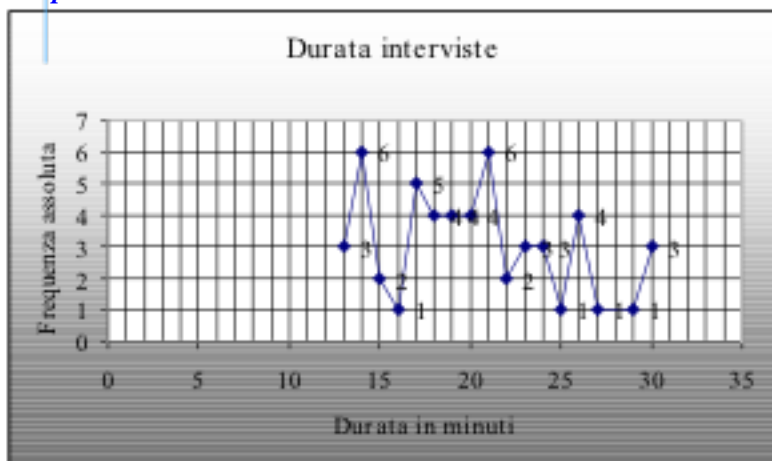
Mediante il pre-test è stato anche possibile provare l'intero percorso preliminare alla realizzazione dell'intervista al fine di ottenere dei suggerimenti sul modo di operare per effettuare l'abbinamento dei numeri telefonici con i nominativi selezionati e per preparare le istruzioni necessarie alla formazione delle intervistatrici. Inoltre è stato possibile ottenere alcune indicazioni di massima sull'andamento dei possibili esiti del contatto telefonico che hanno mostrato la buona propensione dei rispondenti a collaborare e la necessità di utilizzare la lettera di preavviso a firma del Presidente per ridurre la quota di rifiuti più alta nel pre-test che nell'indagine a regime.

Il pre-test, svolto su un numero di rispondenti molto contenuto (53 interviste complete), ma sufficiente a stimare la durata media e a testare il questionario elettronico, è stato effettuato utilizzando la versione *long form* cui erano legate maggiori difficoltà sia in termini di implementazione del questionario elettronico che di conduzione dell'intervista.

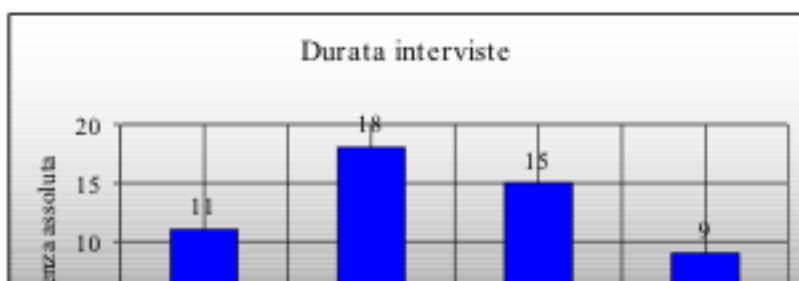
Le interviste del pre-test sono state effettuate da quattro rilevatori donne (in modo conforme a quanto sarebbe accaduto per le interviste della fase estensiva svolte solo da operatori di sesso femminile) nei giorni 21, 24, 25 e 26 settembre 2001 dalle ore 10:00 alle ore 13:30 per la fascia antimeridiana e dalle ore 14:00 alle ore 17:30 per la fascia pomeridiana. La fase di rilevazione è stata preceduta dal *briefing* alle operatrici, istruite sulle finalità e i contenuti dell'indagine e sugli aspetti tecnici di gestione dell'intervista.

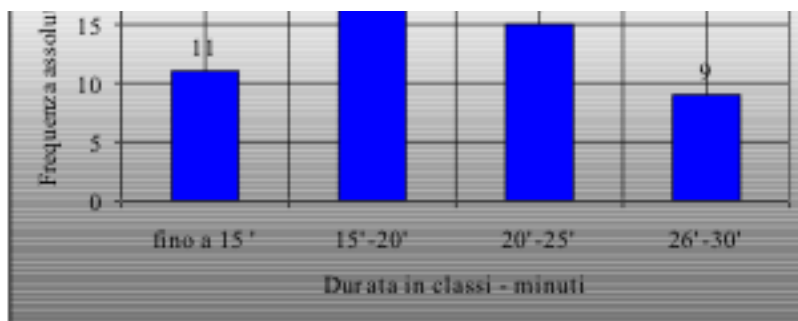
#### 4.3.1.1 Principali risultati del pre-test

Il pre-test ha consentito di stimare la durata delle interviste telefoniche: contro una durata media prevista di 30-35 minuti, se ne è ottenuta una di 20 minuti (versione *long form*), con una variabilità di  $\pm 5$  minuti. E' importante osservare che la differenza tra la durata attesa e quella effettiva è da attribuirsi non solo all'eliminazione di alcune domande che appesantivano l'intervista, ma anche all'inaspettata partecipazione delle intervistate che hanno risposto volentieri alle domande che hanno percepito, a volte, vicine ai loro problemi: rispondere a quesiti riguardanti il bambino e il lavoro della donna ha rappresentato, in alcuni casi, una sorta di "sfogo" per l'intervistata, una *Finestra di scelta della lingua* ad una voce cortese, ma sconosciuta, i propri problemi avrebbero avuto sul tasso di collaborazione all'indagine.



La distribuzione semplice della durata delle interviste del pre-test, riportata nel grafico sottostante, mostra che la punta massima si è attestata sui 30 minuti per 3 casi e la punta minima sui 13 minuti per altrettante interviste; si osservano, inoltre, due valori modali in corrispondenza delle durate "14 minuti" e "21 minuti" che si sono verificate per 6 interviste.





Volendo raggruppare le

frequenze in classi, è possibile osservare come la classe più consistente sia quella delle interviste comprese tra i 15 e i 20 minuti.

Come previsto, la durata registrata nel pre-test si è dimostrata più alta di quella ottenuta nella fase a regime non solo per la diversa lunghezza del periodo di rilevazione che incide, in modo proporzionale, sull'efficienza degli operatori, ma soprattutto a seguito delle modifiche apportate al questionario.

E' questo il secondo e importante risultato conseguito nella fase di pre-test che ha permesso di capire se la struttura del questionario nella sua interezza o delle singole domande dovesse essere modificata e di verificare la presenza di particolari concentrazioni su alcuni *item* di risposta delle domande. In particolare dal pre-test è emerso che:

- non sono state necessarie modifiche riguardanti la struttura del questionario, ossia la sequenza delle sezioni tematiche: l'intervista tocca vari aspetti della vita della donna e del bambino senza creare salti logici nel passare da un argomento ad un altro. E' importante sottolineare che già prima di iniziare le telefonate del pre-test, durante le simulazioni di intervista, sono state eliminate tre domande della Sezione IV sulla tematiche sociali in quanto si è osservato che oltre ad allungare troppo i tempi di intervista risultavano particolarmente pesanti, aumentando il rischio di ottenere "Interruzione Definitiva" come esito definitivo. Le tre domande in questione, sono:

*IV10. Come sappiamo la nascita di un bambino comporta delle spese aggiuntive, sia dirette che indirette. Nel suo caso, quali delle seguenti spese si sono verificate dopo la nascita di (nome)?*

Si tratta di una domanda a risposta sollecitata, per la quale dovevano essere lette al rispondente le 9 modalità di risposta previste.

*IV12. Quali sono le principali voci di spesa che ha ridotto (.. per affrontare nuove spese)?*

Si tratta anche in questo caso di una domanda a risposta sollecitata con 6 *item* di risposta, di cui l'ultimo contenente "Altro, specificare"

*IV13. Come sappiamo la nascita di un bambino comporta spesso dei cambiamenti nello stile di vita delle persone. Nel suo caso, mi può indicare con quale frequenza si dedicava e si dedica alle seguenti attività prima e dopo la nascita di (nome)?*

Domanda a risposta sollecitata con 13 *item* di risposta ognuno dei quali doveva essere proposto due volte, ossia "prima della nascita" e "dopo la nascita" del bambino, come indicato nella seguente tabella:

	PRIMA DELLA NASCITA			DOPO LA NASCITA		
	<i>Mai</i>	<i>Poco</i>	<i>Spesso</i>	<i>Mai</i>	<i>Poco</i>	<i>Spesso</i>
ANDARE AL CINEMA						
ANDARE A TEATRO						
USCIRE CON GLI AMICI						
ASCOLTARE MUSICA						
LEGGERE LIBRI						
VEDERE LA TV						
VOLONTARIATO						
ASCOLTARE LA RADIO						
FREQUENTARE CORSI (CUCINA, MUSICA, LINGUA,)						
ANDARE A PRANZO E/O CENA FUORI CASA						
FARE VACANZE						

ANDARE A PRANZO E/O CENA FUORI CASA						
FARE VACANZE						
FARE SHOPPING						
FARE SPORT						

- erano da migliorare i salti logici tra alcune domande in quanto derivanti da risposte non ipotizzate inizialmente. Ad esempio, la domanda “I24. Suo marito ha assistito al parto? (1=Si, 2=No, 3=Non è stato possibile)” veniva posta a tutte le intervistate. Si è capito, successivamente, che in caso di parto cesareo prevale l’uso di non far assistere i padri; quindi, si è scelto di non porre la domanda alle donne che avevano subito un cesareo, visto che le intervistate appartenenti a questa tipologia rispondevano con un “... ovviamente no” quasi a sottolineare l’ovvietà della domanda;
- occorre apportare modifiche ad alcune domande. Si è trattato di modifiche di tipo diverso, ossia:
  - riformulazione del testo della domanda;
  - cambiamento di struttura della domanda: da sollecitata – lettura degli *item* di risposta – a spontanea; da sollecitata in forma tabellare - singola risposta per *item* - a sollecitata con scelta di uno o più codici di risposta;
  - aggiunta di *item* di risposta non previsti in fase di stesura del questionario o modifica di quelli esistenti.

