

# Un'analisi dell'*attrition* e delle sue determinanti nell'indagine panel europeo sulle famiglie

Francesca Gallo, Sara Mastrovita, Isabella Siciliani<sup>1</sup>  
*Istat*

**Riassunto:** Il fenomeno dell'*attrition* rappresenta uno dei maggiori problemi delle indagini panel e la gravità delle conseguenze è legata non tanto al suo ammontare quanto alla sua natura. Il presente lavoro analizza il fenomeno dell'*attrition* relativamente all'indagine Panel Europeo sulle famiglie: si tratta di un'indagine campionaria longitudinale condotta dal 1994 con cadenza annuale nei Paesi dell'Unione europea. Dopo aver quantificato l'*attrition* del campione italiano nel periodo 1994-2000 viene analizzata la natura dell'*attrition* evidenziando le caratteristiche degli individui che abbandonano definitivamente il campione comparativamente a quelle di chi permane nel campione. Attraverso l'uso congiunto di tecniche di segmentazione binaria e di modelli logit vengono inoltre individuate le variabili che influenzano maggiormente l'*attrition*.

**Parole chiave:** Indagini panel, *attrition*, regressione logistica.

## 1. Introduzione

Il Panel Europeo sulle famiglie (*European Community Household Panel – ECHP*) è un'indagine campionaria che, dal 1994, viene effettuata con cadenza annuale in tutti i Paesi dell'Unione Europea<sup>2</sup>.

L'indagine è realizzata dagli Istituti Nazionali di Statistica o da Istituti di ricerca nazionali con il coordinamento dell'Eurostat: questo fa sì che esistano standard comuni in riferimento alla raccolta e predisposizione dei dati (costruzione delle variabili, registrazione e *editing* dei dati, definizione dei pesi, elaborazione di procedure

---

<sup>1</sup> Il presente lavoro appare nell'ambito del progetto "La ricerca di determinanti del rischio mediante analisi di segmentazione di campioni", cofinanziato dal MIUR (coordinatore nazionale Luigi Fabbris). Il lavoro è frutto della collaborazione delle tre autrici: in particolare F. Gallo ha curato la stesura dei paragrafi 1 e 5, S. Mastrovita dei paragrafi 2 e 4 e I. Siciliani dei paragrafi 3 e 6.

<sup>2</sup> L'Austria, la Finlandia e la Svezia sono entrate a far parte del progetto rispettivamente nel 1995, 1996 e 1997. A partire dal 1997, tre Paesi (Germania, Lussemburgo e Regno Unito) hanno fatto confluire i loro Panel nazionali nell'indagine comunitaria.

di imputazione). Sebbene la predisposizione del file di microdati in un formato standardizzato per tutti i paesi europei risulti particolarmente lunga e complessa, i dati prodotti risultano di elevato valore ed utilità. Il Panel Europeo rappresenta, infatti, una fonte unica di informazioni sul reddito familiare e individuale, e sulle condizioni di vita nei Paesi dell'Unione Europea per la comparabilità dei dati prodotti, per i molteplici temi indagati e per la sua struttura longitudinale che consente lo studio dei mutamenti nel tempo.

L'informazione viene raccolta con un questionario armonizzato. Per ogni unità familiare inclusa nella rilevazione sono individuati tutti i membri che la compongono, le caratteristiche demografiche e le relazioni di parentela. Vi è poi un'intervista familiare che spazia su molteplici temi quali la mobilità territoriale, l'abitazione e le spese sostenute per essa, il possesso di beni durevoli, e soprattutto sui diversi indicatori della situazione finanziaria familiare. Inoltre, a tutti i componenti del nucleo familiare con 16 anni o più è rivolta un'intervista personale. I temi principali di quest'ultima sono l'attività lavorativa e il reddito personale, cui si affiancano quesiti sull'istruzione e la formazione, sulle pensioni e le altre entrate personali, sulla soddisfazione nei confronti della situazione economico-finanziaria e della vita lavorativa, sulla salute e sulle relazioni sociali. Altra caratteristica dell'indagine è quella di cercare di affiancare, per i temi più rilevanti, informazioni di tipo quantitativo con indicazioni di carattere soggettivo, includendo quesiti sul grado di soddisfazione o sul giudizio espresso dalle famiglie, che possono contribuire a dettagliare ulteriormente le informazioni quantitative rilevate.

La letteratura è ricca di contributi sui vantaggi analitici delle indagini panel rispetto ad una singola indagine trasversale reiterata nel tempo o un'indagine trasversale retrospettiva (si veda ad esempio Kasprzyk *et al.*, 1989; Trivellato, 1999).

Le indagini panel, tuttavia, soffrono di alcuni problemi che possono talvolta minare seriamente le loro potenzialità: si ricorda, ad esempio la problematica nota nella letteratura con il nome di *condizionamento da panel* - ovvero la possibilità che la risposta fornita in una occasione di indagine sia influenzata da quelle date nelle occasioni precedenti - e il problema dell'*attrition* non casuale o selettivo - ovvero la perdita progressiva e non casuale di unità del campione nel corso del tempo.