Per maggior chiarezza si riportano degli esempi.

**Esempio di riformulazione del testo e di cambiamento della struttura della domanda:**

III0. Le elencherò ora delle motivazioni per avere un lavoro. Mi può indicare per ciascuna di esse quelle che sono vere per lei?

	Vero	Falso
Il mio lavoro mi rende indipendente	1	2
La mia occupazione mi interessa e mi coinvolge	1	2
Il lavoro mi consente di avere rapporti interpersonali gratificanti	1	2
Il lavoro mi consente di uscire dalla routine domestica	1	2
Lavoro per contribuire al bilancio familiare	1	2

Ad ogni motivazione è stata data la risposta “vero” in circa il 90% dei casi. Questo ci ha indicato che una tale formulazione rendeva la domanda abbastanza scontata. Di conseguenza è stato riformulato il testo e cambiata la struttura del quesito: non più in forma tabellare, ma con scelta di una sola risposta, ossia:

III0. Le elencherò adesso delle motivazioni per avere un lavoro. Mi potrebbe indicare quella che è più importante per lei?

Il mio lavoro mi rende indipendente	1
La mia occupazione mi interessa e mi coinvolge	2
Il lavoro mi consente di avere rapporti interpersonali gratificanti	3
Il lavoro mi consente di uscire dalla routine domestica	4
Lavoro per contribuire al bilancio familiare	5

**Esempio di aggiunta di alternative di risposta:**

III15. Di solito, chi si occupa prevalentemente degli altri bambini?( una sola risposta)

Io stessa	01	Una baby sitter	08
Compagno/marito/padre	02	Asilo nido pubblico	09
I nonni	03	Asilo nido privato	10
Fratelli o sorelle	04	Altri tipi di asilo	11
Altri familiari	05	Scuola materna	12
Degli amici	06	Altro (specificare)	13
I vicini	07		

Ai tredici *item* iniziali è stata aggiunta, dopo il pre-test, la risposta “Nessuno, sono autosufficienti”. Quindi, attraverso la semplice aggiunta di un *item* di risposta è stato possibile considerare anche il caso di bambini che non necessitano di assistenza da parte degli adulti – caso non previsto inizialmente – senza dover “filtrare” questa domanda in funzione dell’età dei bambini – condizione abbastanza complessa in quanto legata alle diverse variabili “data di nascita dei figli conviventi”-;

- non si sono verificate particolari concentrazioni su particolari *item* di risposta rappresentati da:

caso non previsto inizialmente – senza dover “filtrare” questa domanda in funzione dell’età dei bambini – condizione abbastanza complessa in quanto legata alle diverse variabili “data di nascita dei figli conviventi”-;

- non si sono verificate particolari concentrazioni su particolari *item* di risposta rappresentati da:
  - il “Rifiuto”, “Non so/Non ricordo”;
  - le prime o ultime modalità delle domande a risposta sollecitata con un elevato numero di alternative di risposta.

Per i “Rifiuti” e i “Non so/Non ricordo”, citati con una frequenza bassissima, non si sono osservate concentrazioni né su determinate domande, né all’interno di specifiche sezioni del questionario, in quanto si sono distribuiti all’interno di tutto il questionario. Alcune indicazioni in merito a questi due particolari tipi di risposta sono emerse anche dal *de-briefing* con le operatrici (descritto di seguito) le quali hanno constatato non solo la presenza di un “certo” numero di “Non so/Non ricordo” per il quesito “anno di conseguimento del titolo di studio” della madre (evento confermato nell’indagine a regime come sarà descritto nel capitolo “Principali risultati “), ma anche, contrariamente alle nostre aspettative, una certa difficoltà delle donne intervistate di ricordare sul momento la data del matrimonio. Per la fase a regime è stato deciso di:

- mantenere il quesito sulla data di conseguimento del titolo di studio;
- modificare il quesito inerente la “data di matrimonio” in quanto considerato fondamentale per il buon fine dell’intervista (la mancata risposta indica l’interruzione dell’intervista stessa) perché legato alla definizione di “figlio legittimo”. In particolare, per ovviare all’inconveniente della mancata risposta parziale, è stato deciso di ammettere il “Non ricordo” come una delle possibili risposte e di somministrare, in questo caso, una nuova domanda in grado di stabilire la legittimità del bambino. La nuova domanda in questione è: *(Solo se Non ricorda data di matrimonio)*

**18.2 (Bambino) è nato dopo il vostro matrimonio? ( 1= Sì, 2= No)**

Con questa scelta si è quindi data la precedenza alla riduzione del rischio di “Interruzione definitiva” dell’intervista scontando il prezzo di una lieve disomogeneità nel modo di raccogliere le informazioni. Infatti per coloro che ricordano la data del matrimonio, la legittimità del bambino verrà stabilita a posteriori con un confronto tra date (data di nascita del bambino e data del matrimonio), mentre per coloro che non la ricordano sarà stabilita direttamente dalla risposta “Sì” alla precedente domanda.

Anche per il secondo aspetto, relativo alle domande con molte modalità di risposta, non si sono verificate particolari concentrazioni nei risultati. In realtà aggregazioni sulle prime alternative di risposta si sono verificate su tre domande, ossia:

**II6.\_1 Per quale motivo lavora a tempo parziale?**

1. Impossibilità di trovare un lavoro con orario diverso	2 casi
2. Maggior tempo da dedicare alla cura della casa, dei figli e di me stessa	8 casi
3. Studio o formazione professionale	1 caso
4. Motivi di salute personale	--
5. Motivi di salute di altri familiari	--
6. Preferenza personale	--
7. Altro specificare	--

**II.16.1 Perché non è restata a casa ulteriormente?**

1. Ho usufruito del periodo massimo di astensione dal lavoro	2 casi
2. Per esigenze economiche	8 casi
3. Il mio lavoro richiedeva la mia presenza	2 casi
4. Il capo o i colleghi mi hanno fatto dei problemi	--
5. Temevo che un'assenza prolungata potesse pregiudicare la carriera	--
6. Altro	--

**IV7.1 Che tipo di difficoltà economiche ha dovuto affrontare la sua famiglia dopo la nascita del bambino?**

1. Affitto	--
2. Mutuo o debito per la casa	2 casi
3. Spese per la casa	6 casi
4. Acquisti rateali	--

1. Attitto	--
2. Mutuo o debito per la casa	2 casi
3. Spese per la casa	6 casi
4. Acquisti rateali	--
5. Spese per l'automobile	--
6. Spese sanitarie per me o il mio compagno/per il bambino	--
7. Altre spese per il bambino (cibi speciali, pannolini Ecc.)	2 casi
8. Altro (specificare)	-

Si tratta però di situazioni previste in fase di stesura del questionario: le modalità di risposta sono state infatti ordinate mettendo ai primi posti quelle più “probabili”. Questo sistema è stato adottato sia per ovviare alla impossibilità di *Blaise* di ruotare le alternative di risposta, sia per abbreviare i tempi di intervista offrendo alle rispondenti la possibilità di “trovare” subito la risposta più consona alla propria situazione senza dover attendere la lettura di tutte le alternative possibili.

#### 4.3.1.2 Il de-briefing con gli operatori nella fase di pre-test

Al termine delle interviste del pre-test è stato organizzato un de-briefing con le intervistatrici dal quale, nonostante la limitatezza del campione oggetto dello studio, è emerso un denominatore comune alle singole esperienze di intervista del quale prendere atto per la stesura del questionario definitivo. Questo strumento è di estrema importanza in fase di pre-test in quanto consente di cogliere, insieme alla possibilità offerta da *Blaise* di inserire i campi note, anche aspetti puramente qualitativi dell'indagine difficilmente percepibili dalla semplice raccolta dei dati. Le indicazioni fornite dalle rilevatrici hanno consentito di effettuare alcuni cambiamenti sul questionario, nonché di tener conto di particolari aspetti durante l'organizzazione dell'indagine a regime. In particolare:

- secondo le operatrici le madri dimostravano l'esigenza di avere più informazioni a proposito dell'indagine rispetto a quelle fornite nell'introduzione all'intervista, non tanto per diffidenza nei confronti dell'interlocutore quanto per un vero e proprio interesse sull'argomento. Questa esigenza ha portato non solo ad un arricchimento del testo della lettera di preannuncio, ma anche alla redazione di un supplemento introduttivo nel questionario elettronico che le operatrici della fase a regime potevano leggere qualora l'intervistata chiedeva ulteriori dettagli informativi;
- è emersa la necessità di verificare la pertinenza di alcune domande in situazioni particolari e la corretta localizzazione delle stesse all'interno del questionario, onde evitare situazioni di perplessità a rispondere da parte dell'intervistata cui venivano posti quesiti non perfettamente coerenti con quanto dichiarato precedentemente. Le domande in questione riguardavano unicamente la sezione del lavoro della madre e della cura del bambino, ossia:
  - il quesito sul grado di soddisfazione rispetto alla “garanzia del posto di lavoro” e “al rapporto con i colleghi o con i superiori” chiesto anche nel caso di attività lavorativa in proprio;
  - i quesiti su “L'operatore digita il testo nella casella: ad esempio inizia a digitare Rom.. (Roma)”, sul “numero di ore, giornali”, sul “numero di affidamento” e sulla “spesa mensile per il bambino solo occasionalmente”;
  - le domande sul “grado di soddisfazione sull'affidamento del bambino all'asilo nido” poste anche nel caso in cui il bambino ha appena iniziato a frequentare queste strutture;
- le intervistatrici hanno inoltre messo in evidenza la necessità di modificare un controllo di tipo *hard* tra l'anno di “inizio-fine lavoro” e “l'anno di nascita del bambino” in modalità *soft*, in quanto risultava troppo stringente rispetto alle situazioni reali;
- infine, anche dal *de-briefing*, è stata segnalata l'opportunità di modificare la modalità di somministrazione di alcune domande, passando da quesiti a risposta sollecitata a quesiti a risposta spontanea. Tale suggerimento si è rilevato particolarmente importante non solo perché ha permesso di snellire il questionario e di contenere la durata delle interviste, ma anche di avvicinare l'intervista ad un dialogo tra due persone; ciò è stato confermato dall'osservazione delle intervistatrici inerente il fatto che molto spesso, la madre anticipava la risposta prima di aver ascoltato la lettura di tutte le alternative possibili.

In sintesi, è possibile osservare che quanto detto in questi paragrafi dimostra, in generale, l'importanza del pre-test: grazie alla sperimentazione sul campo è possibile, infatti, considerare diverse sfaccettature della realtà di un fenomeno, difficilmente individuabili durante la definizione “a tavolino” del questionario di indagine.

#### 4.3.2 Il briefing agli intervistatori

Prima dell'inizio della rilevazione, sono state organizzate due mezze giornate di *briefing* cui hanno partecipato le rilevatrici, i supervisor di sala e i capi progetto della ditta esterna, i responsabili dell'indagine e dello sviluppo del software di acquisizione dati. La prima giornata è stata suddivisa in due parti:

- nella prima parte è stato organizzato un *briefing* puramente teorico finalizzato a descrivere gli obiettivi di indagine,



Prima dell'inizio della rilevazione, sono state organizzate due mezze giornate di *briefing* cui hanno partecipato le rilevatrici, i supervisori di sala e i capi progetto della ditta esterna, i responsabili dell'indagine e dello sviluppo del software di acquisizione dati. La prima giornata è stata suddivisa in due parti:

- nella prima parte è stato organizzato un *briefing* puramente teorico finalizzato a descrivere gli obiettivi di indagine, il contenuto del questionario e il significato delle domande;
- nella seconda parte è stato effettuato un *briefing* teorico-tecnico per mostrare il funzionamento dei controlli e far conoscere tutte le tipologie di incompatibilità previste dal questionario, al fine di far capire alle intervistatrici l'atteggiamento da tenere verso l'intervistata e il modo di gestire tecnicamente l'intervista in caso di digitazione di un dato anomalo o errato.

La seconda giornata è stata incentrata ad analizzare gli aspetti puramente tecnici dell'indagine. Sono state organizzate anche in questo caso due sessioni di *briefing*:

- la prima ha trattato vari aspetti della metodologia CATI, dalla descrizione di come e quando inserire i possibili esiti di contatto, alla gestione degli appuntamenti, all'uso dei campi note e dei tasti funzione per le mancate risposte parziali;
- nella seconda sono state effettuate simulazioni di intervista dove i rispondenti (generalmente colleghi Istat) sono stati istruiti per dare risposte particolari che potessero generare situazioni, nelle quali diventava di fondamentale importanza la competenza e la professionalità del rilevatore.

#### 4.3.3 La fase pilota

Come già osservato nei paragrafi precedenti, la fase pilota è stata molto importante per migliorare le impostazioni dei parametri di schedulazione e per testare l'impalcatura costruita intorno al questionario elettronico, vale a dire la codifica assistita, la produzione degli indicatori di monitoraggio e le *utility* di scarico dati. Pochissimi i cambiamenti apportati al questionario elettronico a seguito della pilota: a parte la modifica di un percorso per comprendere una casistica particolare e l'ampliamento del *range* di validità di due variabili, il questionario, già modificato e validato nella fase di pre-test, non ha subito modifiche né contenutistiche né strutturali. In maggior dettaglio gli interventi migliorativi effettuati dopo la fase pilota sono stati i seguenti:

- modifiche ai parametri di schedulazione:
  - è stata ampliata la dimensione del *day batch*, vale a dire del numero di nominativi lavorabili giornalmente, ossia messi in linea con il questionario elettronico lavorabili. Questo ampliamento è dipeso sostanzialmente da una durata media delle interviste più contenuta di quanto stimato nel pre-test che ha ovviamente consentito la lavorazione giornaliera di un maggior numero di nominativi;
  - sono stati diminuiti i possibili tentativi per i numeri con esiti di chiamata non definitivi (libero, occupato, segreteria): in questo modo sono state ridotte le code di attesa dei singoli record la cui entità nella fase pilota aveva impedito a qualche nominativo di essere provato il massimo numero di volte stabilito;
  - è stato ampliato il numero di giorni di lavorazione dei numeri da 4 a 7, non solo per fare in modo che i numeri venissero provati tutti i giorni della settimana, ma anche per incrementare il tasso di risposta in quanto molti contatti si sono trasformati in esiti positivi proprio intorno al terzo-quarto giorno di rilevazione;
- modifiche al sistema di codifica assistita: per evitare rallentamenti e, in qualche caso, blocchi di rete è stata modificata la struttura di base della codifica delle variabili "Comune" e "Stato estero", spostando dal *server* ai singoli *client* le tabelle dei dati alla base della codifica;
- modifiche agli indicatori di monitoraggio: è stato migliorato il contenuto di due tavole:
  - sulla tavola degli esiti di contatto è stata inserita la specificazione di "Numero errato", "Numero inesistente" e "Famiglia trasferita", come ampliamento degli esiti di contatto "Numero telefonico non corrispondente";
  - sulla tavola di calcolo della durata media delle interviste per operatore è stato inserito il numero di interviste al fine di tener conto anche dei giorni di presenza/assenza di ciascun intervistatore.

#### 4.4 L'indagine a regime

Al termine di tutte le fasi descritte precedentemente si potuto iniziare l'indagine a regime, che ha visto coinvolte quotidianamente 60 intervistatrici, delle quali 20 per le interviste in versione *long* e 40 per le *short*.

Lo schema seguente riassume, da un punto di vista tecnico, tutti i passaggi effettuati prima di poter entrare nel vivo nell'indagine, insieme ai *software* utilizzati per la loro realizzazione.

L'insieme di questi strumenti, già disponibile in Istat, costituisce un pacchetto facilmente personalizzabile a seguito della modifica di appositi parametri.