In questo lavoro si concentrerà l'attenzione sul fenomeno dell'*attrition* nell'indagine Panel Europeo sulle famiglie avendo quale obiettivo principale l'esame dell'eventuale impatto sulla qualità del campione.

L'*attrition* rappresenta uno dei maggiori problemi delle indagini panel e la gravità delle conseguenze di tale fenomeno è legata non tanto al suo ammontare quanto alla sua natura: se, infatti, un *attrition* casuale produce una riduzione aleatoria della numerosità campionaria, che si traduce esclusivamente in una perdita di efficienza degli stimatori, un *attrition* non casuale o selettivo, colpendo in modo differenziato segmenti diversi della popolazione, può causare serie distorsioni.

Dopo aver descritto brevemente nel par. 2 le caratteristiche principali dell'indagine in oggetto, il lavoro prosegue prendendo in esame 7 occasioni di indagine panel (1994-2000) e valutando, nel par. 3, il fenomeno dell'*attrition* per verificare inoltre, se quest'ultimo si sia concentrato in particolari occasioni di indagine, ovvero sia uniformemente distribuito. Modellando la probabilità di un individuo di rispondere a 7 occasioni consecutive di indagine, il par. 4 intende esaminare la natura dell'*attrition* ed inoltre evidenziare i segmenti della popolazione interessati in modo più acuto dal fenomeno mediante l'utilizzo congiunto di tecniche di segmentazione binaria e di modelli logit. Il par. 5 sposta l'attenzione sull'analisi della dinamica temporale dell'*attrition*: più precisamente, permette di individuare le peculiarità degli individui che permangono<sup>3</sup> nel panel a diverse durate, rispetto a coloro che ne escono.

## 2. Panel Europeo sulle Famiglie: principali caratteristiche

L'universo di riferimento del Panel Europeo è l'insieme delle famiglie residenti in abitazioni private. Le unità di analisi sono le famiglie e tutti gli individui di 16 anni e più che le compongono.

L'indagine è basata su un disegno di campionamento complesso di tipo probabilistico la cui struttura, pur dovendo assicurare comparabilità internazionale, presenta un elevato grado di flessibilità per adeguarsi alle specifiche realtà nazionali.

Per la selezione del campione italiano l'Istat si è avvalso di un disegno di campionamento a due stadi con stratificazione delle unità primarie, cioè i comuni, in base alla regione e alla numerosità della popolazione. All'interno di ciascuno strato è stato selezionato un unico comune campione con probabilità proporzionale alla dimensione demografica. I comuni di maggiore dimensione demografica (23 in tutto) sono stati considerati autorappresentativi, costituenti cioè uno strato a sé. Per ciascun comune è stato poi selezionato un insieme di famiglie con campionamento sistematico direttamente dalle liste anagrafiche.

In Italia, il campione iniziale comprende complessivamente 7.989 famiglie e 24.063 individui. Il numero di famiglie e di individui intervistati alla prima occasione di indagine (1994) è stato pari rispettivamente a 7.115 e 17.729; questo fa sì che l'Italia, assieme alla Spagna, è caratterizzata dal numero più elevato di individui intervistati (Istat, 2001, pag. 51).

Come è noto, la peculiarità della tecnica *panel* consiste nel seguire lo stesso campione nel tempo. A differenza di un'indagine trasversale, in aggiunta all'estrazione del campione iniziale, si richiede pertanto, la definizione e l'applicazione di

---

<sup>3</sup> Per permanenza nel campione si intende, in questo lavoro, la partecipazione attiva all'indagine mediante realizzazione dell'intervista.

regole di inseguimento (*follow-up rules*) che permettano di pervenire ad una definizione dinamica della popolazione.

Ai fini della definizione delle regole di inseguimento è importante la distinzione, all'interno di ciascuna famiglia, tra "componente campione" e "componente non campione". Un componente campione è qualsiasi persona appartenente al campione iniziale che abbia ancora i requisiti per essere seguito dall'indagine, ossia non risulti in istituzione<sup>4</sup> o in un Paese al di fuori dell'Unione Europea. In aggiunta sono considerati componenti campione tutti i figli nati, dopo il 1994, da madre componente campione.

È componente non campione ogni altro individuo che non appartiene all'insieme precedente, ma che comunque risiede nella stessa famiglia dove sono presenti uno o più componenti campione (ad esempio, il coniuge di un componente campione che si sia sposato dopo la prima onda, è componente non campione).

Mentre i componenti campione sono seguiti ed intervistati (a condizione che non si trasferiscano in un'istituzione o in un Paese al di fuori dell'Unione Europea), i componenti non campione sono inclusi nel campione - e quindi intervistati - fino a che vivono in famiglie con componenti campione; non sono più seguiti se si trasferiscono in famiglie senza componenti campione.