Reperimento numeri  
di telefono

Rilevazione



#### *4.4.1 Alcuni problemi sorti in fase di rilevazione*

Durante la rilevazione si sono verificati alcuni problemi di carattere tecnico dipendenti sia dalla struttura di rete offerta dalla ditta che da alcune procedure di *Blaise*. Non si può negare il fatto che la loro soluzione abbia richiesto un certo impegno da parte dell'Istat, ma bisogna anche puntualizzare che: *a)* trattandosi della prima volta in cui questa strategia veniva applicata non era possibile prevedere a monte tutti i fattori problematici; *b)* che sulla base delle esperienze fatte le prossime applicazioni richiederanno sicuramente uno sforzo ed un impegno in costante diminuzione.

Dal punto di vista degli operatori, l'impatto con un nuovo software di acquisizione dati è stato positivo: le giornate di *training* sono state sufficienti a far capire il funzionamento del sistema.

Si sono verificati problemi in relazione alla struttura di rete della ditta esterna, la cui risoluzione ha richiesto l'intervento da parte dell'Istat di un esperto di rete e di un esperto di Windows che hanno provveduto a:

- dare indicazioni per ottimizzare l'intera architettura della rete della ditta esterna;
- suggerire lo spostamento delle tabelle per la codifica assistita dal *server* ai singoli *client*;
- consigliare la necessità di incrementare lo spazio sul disco rigido del *server* per consentire una più efficiente gestione del traffico di rete.

Inoltre è stata implementata dal servizio Istat responsabile di *Blaise* un'altra *utility* che, dopo lo scarico dati e l'avvio della procedura *Blaise-to-Blaise*, accerta che questi due passaggi abbiano avuto un esito positivo (conteggio dei record memorizzati nell'archivio delle interviste) e invia automaticamente all'Istat una *e-mail* contenente l'esito della procedura. In questo modo è possibile tenere sotto controllo la delicata fase di salvataggio dati senza richiedere l'intervento, non sempre garantito e più che altro non sempre efficiente, del personale della ditta esterna.

## **5 Principali risultati**

garantito e più che altro non sempre efficiente, del personale della ditta esterna.

## 5 Principali risultati

### 5.1 Analisi degli esiti finali di contatto

La prima fase dell'indagine campionaria sulle nascite, che ha avuto una durata di 43 giorni per le interviste *long form* e 41 per le interviste *short*, si è conclusa con il 69% e 67,6% di interviste complete rispettivamente per la versione *short* e *long* del questionario (cfr. tabella 5.1).

Tabella 5.1 Esiti finali del contatto telefonico

Esiti	Short		Long	
	V.A.	%	V.A.	%
1) INTERVISTE CONCLUSE	11734	69.0	5735	67.6
2) NUMERO NON CORRISPONDENTE	3388	19.9	1765	20.8
<i>di cui</i>				
<i>numero di telefono inesistente</i>	214	1.3	103	1.2
<i>numero di telefono errato</i>	2873	16.9	1497	17.7
<i>famiglia trasferita</i>	298	1.7	165	1.9
3) RIFIUTO A RISPONDERE	668	3.9	393	4.6
4) INTERVISTA INTERROTTA DEFINITIVAMENTE	209	1.2	115	1.4
5) NUMERI CADUTI PER RAGGIUNTO LIMITE DEI CONTATTI	1009	6.0	472	5.6
<b>TOTALE</b>	<b>17006</b>	<b>100.0</b>	<b>8480</b>	<b>100</b>

In generale non si sono verificate differenze significative tra gli esiti di contatto delle due versioni. In particolare:

- sul totale degli esiti finali, circa il 20% è costituito da numeri non corrispondenti. Analizzando le tre modalità che hanno concorso a determinare una percentuale così elevata, notiamo che circa l'85% dei numeri non corrispondenti è costituito da telefoni errati. Nelle prossime fasi di indagine si provvederà ad analizzare i nominativi con numero errato in funzione del criterio di ricerca utilizzato dalla ditta per l'abbinamento del telefono. Per la prima fase di indagine non è stato possibile effettuare questo tipo di controllo per la mancanza dell'informazione sul criterio di ricerca su tutti i nominativi abbinati. Inoltre, la standardizzazione degli indirizzi, descritta precedentemente (c.f.r. par. 4.2.1), consentirà non solo di incrementare la percentuale di abbinamento, ma anche di ridurre, probabilmente, l'entità dei numeri errati;
- la percentuale di Rifiuti a collaborare è molto bassa: si attesta infatti intorno al 4%. Analogamente, la percentuale di interviste interrotte è risultata significativamente bassa (circa l'1% del totale nominativi contattati). Inoltre, poiché le interruzioni si sono verificate soprattutto sulle prime domande del questionario, è possibile assimilarle a dei rifiuti a collaborare. Da questi risultati che, come già detto, prescindono dalle versioni *long* e *short*, può dedursi che la lunghezza dell'intervista non è stata un problema; pertanto, essendo evidentemente l'argomento gradito e

l'intervista scorrevole nonostante la delicatezza di alcuni dei temi trattati, se si disponesse di un *budget* sufficiente, sarebbe possibile utilizzare esclusivamente il questionario nella sua versione estesa.

Scendendo in dettaglio sulle motivazioni di rifiuti a collaborare ed interruzioni definitive (cfr. tabelle 5.2 e 5.3), si osserva che:

- per entrambe le versioni il motivo prevalente è "l'assenza di interesse sul tema d'indagine", ulteriore conferma dell'analogia tra rifiuto e interruzione e quindi del fatto che il questionario viene portato a termine nella maggior parte dei casi;
- in particolare poi, all'interno dei rifiuti a collaborare, un'altra motivazione che ha pesato sulle mancanze di volontà a collaborare è stata la "mancanza di tempo" (11% *short* e 21% *long*);
- nell'ambito delle interruzioni definitive dell'intervista, invece, motivazioni frequenti sono state anche "il momento poco opportuno" (10% *short* e 13% *long*) e "la lunghezza del questionario" (13% *long*).

Tabella 5.2. Motivi del rifiuto a collaborare

Motivo del rifiuto	Long	Short
--------------------	------	-------

long) e “la lunghezza del questionario” (13% long).

Tabella 5.2. Motivi del rifiuto a collaborare

Motivo del rifiuto	Long		Short	
	v.a	%	v.a	%
Assenza di interesse circa il tema	146	37.1	264	39.6
Dialetto / Scarsa cultura	13	3.3	22	3.3
Disturbati / Momento poco opportuno	26	6.6	47	7.0
Malattia / Sordità	2	0.5	6	0.9
Mancanza di tempo	82	20.9	75	11.2
Opinioni negative e sfiducia su indagini	18	4.6	44	6.6
Problemi familiari / Lutto	13	3.3	15	2.2
Bambino deceduto	14	3.6	23	3.4
Bambino fuori target/non riconosciuto	33	8.4	72	10.8
Sfiducia / Diffidenza / Paura	24	6.1	48	7.2
Stranieri che non parlano italiano	11	2.8	40	6.0
Non so	11	2.8	12	1.8
<b>Totale</b>	<b>393</b>	<b>100</b>	<b>668</b>	<b>100</b>