Secondo le indicazioni fornite a livello europeo (Eurostat, 1994), le famiglie incluse nel campione al tempo  $t+1$  sono, in primo luogo, quelle intervistate al tempo  $t$ , alle quali si aggiungono le famiglie facenti parte del campione al tempo  $t$ , ma non intervistate per mancato contatto, incapacità a rispondere o rifiuto non definitivo.

Non sono più seguite le famiglie che hanno fornito un rifiuto definitivo al tempo  $t$  e tutte quelle che si sono spostate in un Paese fuori Unione Europea o si sono interamente istituzionalizzate. A partire dalla terza fase dell'ECHP, anche le famiglie che per due occasioni di indagine consecutive non sono state intervistate, indipendentemente dalla motivazione, sono state escluse dal campione.

Oltre alle famiglie già presenti nel campione e inseguite secondo le regole descritte, si aggiungono al campione tutte le famiglie di nuova formazione (cosiddette famiglie *split*), generate dall'uscita di uno o più componenti campione da una famiglia pre-esistente. Tenendo conto di eventi come nascite, morti, matrimoni o divorzi, l'ECHP cerca di cogliere, attraverso le modifiche del campione, i cambiamenti che intervengono nella popolazione.

Le regole di inseguimento permettono di sfruttare appieno le potenzialità dell'indagine panel e di limitare i problemi derivanti dalla perdita di unità nel corso del tempo. Ovviamente, maggiore è la disponibilità delle famiglie e degli individui intervistati durante la prima fase ad essere seguiti nel tempo, migliore è la qualità dell'indagine e quindi dell'informazione prodotta. L'introduzione di famiglie nuove e l'elimi-

---

<sup>4</sup> Per istituzione si intende la casa per anziani, la casa di cura, il convento o monastero, la caserma, il carcere, ecc..

	1994	1995	1996	1997	1998	1999	2000
famiglie intervistate all'onda precedente		7115	7128	7132	6713	6571	6370
+ famiglie non intervistate all'onda precedente ma da ricontattare		482	328	218	345	301	245
<b>= famiglie da onda precedente</b>		<b>7597</b>	<b>7456</b>	<b>7350</b>	<b>7058</b>	<b>6872</b>	<b>6615</b>
+ famiglie nuove	7989	305	291	312	419	264	250
- famiglie non più eleggibili (senza comp camp, istituz, fuori EU)		60	56	69	92	90	94
- famiglie che hanno rifiutato definitivamente	35	185	208	479	415	329	390
- famiglie non intervistate per 2 onde consecutive	357 (*)	201	133	56	98	102	86
<b>= famiglie che passano all'onda successiva</b>	<b>7597</b>	<b>7456</b>	<b>7350</b>	<b>7058</b>	<b>6872</b>	<b>6615</b>	<b>6295</b>

(\*) dal 94 al 95 sono state escluse tutte le famiglie non intervistate per un motivo diverso dal mancato contatto

nazione di quelle non più intervistabili determina, chiaramente, un cambiamento nella numerosità del campione. L'evoluzione del campione scaturisce, quindi, dal meccanismo combinato costituito dalle regole d'inseguimento e dai tassi di risposta. Nella Tavola 1 viene riportata l'evoluzione del campione di famiglie dal 1994 al 2000.

### 3. Il problema dell'*attrition* e una sua quantificazione nell'indagine Panel Europeo sulle famiglie

L'*attrition*, ovvero la perdita progressiva di unità del campione nel corso del tempo, può essere causato da fattori naturali (come ad esempio morti ed emigrazioni) o da comportamenti individuali (mobilità territoriale, irreperibilità, rifiuto a continuare la partecipazione all'indagine), che in varia misura contribuiscono alla diminuzione della dimensione campionaria, spesso in modo non casuale, man mano che la vita del panel si allunga. Tale fenomeno inoltre si accentua poiché, fatte salve le regole di inseguimento, generalmente non sono previste sostituzioni delle unità non rispondenti.

Come già evidenziato precedentemente, l'*attrition* in un'indagine panel costituisce un problema non solo perché determina una riduzione della consistenza campionaria, ma soprattutto perché può produrre una distorsione se il meccanismo che porta alla selezione è non ignorabile (Little e Rubin, 1987), colpendo in maniera dif-

ferenziata segmenti diversi della popolazione. La sola riduzione della numerosità campionaria determinerebbe una diminuzione dell'efficienza delle stime, ma la distorsione che si rischia di introdurre potrebbe essere ben più seria e non sanabile con le usuali tecniche di ponderazione o imputazione.