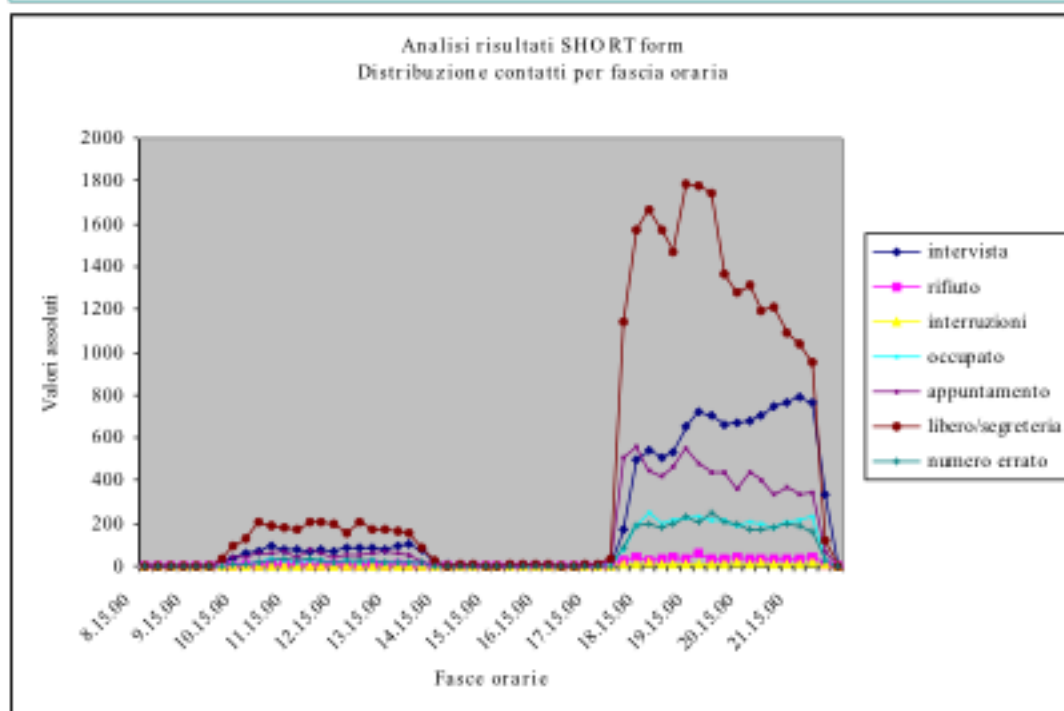
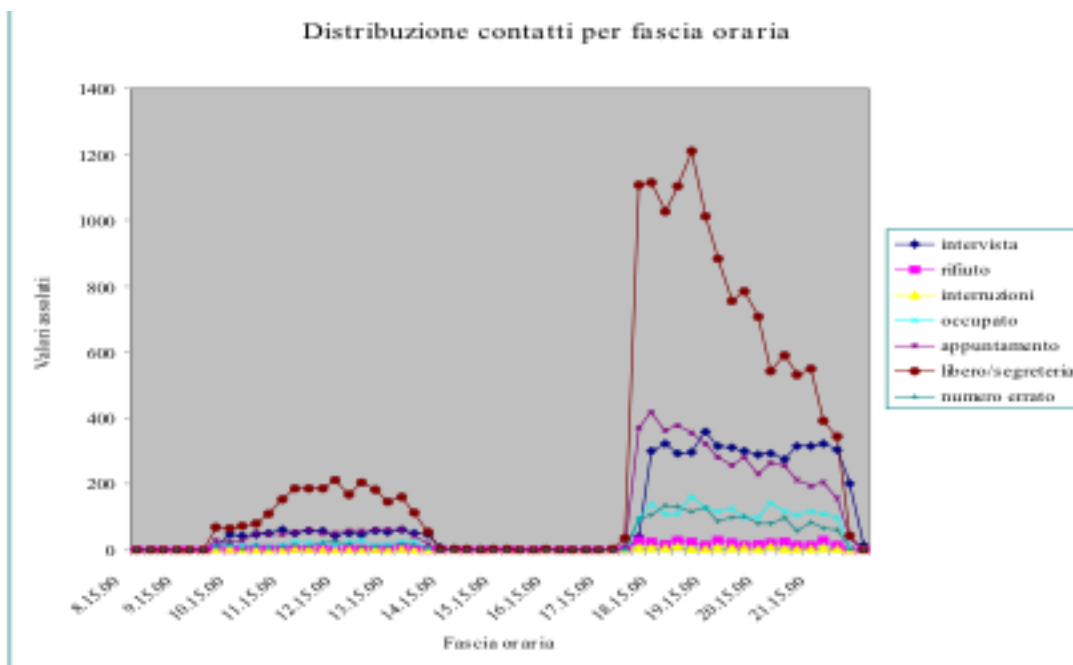
Tabella 5.3. Motivi dell'interruzione definitiva

Motivo dell'interruzione definitiva	Long		Short	
	v.a	%	v.a	%
Assenza di interesse circa il tema	46	40.0	75	35.9
Disturbati / Momento poco opportuno	15	13.0	21	10.0
Questionario troppo lungo	15	13.0	11	5.3
Quesito delicato	4	3.5	13	6.2
Quesito difficile da comprendere	2	1.8	2	1.0
Bambino deceduto	6	5.2	16	7.6
Bambino fuori target/non riconosciuto	9	7.8	34	16.3
Stranieri che non parlano/parlano poco italiano	4	3.5	8	3.8
Non so	14	12.2	29	13.9
<b>Totale</b>	<b>115</b>	<b>100</b>	<b>209</b>	<b>100</b>

L'indicazione che può trarsi da questi risultati è quella di insistere particolarmente sulla formazione delle operatrici per la fase di presentazione dell'indagine e, soprattutto quando l'intervistata dichiara di non aver tempo o di trovare l'intervista eccessivamente lunga, nell'incentivazione a proporre appuntamenti, presentando anche la prospettiva di realizzare l'intervista in più telefonate.

- anche la percentuale di caduti per raggiunto limite di contatti risulta accettabile (circa il 6%). Infatti l'analisi più dettagliata di tutti gli esiti di contatto ha dimostrato che i parametri di schedulazione delle chiamate impostati hanno consentito un'efficiente distribuzione dei tentativi per giorno e fascia oraria di rilevazione funzionale al raggiungimento di un tasso risposta elevato. Come risulta dai grafici seguenti, l'andamento degli esiti nelle varie fasce orarie non presenta irregolarità e raggiunge un picco elevato solo in corrispondenza dell'esito "libero, segreteria" nelle prime ore pomeridiane stabilite per l'intervista.
- la produttività oraria e giornaliera osservata è buona, il che conferma una buona gestione dei tentativi di contatto; infatti, le intervistatrici hanno prodotto una media di 15 contatti all'ora per le interviste *long form* e 23 per quelle *short form*. Dei 15 contatti all'ora per le interviste *long form*, una media di 3 contatti ha portato ad un'intervista completa; analogamente, per le interviste *short form*, dei 23 tentativi di contatto all'ora, 5 hanno portato ad un'intervista completa. La produttività giornaliera è stata di 7.9 interviste complete al giorno per intervistatrice sul questionario *long form* e 14.2 per il questionario *short*.

Analisi risultati LONG form  
Distribuzione contatti per fascia oraria



A ulteriore conferma del buon andamento della rilevazione sono i valori ottenuti per i tassi di risposta e non risposta calcolati per questa indagine (cfr. tabella 5.4). Si può infatti osservare che, prescindendo dagli altri esiti che non comportano un contatto con l'intervistata (numeri non corrispondenti, numeri caduti per raggiunto limite di contatti), è assolutamente preponderante la proporzione di donne che portano a termine l'intervista (92% per le *long* e 93% per le *short*). Tale risultato è conforme con quanto già osservato in fase di pre-test, ed è attribuibile alla disponibilità e al piacere delle madri nel rispondere ai quesiti sul bambino e sulla propria vita lavorativa.

Tabella 5.4. Tassi di risposta e tassi di caduta

	Long Form	Short Form
Tasso di risposta	91.9	93
Tasso di rifiuto	6.3	5.3
Tasso di interruzione definitiva	1.8	1.7
Tasso di caduta	32.4	31.0

L'elevato tasso di caduta è invece dovuto alla forte incidenza dei numeri non corrispondenti, in particolare ai numeri errati; come già accennato sono in corso azioni migliorative (analisi dei criteri di abbinamento e normalizzazione degli indirizzi) che ci consentiranno verosimilmente di ridurre l'entità a favore della rappresentatività del campione rispetto all'universo.

Volendo fare un'analisi comparata con altre indagini Istat condotte con metodologia CATI, si

particolare ai numeri errati; come già accennato sono in corso azioni migliorative (analisi dei criteri di abbinamento e normalizzazione degli indirizzi) che ci consentiranno verosimilmente di ridurre l'entità a favore della rappresentatività del campione rispetto all'universo.

Volendo fare un'analisi comparata con altre indagini Istat condotte con metodologia CATI, si osserva che le percentuali di interviste complete, rifiuti a collaborare e interruzioni definitive sono sostanzialmente in linea con quanto ottenuto nell'indagine in questione (Muratore, Quattrococchi, Sabbadini, in corso di pubblicazione) (cfr. tabella 5.5).

Tabella 5.5 Esiti del contatto telefonico di alcune indagini Istat

	Nascite Short	Nascite Long	Vittimizzazione	Viaggi e vacanze	Diplomati
Interviste complete	69.0%	67.6%	57.0%	61%	60.8%
Rifiuti	3.9%	4.6%	15.5%	13%	2.7%
Interruzioni definitive	1.2%	1.4%	0.5%	2.5%	0.6%

### 5.1.1 Analisi degli esiti per ricezione lettera e ripartizione geografica

La lettera, a firma del Presidente, di presentazione dell'intervista ha notoriamente un impatto positivo sulla collaborazione all'indagine. Infatti, anche in questo caso, la percentuale di interviste concluse è molto più alta fra le donne che hanno ricevuto la lettera (hanno ricevuto la lettera l'82% delle intervistate – *long* e *short*). In maggior dettaglio (cfr. tabelle 5.6 e 5.7):

- nella *long form*:
  - il 96% delle madri che hanno ricevuto la lettera ha concluso l'intervista contro il 79.1% di quelle che dichiarano di non averla ricevuta o non lo ricordano;
  - circa il 3% di chi ha ricevuto la lettera ha rifiutato di collaborare contro un 15.5% di chi non l'ha ricevuta;
  - le interruzioni definitive sono state circa dell'1% e del 5% rispettivamente fra la madri che hanno ricevuto la lettera e quelle che non l'hanno ricevuta;
- analogamente nella *short form*:
  - il 97.3% delle madri che hanno ricevuto la lettera ha concluso l'intervista contro l'80.9% di quelle che dichiarano di non averla ricevuta o non lo ricordano;
  - circa il 2% di chi ha ricevuto la lettera ha rifiutato di collaborare contro un 14.4% di chi non l'ha ricevuta ;
  - le interruzioni definitive sono state dell'1% e del 4,7% rispettivamente fra la madri che hanno ricevuto la lettera e quelle che non l'hanno ricevuta.