In termini più generali la letteratura (Kasprzyk *et al.*, 1989) parla del fenomeno della "non risposta panel" (*panel non response*), dove con questo termine vengono indicate 3 differenti tipologie di non risposta specifiche delle indagini panel:

1. la non risposta iniziale, ovvero la presenza di unità che rifiutano o non sono in grado di cooperare alla prima onda, accettando però di rispondere successivamente;
2. l'*attrition*, o caduta definitiva dal panel, che si manifesta con la presenza di unità che, dopo aver risposto ad una o più onde, lasciano il panel senza più farvi ritorno;
3. la non risposta saltuaria, data dalla presenza di unità con un comportamento irregolare di risposta alle varie onde.

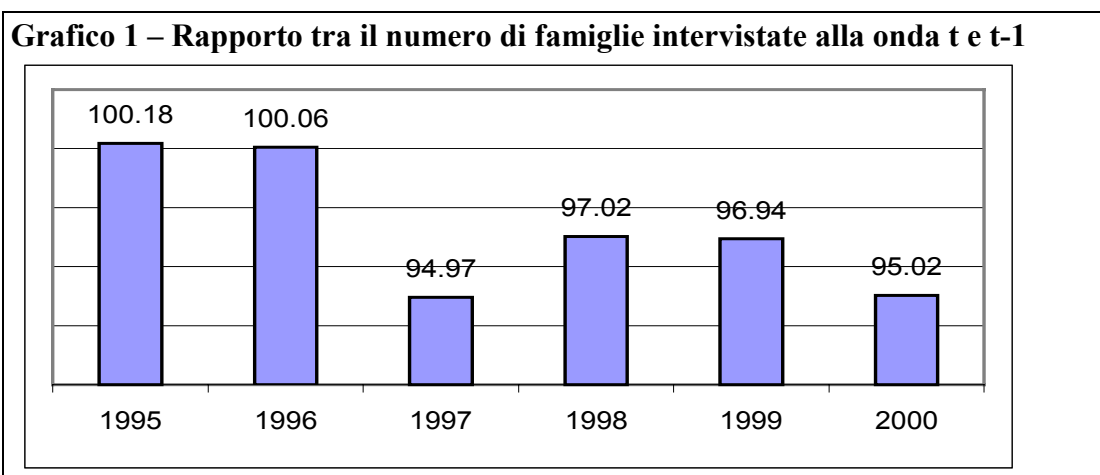
In ciascuna onda del panel il campione si assottiglia per molteplici ragioni quali il mancato contatto, la mancata risposta, l'insuccesso nel seguire il campione e così via. Questa diminuzione campionaria è compensata dall'inclusione di famiglie *split* o da famiglie che, non avendo partecipato al massimo ad un'onda, ritornano successivamente a far parte del campione. Nel Grafico 1 si riporta l'effetto netto di queste componenti. Si può notare come nei primi due anni l'uscita delle famiglie è più che compensata dall'ingresso di famiglie *split* o non intervistate l'anno precedente.

Una caduta pronunciata, invece, si registra in corrispondenza della quarta fase e della settima; il rapporto scende rispettivamente a 94,97% e 95,02%. Si ricorda a tale proposito che alle famiglie contattate all'inizio dell'indagine era stato prospettato un ciclo triennale di interviste; molto probabilmente la caduta di interviste nel 1997 e nel 2000 si verifica a seguito della decisione di Eurostat di proseguire l'indagine ECHP, inizialmente per un ciclo triennale (1997-1999) e poi fino al 2001.

Concentrando l'attenzione esclusivamente sull'*attrition* e sulla non risposta saltuaria e spostando l'analisi sugli individui piuttosto che sulle famiglie<sup>5</sup>, si vuole esaminare fino a che punto gli individui intervistati alla prima onda dell'indagine sono stati seguiti nel corso nel tempo. Si fa presente che ci possono essere individui non intervistati, per rifiuto o per impossibilità nel contatto, all'interno di una famiglia che comunque ha partecipato all'indagine. Quindi per un individuo la mancata partecipazione potrebbe prolungarsi anche per più di 2 onde, se è inserito in una famiglia,

---

<sup>5</sup> Dal momento che, dopo il primo anno di rilevazione, la composizione della famiglia può subire dei cambiamenti a seguito di eventi come trasferimenti, decessi, nascite, matrimoni, si pone il problema di decidere quale nucleo rappresenta la continuazione della famiglia originaria. In letteratura non sembra esserci un'uniformità di consensi nella definizione di famiglia "longitudinale" (Folsom, La Vange, Williams, 1989; Ernst, 1989), vi è peraltro unanimità nel ritenere le persone, piuttosto che le famiglie, le reali unità longitudinali in un panel.



che, secondo le regole di inseguimento, non esce dal campione. Sono stati esclusi dall'analisi tutti gli individui che rispondono per la prima volta a partire dal 1995, siano essi componenti non campione o componenti campione che diventano eleggibili poiché presentano il requisito dell'età.

Una prima analisi ha suddiviso il campione in tre gruppi denominati 'non attritors', 'attritors' (II tipologia di non risposta) e 'rientrati' (III tipologia di non risposta). Al primo gruppo appartengono gli individui, con 16 anni o più nel 1994, che presentano un *pattern* continuativo di intervista nel tempo. Come si evince dalla Tavola 2, circa il 61% è stato intervistato in tutte le onde del panel, dal 1994 al 2000.

Al secondo gruppo appartengono gli individui, con 16 anni o più nel 1994, non più intervistati a partire dall'i-esimo anno di indagine ( $i=1995, \dots, 2000$ ). Gli *at-*

**Tavola 2 – Non attritors, attritors e rientrati nell'indagine Panel europeo – Anni 1995-2000 (valori percentuali)**

Anni	<i>Attritors</i>	Assenti ad almeno un'onda e rientrati successivamente nel campione
1995	4.46	1.77
1996	3.94	1.03
1997	6.91	1.93
1998	5.98	1.32
1999	4.77	0.85
2000	5.81	-
<b>Non attritors:</b>		<b>61.28</b>
<b>Totale individui intervistati nella prima onda:</b>		<b>17.729</b>

*tritors* non superano il 6% ogni anno, fatta eccezione per il 1997, in cui la percentuale è pari a 6,91%.

Nel terzo gruppo sono presenti le persone, con 16 anni o più nel 1994, non intervistate in almeno una onda *i*-esima ( $i=1995, \dots, 2000$ ) del panel e successivamente re-intervistati. Come si evince dalla Tavola 2, la percentuale di individui che rientrano nel corso della vita del panel è, in generale, inferiore al 2%.

La Tavola 2 mette, inoltre, in evidenza che la percentuale di *attritors* non è distribuita uniformemente nei vari anni, ma ha un picco nella quarta rilevazione, in corrispondenza dell'avvio del nuovo ciclo, non previsto inizialmente nel 1994. All'inizio di ogni triennio si riscontra un andamento decrescente dell'*attrition* nel corso del tempo.

Per verificare se l'*attrition* sia di tipo casuale oppure selettivo sono state prese in esame alcune caratteristiche degli individui e delle famiglie del campione, concentrando dapprima l'attenzione su variabili suggerite dai risultati della letteratura su ricerche simili (Fitzgerald, Gottschalk, Moffitt, 1998; Pannenberg, 2000).

La Tavola 3 riporta la composizione percentuale di coloro i quali nel 2000 risultano 'non *attritors*', '*attritors*' e 'rientrati' secondo alcune caratteristiche possedute dall'individuo o dalla famiglia nel 1994 e i valori del test del chi-quadrato e del relativo *p*-value per la verifica dell'ipotesi di indipendenza.

Da una prima analisi delle distribuzioni percentuali si può concludere che, fra le variabili prese in esame, solamente il sesso non abbia alcuna influenza sul fenomeno dell'*attrition*: maschi e femmine ne sono infatti interessati allo stesso modo. Tendono, invece, ad uscire maggiormente dal campione gli anziani (classe 65 e più); i vedovi<sup>6</sup> e i separati; gli individui con titolo di studio basso; i ritirati dal lavoro, i lavoratori autonomi e i sottoccupati; anche chi ha dichiarato di versare in cattive o pessime condizioni di salute o chi aveva relazioni sociali quasi inesistenti è caratterizzato da una maggiore probabilità di uscire dal campione. Per quanto riguarda le caratteristiche familiari, sono propensi ad abbandonare il campione gli individui che si trovano in famiglie la cui fonte principale di reddito è costituita da investimenti e proprietà o dalle pensioni e caratterizzate da un reddito familiare equivalente<sup>7</sup> elevato; i residenti nel Nord-Est; gli individui appartenenti a famiglie di persone sole; chi vive in appartamento e gli affittuari o subaffittuari.

<sup>6</sup> Questo risultato era peraltro prevedibile in quanto il gruppo degli ultrasessantacinquenni è interessato da una maggiore mortalità. Il 50% degli *attritors* di questo sottogruppo risulta infatti deceduto nei 7 anni considerati nell'analisi (i deceduti nel corso dell'indagine panel rappresentano il 14,3% del totale degli *attritors*). Analogo fenomeno interessa il sottogruppo dei vedovi: il 54,4% degli *attritors* di questo sottogruppo muore nel corso dell'indagine panel.

<sup>7</sup> Per tenere conto delle differenze nella dimensione e composizione familiare, il reddito familiare è stato reso equivalente usando la scala OCSE modificata, che assegna un punteggio pari ad 1 al primo individuo, 0,5 ad ogni altro componente di 14 anni o più e 0,3 ai minori di 14 anni.