Tabella 5.6. Esiti finali del contatto telefonico per ricezione della lettera - LONG FORM

Valori percentuali	RICEZIONE LETTERA	
	Si	No
Interviste concluse	96.0	79.1
Rifiuto a collaborare	2.8	15.5
Interruzione definitiva	1.2	5.3

Tabella 5.7. Esiti finali del contatto telefonico per ricezione della lettera - SHORT FORM

Valori percentuali	RICEZIONE LETTERA	
	Si	No
Interviste concluse	97.3	80.9
Rifiuto a collaborare	1.7	14.4
Interruzione definitiva	1.0	4.7

Analizzando la distribuzione degli esiti di contatto per ripartizione geografica, la percentuale di interviste concluse risulta superiore al 76% per l'Italia settentrionale e centrale (circa 78% per le *long* e 79% per le *short*), mentre non supera il 60% degli esiti nel Mezzogiorno (58% per le *long* e 60% per le *short*). Questa differenza è attribuibile alla proporzione elevata di numeri errati (30% per le *long* e 28% per le *short*) riscontrati in questa ripartizione pari a circa il triplo di quella delle rimanenti aree geografiche. In prospettiva, quindi, per quest'area dovranno essere affinati i criteri di abbinamento dei numeri telefonici con gli anagrafici.

Tabella 5.8. Esiti finali del contatto telefonico per ripartizione geografica - LONG FORM

Valori percentuali	Ripartizione geografica	Totale
--------------------	-------------------------	--------

Tabella 5.8. Esiti finali del contatto telefonico per ripartizione geografica - LONG FORM

Valori percentuali	Ripartizione geografica				Totale
	Nord-Ovest	Nord-Est	Centro	Sud/Isole	
Interviste concluse	76.6	79.5	76.3	58.1	67.6
Non risponde nessuno	3.4	3.2	3.9	4.9	4.2
Occupato	0.4	0.1	0.1	0.4	0.3
Appuntamento	0.6	0.4	0.5	0.5	0.5
Rifiuto	5.9	4.2	4.3	4.5	4.6
Segreteria	0.8	0.4	0.7	0.4	0.5
Errato	10.9	10.8	12.5	30.0	20.8
Interruzione definitiva	1.4	1.4	1.7	1.2	1.4
<b>Totale</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

Tabella 5.9. Esiti finali del contatto telefonico per ripartizione geografica - SHORT FORM

Valori percentuali	Ripartizione geografica				Totale
	Nord-Ovest	Nord-Est	Centro	Sud/Isole	
Interviste concluse	77.8	80.6	77.6	60.1	69.0
Non risponde nessuno	5.1	2.7	5.0	5.4	4.9
Occupato	0.2	0.0	0.3	0.1	0.2
Appuntamento	0.3	0.5	0.4	0.6	0.5
Rifiuto	4.9	4.1	3.5	3.8	3.9
Segreteria	0.5	0.7	0.5	0.3	0.4
Errato	10.0	10.2	11.6	28.4	19.9
Interruzione definitiva	1.2	1.2	1.1	1.3	1.2
<b>Totale</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

### 5.1.2 Analisi della durata delle interviste

La durata media delle interviste è stata di 13,4 e 6.0 minuti per le interviste svolte rispettivamente con la versione *long* e *short* del questionario. Tale durata è risultata inferiore a quella riscontrata nel pre-test svolto nel laboratorio CASIC dell'Istat e ciò può essere attribuito principalmente ai miglioramenti apportati al questionario alla luce del pre-test stesso, nonché al fatto che la maggiore lunghezza del periodo di rilevazione ha indubbiamente inciso positivamente sull'efficienza delle intervistatrici. Il pre-test ha suggerito, infatti (cfr. par. 4.3.1.1), importanti modifiche ad alcune domande per migliorare la comprensione e la fluidità del questionario. Le modifiche apportate sono state di tre tipi:

- riformulazione del testo di alcune domande;
- cambiamento di struttura di alcune domande: da sollecitata – lettura degli item di risposta – a spontanea; da sollecitata in forma tabellare - singola risposta per item - a sollecitata con scelta di uno o più codici di risposta;
- aggiunta di item di risposta non previsti in fase di stesura del questionario o modifica di quelli esistenti.

Inoltre, relativamente all'esperienza acquisita dalle singole rilevatrici, è da ritenere che lo *skill* degli intervistatori, all'aumentare delle interviste fatte, abbia concorso alla riduzione della durata delle interviste.

Tabella 5.10. Durate medie delle interviste rispetto ad alcuni fattori

Durate in minuti	Short	Long
<u>Ripartizione geografica</u>		
Nord Ovest	5.5	14.3
Nord Est	6.2	14.5
Centro	6.0	14.0
Sud e Isole	5.5	13.0
<u>Condizione professionale</u>		
Occupata	5.5	15.2

<u>Condizione professionale</u>		
Occupata	5.5	15.2
Non occupata	6.1	11.4
<u>Numero di figli</u>		
Un figlio	5.3	13.0
Due figli	6.0	14.0
Tre figli	6.3	14.1
Quattro e oltre	6.5	15.4

Se si analizza la durata delle interviste alla luce di alcune caratteristiche (cfr. tabella 5.10) emerge che:

- non si scorgono forti differenze territoriali; infatti, per la versione *long form* la durata media varia di circa 2 minuti passando da un minimo di 13 minuti nel sud/isole ad un massimo di 14,5 minuti nel nordest, mentre per la versione *short form* la durata media è sostanzialmente identica;
- ciò che maggiormente incide sulla durata delle interviste *long* è la condizione professionale delle madri intervistate: l'essere occupate allunga l'intervista di oltre 3 minuti. Inoltre, tra le donne occupate il picco più alto di durata spetta a coloro che nella loro vita lavorativa hanno cambiato lavoro in seguito alla gravidanza: per questo gruppo la durata media si attesta infatti a 16 minuti contro 11 minuti delle non occupate e i 13 della media poiché a queste tipologie di donna vengono somministrati più quesiti.
- il numero di figli conviventi, invece, ha un'incidenza marginale nelle interviste *long*, dove, un figlio in più pesa pochi secondi in termini di durata. Lo scarto più grande si verifica oltre i 4 figli e ciò, si verifica inaspettatamente anche nelle *short*, in cui non vengono raccolte informazioni sui figli conviventi; ciò potrebbe essere imputato alla maggiore disattenzione che l'intervistata presta per l'intervista a causa di concorrenti impegni familiari.

### 5.1.3 Analisi dei "non so, non ricordo", "rifiuto a rispondere" e "Altro-specificare"

Esaminando come si sono distribuiti i "non so, non ricordo" e i "rifiuto a rispondere" tra i quesiti che ne ammettevano il ricorso, è stato osservato che, in generale, raramente si è ricorso a queste modalità di risposta e, in particolare, che:

- la percentuale di "non so, non ricordo" è stata compresa tra un minimo di 0,01%, peraltro osservato nella maggior parte dei casi, a un massimo di 29.9% nelle interviste *short form*, e tra 0,01% e 27.6% nelle interviste *long form*.
- le percentuali dei "rifiuti a rispondere" è stata molto bassa nelle interviste *short*, andando da un minimo di 0,01% a un massimo di 1,5%; nelle interviste *long form*, la percentuale di rifiuto a rispondere varia da un minimo di 0,01% a un massimo di 10.9%.

Analizzando le risposte ai singoli quesiti delle interviste concluse è emerso che:

- i quesiti inerenti tempi e durate hanno presentato le percentuali più alte di "non so, non ricordo". Tra le donne intervistate col questionario *short form*, il 29,8% ha dichiarato di non ricordare l'anno di conseguimento del titolo di studio, il 4,3% la durata del soggiorno in Italia prima della residenza e il 2,9% le settimane di gravidanza in caso di parto prematuro; analogamente, tra le intervistate col questionario *long form*, il 27.6% delle donne ha dichiarato di non ricordare l'anno di conseguimento del titolo di studio, il 20% la durata del soggiorno del padre del bambino in Italia prima della residenza e il 18.9% la durata del congedo parentale di cui ha usufruito la madre.