**Tavola 3 - Composizione percentuale nel 2000 dei 'non attritors', 'attritors' e 'rientrati' secondo alcune caratteristiche possedute dall'individuo nel 1994 - (a)**

Caratteristiche degli individui nel 1994 (Caratteristiche individuali)	Non attritors	Attritors	Rientrati	Valore della statistica $\chi^2$	p-value
<b>Sesso</b>					
Maschio	60,6	32,6	6,8		
Femmina	61,9	31,2	6,9	4,0	0,1300
<b>Età</b>					
16-24	61,0	30,4	8,6		
25-39	62,6	29,7	7,7		
40-54	63,7	29,7	6,6		
55-64	65,1	29,1	5,9		
65 e più	51,6	43,9	4,6	<b>240,3</b>	<b>&lt;0,0001</b>
<b>Titolo di studio</b>					
Titolo universitario	59,8	30,4	9,8		
Diploma scuola media superiore	62,1	30,1	7,9		
Licenza media o avviamento professionale	62,0	30,7	7,4		
Licenza elementare, nessun titolo	60,3	34,6	5,1	<b>72,4</b>	<b>&lt;0,0001</b>
<b>Condizione professionale</b>					
Lavoratore dipendente	62,1	30,3	7,6		
Apprendista, Contratto Formazione Lavoro	61,5	32,6	6,0		
Lavoratore autonomo	55,0	38,5	6,5		
Studente	64,3	27,3	8,4		
Disoccupato, in cerca di prima occupazione	62,4	29,6	8,0		
Ritirato dal lavoro	56,3	38,8	4,8		
Casalinga	64,3	29,0	6,7		
Altro	42,3	42,3	15,5	<b>157,6</b>	<b>&lt;0,0001</b>
<b>Stato occupazionale</b>					
Un lavoro	61,5	31,4	7,1		
Due o più lavori	68,9	22,7	8,4		
Lavora meno di 15 ore a settimana	55,5	38,5	6,0		
Non lavora	61,2	32,2	6,6	<b>13,4</b>	<b>0,0366</b>
<b>Frequenza di conversazione con i vicini</b>					
Sempre	64,0	29,3	6,7		
Qualche volta	59,5	33,7	6,8		
Raramente o mai	54,5	37,8	7,7	<b>86,9</b>	<b>&lt;0,0001</b>
<b>Stato di salute</b>					
Ottima o buona	62,4	30,0	7,6		
Normale	62,1	31,8	6,1		
Cattiva o pessima	54,1	41,0	4,9	<b>112,8</b>	<b>&lt;0,0001</b>

(a) - In grassetto sono evidenziati i valori del test che portano a rifiutare, ad un livello di significatività del 5%, l'ipotesi di indipendenza.

**Tavola 3 continua - Composizione percentuale nel 2000 dei 'non attritors', 'attritors' e 'rientrati' secondo alcune caratteristiche possedute dall'individuo nel 1994 - (a)**

Caratteristiche degli individui nel 1994 (Caratteristiche familiari)	Non attritors	Attritors	Rientrati	Valore della statistica $\chi^2$	p-value
<b>Ripartizione geografica</b>					
Nord - Ovest	51,3	40,9	7,8		
Nord - Est	49,9	43,3	6,9		
Centro	58,3	34,3	7,4		
Sud	72,3	21,5	6,2	<b>780,9</b>	<b>&lt;0,0001</b>
<b>Dimensione familiare</b>					
Un componente	52,1	40,5	7,5		
Due componenti	56,9	37,1	6,0		
Tre componenti	58,9	33,5	7,6		
Quattro componenti	63,2	30,0	6,9		
Cinque o più componenti	67,2	26,4	6,5	<b>158,1</b>	<b>&lt;0,0001</b>
<b>Tipologia familiare</b>					
Single di età < 30 anni	50,8	41,5	7,7		
Single di età 30-64	59,5	30,3	10,2		
Single di età $\geq$ 65 anni	46,6	48,0	5,4		
Coppia senza figli	57,9	36,6	5,5		
Coppia con figli	64,0	29,2	6,8		
Monogenitore	54,4	37,2	8,4		
Altro	61,0	32,9	6,1	<b>165,0</b>	<b>&lt;0,0001</b>
<b>Fonte di reddito principale</b>					
Salari e stipendi	63,7	29,3	7,0		
Redditi da lavoro indipendente o autonomo	60,9	32,9	6,2		
Pensioni	57,4	36,1	6,5		
Assegni disoccupazione, cassa integrazione guadagni e altre indennità	65,6	26,9	7,5		
Investimenti/proprietà e altro	56,2	37,8	6,0	<b>73,1</b>	<b>&lt;0,0001</b>
<b>Classe di reddito familiare equivalente</b>					
Fino al 1° quartile	68,2	25,6	6,2		
Fino al 2° quartile	62,3	31,2	6,5		
Fino al 3° quartile	56,7	36,2	7,1		
Oltre il 3° quartile	57,5	35,6	6,9	<b>164,3</b>	<b>&lt;0,0001</b>
<b>Tipo di abitazione</b>					
Villa, villino e casa rurale	62,0	31,6	6,4		
Appartamento	60,5	32,3	7,2		
Altro	66,7	27,4	5,9	<b>16,2</b>	<b>0,0027</b>
<b>Titolo di godimento dell'abitazione</b>					
Proprietà o usufrutto	62,6	30,8	6,6		
Affitto o subaffitto	56,5	35,2	8,3		
Altro titolo	59,9	34,7	5,5	<b>49,8</b>	<b>&lt;0,0001</b>
<b>Totale</b>	<b>61,3</b>	<b>31,9</b>	<b>6,9</b>		