Per i quesiti relativi al durata del congedo parentale e della gravidanza si è quindi pensato di incentivare la possibilità di rispondere in termini di mesi, piuttosto che di settimane; per l'anno di conseguimento del titolo di studio, si sta invece pensando di dare la possibilità di rispondere in termini di età dell'intervistata al momento del conseguimento del titolo;

- tra le interviste col questionario *short form*, le percentuali di "rifiuto a rispondere" sono state molto più basse delle aspettative e non hanno quindi richiesto particolari interventi correttivi; le più elevate sono state registrate sul quesito del riconoscimento del figlio da parte del padre (1.5%) e sulla diagnosi di eventuali malformazioni alla nascita (0,8%). Nelle interviste col questionario *long form* le domande che sono risultate più delicate, registrando le percentuali più



- molto più basse delle aspettative e non hanno quindi richiesto particolari interventi correttivi; le più elevate sono state registrate sul quesito del riconoscimento del figlio da parte del padre (1.5%) e sulla diagnosi di eventuali malformazioni alla nascita (0,8%). Nelle interviste col questionario *long form* le domande che sono risultate più delicate, registrando le percentuali più consistenti di "rifiuto a rispondere", sono state quelle sul lavoro; in particolare, sull'eventualità di una regolare retribuzione in caso di assenza dal lavoro per motivi di salute del bambino da parte del padre (10.9%) e della madre (5.5%) e sulla volontà di cercare un lavoro in futuro (4.9);
- i quesiti delicati sulla storia riproduttiva della donna non sono risultati problematici, a prescindere dalla versione del questionario. L'interruzione dell'intervista che si sarebbe verificata in caso di "rifiuto a rispondere" e "non so, non ricordo" a questi quesiti, si è registrata mediamente, contrariamente alle aspettative, solo in due casi.

Dallo studio delle modalità "Altro-specificare" presenti nelle domande a risposta predefinita, è emerso che alcuni quesiti non sono stati compresi correttamente, spesso già dalle intervistatrici. Tra gli 'Altro', infatti, sono state spesso registrate:

- risposte immediatamente riconducibili ad una delle modalità predefinite;
- risposte ambigue non coerenti con la domanda, che necessitavano di un chiarimento o di un'ulteriore domanda da parte dell'intervistatrice.

Un quesito che è risultato particolarmente problematico è stato quello sulla posizione nella professione in cui le intervistatrici, anziché chiarire all'intervistata che non si richiedeva qui la professione, ma, qualora si fosse dipendenti, se si fosse inquadrati come dirigenti, direttivi, ecc....(e corrispondente casistica per i lavoratori in proprio), hanno descritto in "Altro-specificare" la professione svolta dalla madre o dal padre del bambino.

Altre domande che hanno presentato un uso errato della modalità "Altro-specificare" con una frequenza rilevante sono stati:

➤ *Mi può indicare che grado di parentela hanno con lei gli altri familiari con cui vive?*

GENITORE DELLA DONNA	1
GENITORE DEL CONIUGE/COMPAGNO	2
FRATELLI DELLA DONNA O DEL COMPAGNO	3
ALTRI FAMILIARI	4
COLLABORATORI FAMILIARI	5
ALTRO (SPECIFICARE)	6

In altro specificare sono state riportate le risposte già presenti nelle modalità:

- Cognata\o [fratelli della donna o del compagno]
- zii della madre [altri familiari]
- baby-sitter [collaboratori familiari]

➤ *Perché non vuole altri figli in futuro?*

PER MOTIVI DI ETA'	1
PERCHE' HO RAGGIUNTO IL NUMERO DI FIGLI DESIDERATO	2
I FIGLI SONO COSTOSI, SPECIALMENTE QUANDO CRESCONO	3
I FIGLI RENDONO PIÙ DIFFICILE AVERE UN LAVORO	4
NON CI SAREBBE TEMPO PER ALTRE COSE IMPORTANTI DELLA VITA	5
LA GRAVIDANZA, IL PARTO E LA CURA DEI FIGLI SONO PESANTI PER UNA DONNA	6
LA CRESCITA DEI FIGLI PORTA MOLTE PREOCCUPAZIONI E PROBLEMI	7
AVERE UNA CASA NON ADATTA PER UNA FAMIGLIA PIÙ NUMEROSA	8
NON POTER CONTARE SULL'AIUTO COSTANTE DI PARENTI ED AMICI	9
ESSERE ASSORBITO DA ONERI DI CURA DI ALTRI FAMILIARI	10

ED AMICI

ESSERE ASSORBITO DA ONERI DI CURA DI ALTRI	10
FAMILIARI	
IL LEGAME DELLA COPPIA NON E' ABBASTANZA FORTE	11
Altro (specificare)	12

In altro specificare sono state riportate le risposte già presenti nelle modalità:

- Gravidanza difficile [gravidanza, parto e cura dei figli sono pesanti]
- Motivi di parto
- Molto faticoso crescerli
  
- Sono suff. quelli che ho [Ho raggiunto il numero di figli desiderato]
- Va bene così

➤ *Perché si è licenziata*

NON ERA SODDISFATTA DELLA POSIZIONE LAVORATIVA	1
NON ERA SODDISFATTA DELLA RETRIBUZIONE O DEI 2	
GUADAGNI	
NON ERA SODDISFATTA DEL TIPO DI LAVORO CHE SVOLGEVA	3
L'AMBIENTE DI LAVORO NON LA FACEVA SENTIRE A SUO AGIO	4
GLI ORARI DEL PRECEDENTE LAVORO ERANO INCONCILIABILI	5
CON L'ORGANIZZAZIONE FAMILIARE	
LA SEDE DI LAVORO ERA TROPPO DISTANTE DALLA SUA 6	
ABITAZIONE	
CI SONO ALTRI MOTIVI? (Specificare)	7

I seguenti "altri" motivi dovevano essere registrate nella modalità 5:

- Stare con figlia
- Accudire bambina
- Per la bimba [orari inconciliabili con  
l'organizzazione familiare]
- Per la nascita del bambino
- Per stare dietro alla famiglia
- Perché ha avuto la bimba

➤ *Perché ha scelto di affidare il bambino a.....*

PER L'APPROCCIO EDUCATIVO	1
PER FAR SOCIALIZZARE IL BAMBINO	2
PER LA CONVENIENZA ECONOMICA	3
PER LA FIDUCIA	4
PER LA FLESSIBILITÀ DEGLI ORARI	5
PER COMODITÀ	6
PER NON ESPORRE IL BAMBINO AL RISCHIO DI FREQUENTI	7
MALATTIE	
PER L'IGIENE	8
PER LA QUALITÀ DELLE CURE DATE AL BAMBINO	9
NON AVEVO ALTRE ALTERNATIVE	10
CI SONO ALTRI MOTIVI? (specificare)	

Dovevano essere ricondotte alla modalità preesistenti i seguenti "altri" motivi:

- per non gravare sui nonni
- sta bene col nonno [comodità]
- venire incontro al desiderio del bambino

- per non gravare sui nonni  
sta bene col nonno [comodità]
- venire incontro al desiderio del bambino
- costo alto dei nidi  
economica] [convenienza]
- le strutture sono più costose
- non avevo familiari cui lasciarli, mancanza nidi  
alternative] [non avevo altre
- non avevo altre alternative

L'analisi di queste risposte ha portato ad un ulteriore approfondito *briefing* alle intervistatrici e, in alcuni casi, a migliorare il *wording* delle modalità di risposta.

### 5.2 Confronto con i dati della rilevazione degli iscritti in anagrafe per nascita

Le variabili che provengono dalla rilevazione dei nati iscritti in anagrafe (mod. Istat P4) contenute nell'archivio dati alla base della rilevazione e che vengono sottoposte a controllo, come precedentemente descritto, sono:

- la data di nascita della madre;
- la data di nascita del padre;
- la data di nascita del bambino;
- il sesso del bambino;
- lo stato civile della madre;
- lo stato civile del padre.

E' stata fatta un'analisi per comune della frequenza con cui le informazioni provenienti dalla rilevazione anagrafica sono risultate divergenti rispetto a quelle rilevate in corso di intervista. Si ricorda che, dal momento che in caso di divergenza questa veniva fatta presente all'intervistata, chiedendole di confermare o meno quanto risposto, si può assumere che l'informazione corretta sia quella rilevata, mentre il dato di fonte anagrafica possa essere affetto da errore.

Nel computo dei possibili errori non abbiamo considerato la variabile "stato civile" che potrebbe essere effettivamente cambiata nel periodo intercorso tra la rilevazione anagrafica e l'indagine. Tale analisi è finalizzata all'ottimizzazione del processo di acquisizione dell'indagine anagrafica, in quanto, tramite il monitoraggio del lavoro dei comuni nella fase di acquisizione dei dati, è possibile intervenire su quelli che hanno presentato la percentuale più elevata di dati errati. Dall'analisi effettuata è emerso che la variabile affetta da errore nella più elevata percentuale dei casi è stata la data di nascita del padre del bambino, seguita dalla data di nascita della madre, del bambino e dal sesso di quest'ultimo.

Tabella 5.11 Errori medi riscontrati nell'indagine di fonte anagrafica

Variabile	Percentuale di errore
la data di nascita della madre	2.9
la data di nascita del padre	5.6
la data di nascita del bambino	1.2
il sesso del bambino	0.5

Sono state inoltre calcolate le percentuali di errore per singolo comune, così da intervenire eventualmente su quelli più problematici. Al di là poi dei singoli comuni, è stata effettuata un'analisi per tipologia di comune, diversificandoli a seconda che si tratti di capoluoghi di regione oppure no. Si è infatti ritenuto che le procedure di raccolta e acquisizione dei dati dei grandi comuni possano diversificarsi da quelle dei piccoli, in funzione della maggiore mole di lavoro e delle diverse professionalità coinvolte nel processo. Come emerge dalla seguente tabella, infatti, i comuni capoluoghi di regione riescono più difficilmente a controllare il processo di acquisizione in modo da garantire elevati livelli di qualità.

Tabella 5.12 Percentuali media di errori riscontrate nell'indagine di fonte anagrafica per tipologia di comune

	Totale interviste	Percentuale media di errore

di comune

	Totale interviste	Percentuale media di errore
Comuni capoluoghi di regione	2376	8.8
Altri comuni	15093	3.9

### 5.3 Modifiche di alcuni controlli soft (signal)

Come descritto nel paragrafo 4.1.2, il programma prevede dei *signal* in corrispondenza di controlli che manifestano o la presenza di incompatibilità tra le informazioni raccolte o l'inserimento di dati che, seppur possibili, risultano poco probabili e fanno quindi ipotizzare che l'informazione raccolta sia frutto di un errore di digitazione.

Analizzando i dati rilevati è emerso che, per alcune variabili, i controlli *signal* impostati sono risultati un po' stringenti, in quanto erano stati considerati come 'particolari' valori, mentre si sono verificati in numero cospicuo di casi. Pertanto sono state ridefinite alcune soglie di controllo per l'effettiva individuazione degli eventi rari. In particolare

- è stato abbassato da 13 a 12 anni il limite di età per il conseguimento del titolo di licenza media;
- è stato alzato da 4 a 5 il numero di gravidanze avute prima di aspettare il bambino presente nel campione;
- è stato alzato da 6 a 7 il limite sul numero di stanze di cui è composta l'abitazione in cui vive la madre del bambino;
- è stato abbassato da 30 a 25 il numero di ore lavorate a settimana dalle madri che lavorano a tempo pieno.

Gli interventi effettuati hanno prodotto un miglioramento nella fase di acquisizione dei dati in quanto hanno reso l'intervista più scorrevole sia per le rispondenti che per gli operatori.

## 6. Considerazioni conclusive

L'esperienza maturata nell'ambito dell'indagine sulle nascite ha permesso di apprezzare sul campo i vantaggi della soluzione *in-house* che si è adottata.

In primo luogo, i tempi necessari per l'avvio dell'indagine dal momento in cui si è resa disponibile la ditta sono stati estremamente ridotti - circa un mese - e ciò rappresenta un risultato notevole se si considera la natura nuova dell'indagine e la sua complessità. Tale riduzione dei tempi è stata possibile in quanto tutte le fasi di progettazione, sviluppo e test del questionario elettronico erano state precedentemente realizzate internamente all'Istat: il questionario elettronico adottato per l'indagine a regime è stato, infatti, quello sviluppato per il pre-test integrato da alcuni interventi migliorativi. Questa opportunità consente di ottimizzare la schedulazione dei tempi necessari per l'avvio dell'indagine; tali fasi, infatti, possono essere espletate nel periodo di attesa che intercorre tra la richiesta di gara e la disponibilità della ditta (circa sei mesi).

Un altro vantaggio apprezzabile, qualora si debba replicare l'indagine, deriva dalla possibilità di riutilizzo del questionario elettronico e del *software*, senza dover effettuare ex-novo le fasi di progettazione, sviluppo e test dello strumento.

Notevoli anche gli aspetti positivi derivati dalla stretta collaborazione tra il responsabile dell'indagine e gli esperti della tecnica di rilevazione. È stato possibile, infatti, procedere per affinamenti successivi dello strumento rendendolo il più flessibile e accurato possibile. L'elevata qualità del prodotto sviluppato e la sua assoluta conformità alle specifiche di progettazione sono garanzia di un alto livello di qualità delle informazioni rilevate.

Il poter utilizzare al meglio la potenzialità della tecnica di rilevazione CATI inerente la prevenzione degli errori di rilevazione consente di minimizzare la fase di controllo e correzione, comportando una notevole riduzione dei tempi necessari al rilascio delle informazioni.

Più in generale, si sottolinea inoltre che i moduli sviluppati per l'Indagine Campionaria sulle Nascite, a meno del questionario elettronico, sono stati progettati tenendo in considerazione sia le esigenze specifiche della stessa, che la possibilità di essere riutilizzati per ulteriori indagini. La procedura può quindi costituire uno standard per le rilevazioni CATI da realizzarsi secondo la strategia *in-house*, in quanto costituita da moduli facilmente personalizzabili a seguito della modifica di appositi parametri.

La soluzione implica, tuttavia, degli oneri aggiuntivi. La responsabilità dell'attività di manutenzione di tutti i moduli sviluppati, infatti, impone la disponibilità di risorse dell'Istat per apportare aggiornamenti o miglioramenti all'applicazione ogni qualvolta ciò si renda necessario.

La soluzione implica, tuttavia, degli oneri aggiuntivi. La responsabilità dell'attività di manutenzione di tutti i moduli sviluppati, infatti, impone la disponibilità di risorse dell'Istat per apportare aggiornamenti o miglioramenti all'applicazione ogni qualvolta ciò si renda necessario.

A ciò si aggiunga che l'attività di formazione del personale esterno addetto all'indagine - soprattutto del responsabile del progetto e dei supervisor di sala - risulta anch'essa più impegnativa, in quanto occorre istruire risorse che dovranno coordinare e monitorare tutta la fase delle interviste, le quali, pur avendo dimestichezza con le rilevazioni CATI, non conoscono nello specifico lo strumento da utilizzare.

Infine, dal punto di vista strategico, qualora questa soluzione venisse adottata più diffusamente, si dovrebbe continuare a far riferimento alla struttura centralizzata per l'assistenza nelle fasi di progettazione e messa in produzione delle singole applicazioni e si dovrebbe prevedere la formazione, presso i settori responsabili delle indagini, di figure professionali esperte dei pacchetti *software* utilizzati. Nell'ambito di questa soluzione organizzativa si dovrebbe, inoltre, prevedere la collaborazione con il settore informatico per avvalersi di esperti di rete e sistemisti Windows, soprattutto durante la fase di avvio dell'indagine.

## **Bibliografia**

ACTRv3 User Guide (1998), Statistics Canada

Blaise Developer's Guide (1999), Statistics Netherlands

Appel M. and Hellerman E. (1983). "Census Bureau experience with Automated Industry and Occupation Coding". In American Statistical Association, *Proceedings of Section on Survey Research Methods*, pages 32-40.

Chiaro M. (1996) *I Sondaggi Telefonici*, CISU, Roma.

Couper M. P., Baker R. P., Bethlehem J., Clark C. Z. F., Nicholls II W. L., O'Reilly J. M. (Eds.) (1998) *Computer Assisted Survey Information Collection*, Wiley, New York.

Groves R. M., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls II W. L., Waksberg J. (Eds.) (1988) *Telephone Survey Methodology*, Wiley, New York.

Macchia S., Mastroluca S., Reale A., (2001) 'Planning the quality of the automatic coding process for the next Italian General Population Census', Q2001 International Conference on Quality in Official Statistics, Stockholm, May 14-15, 2001

Macchia S. and Murgia M. (2002). "Coding of textual responses: various issues on automated coding and computer assisted coding". JADT 2000: 6es Journées Internationales d'Analyse Statistique des Données Textuelles, St Malo 13-15 marzo 2002

Muratore M.G., Quattrociochi L. and Sabbadini L.L. (2000 -in corso di pubblicazione) "Indagini sociali telefoniche: metodologia ed esperienze della statistica ufficiale", Istat

Saris W. E. (1991) *Computer Assisted Interviewing*, SAGE Publ., Newbury Park.

Wenzowski M.J. (1988), ACTR – A Generalised Automated Coding System. *Survey Methodology*, vol. 14: 299-308.