#### 4. Individuazione delle principali determinanti dell'*attrition*

Un'analisi più accurata dell'*attrition* nei diversi segmenti della popolazione non può non tener conto dell'esistenza di interrelazioni fra le potenziali variabili esplicative del fenomeno.

Nel compiere questa analisi si è preferito escludere i 'rientrati' che, partecipando all'indagine in maniera discontinua, non sembrano assimilabili al gruppo degli 'attritors', i quali contrariamente manifestano un comportamento continuativo di non risposta a partire da una certa occasione d'indagine. Si fa notare, inoltre, che la consistenza numerica dei 'rientrati' è piuttosto contenuta (paragrafo 3).

È stato quindi specificato un modello logit binomiale che stima la probabilità di permanenza di un individuo nel panel ripetutamente nelle 7 onde considerate – come contrasto la probabilità dell'individuo di lasciare definitivamente il campione in una qualsiasi delle 7 onde – condizionatamente ad un insieme di variabili socio-economiche riferite alla prima onda del panel.

La specificazione del modello è stata preceduta da un'analisi dei dati attraverso le tecniche di segmentazione binaria, utilizzate in questo contesto per selezionare le variabili esplicative da includere nel modello. Le variabili che influenzano maggiormente la permanenza di un individuo nel panel possono essere raggruppate nelle seguenti 4 tipologie:

- la ripartizione geografica di residenza della famiglia;
- l'età e lo stato di salute degli individui intervistati;
- la classe, la fonte di reddito e la tipologia di abitazione della famiglia;
- il "grado di socievolezza" dell'intervistato, misurato dalla variabile "frequenza con cui l'intervistato parla con persone non della famiglia".

Se  $Y$  indica una variabile dicotomica che assume il valore 1 se l'individuo ha partecipato a tutte le 7 onde di indagine e 0 altrimenti, il modello utilizzato è il seguente:

$$P(Y=1|X_{1994}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_q X_q)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_q X_q)} \quad (1)$$

La Tavola 4 riporta i risultati del modello stimato a partire dal set di variabili  $X_1, \dots, X_q$  suggerite dall'analisi esplorativa.

Per un individuo che non risiede al Sud, la probabilità di permanere nel campione risulta inferiore a quella di chi risiede in questa ripartizione; in particolare risulta confermata la maggiore propensione all'*attrition* per i residenti nel Nord-Est.

Gli individui con meno di 65 anni hanno una maggiore probabilità di rimanere nel campione, probabilità che risulta più accentuata per le classi intermedie. Per -

**Tavola 4 - Risultati di un modello di regressione logistica sulla probabilità di permanenza nel panel fino all'anno 2000**

Variabili individuali e familiari	Probabilità di permanenza nel panel fino all'anno 2000 condizionata alle variabili dell'anno 1994			
	Stima param.	Valore del $\chi^2$	p-value	Odds ratio
<b>Intercetta</b>	-0,0804	0,2	0,6486	
<b>Caratteristiche individuali</b>				
<b>età</b>		89,0	<0,0001	
16-24	<b>0,2200</b>	8,6	0,0034	1,2461
25-39	<b>0,4091</b>	36,1	<0,0001	1,5055
40-54	<b>0,4119</b>	38,1	<0,0001	1,5097
55-64	<b>0,5715</b>	73,2	<0,0001	1,7709
65 e più	0,0000	.	.	1,0000
<b>frequenza di conversazione con i vicini</b>		70,6	<0,0001	
Sempre	<b>0,4095</b>	58,2	<0,0001	1,5061
Qualche volta	<b>0,1892</b>	11,3	0,0008	1,2082
Raramente o mai	0,0000	.	.	1,0000
<b>stato di salute</b>		43,9	<0,0001	
Ottima o buona	<b>0,4035</b>	42,1	<0,0001	1,4970
Normale	<b>0,3540</b>	33,5	<0,0001	1,4248
Cattiva o pessima	0,0000	.	.	1,0000
<b>Caratteristiche familiari</b>				
<b>ripartizione geografica</b>		442,9	<0,0001	
Nord-Ovest	<b>-0,9279</b>	349,2	<0,0001	0,3954
Nord-Est	<b>-1,0036</b>	399,3	<0,0001	0,3666
Centro	<b>-0,5902</b>	132,5	<0,0001	0,5542
Sud	0,0000	.	.	1,0000
<b>fonte di reddito principale</b>		17,6	0,0015	
Salari e stipendi	<b>0,3404</b>	4,6	0,0317	1,4054
Redditi da lavoro indipendente o autonomo	0,1767	1,2	0,2731	1,1932
Pensioni	0,2132	1,8	0,1850	1,2377
Ass. disoccupazione, cassa integr. e altre indennità	0,2923	1,5	0,2137	1,3395
Investimenti/proprietà e altro	0,0000	.	.	1,0000
<b>classe di reddito familiare equivalente</b>		28,0	<0,0001	
Fino al 1° quartile	<b>0,1727</b>	10,0	0,0015	1,1885
Fino al 2° quartile	0,0645	1,5	0,2142	1,0666
Fino al 3° quartile	<b>-0,1005</b>	4,1	0,0431	0,9043
Oltre il 3° quartile	0,0000	.	.	1,0000
<i>Devianza (b)</i>	<i>1708 (18)</i>			
<i>p-value</i>	<i>&lt;0,0001</i>			
<i>Numerosità campionaria</i>	<i>16513 (c)</i>			

(a) I parametri significativamente diversi da zero, a livello di significatività del 5%, sono in grassetto - (b) La devianza è pari a  $-2\log(\text{verosimiglianza del modello con la sola intercetta}/\text{verosimiglianza del modello stimato})$ . Tra parentesi sono riportati i gradi di libertà - (c) Dall'analisi sono stati esclusi gli individui usciti in alcune onde del panel, ma poi rientrati, in quanto non è stato ritenuto corretto assimilare il loro comportamento a quello degli attritors.

chi vive in famiglie in cui la fonte prevalente di reddito è costituita da salari e stipendi, la propensione a rimanere nell'indagine aumenta rispetto a chi vive prevalentemente con redditi da proprietà o investimenti finanziari. Si nota anche la maggiore permanenza nel campione di individui caratterizzati da livelli bassi e medio-bassi di reddito familiare equivalente, mentre chi si colloca in una classe medio-alta sembra uscire dal campione con più facilità rispetto a chi ha livelli di reddito più alti. Anche chi gode di buone condizioni di salute e chi mantiene buone relazioni sociali con i vicini di casa risulta più propenso a permanere continuativamente nel campione durante le 7 onde di indagine.

Il valore della devianza (1.708), presa come statistica che permette di giudicare la bontà del modello stimato, indica che il fenomeno dell'*attrition*, che ha interessato complessivamente nelle 7 onde il campione dell'indagine panel, non è di tipo casuale.

## 5. Un'analisi della dinamica dell'*attrition*

Il modello (1) esposto nel par. 4 descrive il processo di selezione avvenuto complessivamente dal 1994 al 2000.

Per approfondire la dinamica del processo che ha portato nel tempo alla perdita delle unità appartenenti al campione iniziale sono stati analizzati sei insiemi di dati: il primo, relativo all'anno 1995, il secondo al 1996 e così via fino al sesto relativo all'anno 2000. L'insieme relativo al generico anno  $t$  ( $t=1995, 1996, 1997, 1998, 1999, 2000$ ) ha numerosità pari al numero di individui che hanno sempre partecipato attivamente all'indagine fino all'anno  $t-1$  e contiene le seguenti variabili:

- ◆  $Y_t$ , variabile dicotomica che assume il valore 1 (rispondenti) se l'individuo intervistato fino all'anno  $t-1$  è intervistato anche all'anno  $t$  e 0 (non rispondenti) altrimenti;
- ◆ le principali caratteristiche socio-economiche possedute dalla famiglia e dagli individui nell'anno  $t-1$ .

Le Tavole 5 e 6 riportano i risultati di una prima analisi descrittiva dei sei insiemi di dati. Più precisamente, nella Tavola 5 sono evidenziate in grassetto le variabili per le quali viene rifiutata, ad un livello di significatività del 5%, l'ipotesi di indipendenza dalla variabile  $Y_t$  ( $t=1995, 1996, 1997, 1998, 1999, 2000$ ). La Tavola 6, invece, presenta la distribuzione condizionata della variabile  $Y_t$  ( $t=1995, 1996, 1997, 1998, 1999, 2000$ ) alle caratteristiche possedute dalla famiglia o dall'individuo all'anno  $t-1$ .

Tra le variabili individuali, l'*attrition* sembra essere indipendente dal sesso, come già evidenziato nell'analisi relativa a tutto il periodo delle 7 occasioni di

**Tavola 5 - Risultati del test sull'ipotesi di indipendenza tra la variabile risposta/non risposta e le caratteristiche possedute dagli individui nell'anno precedente all'indagine (a)**

Caratteristiche degli individui nell'anno precedente all'indagine	gradi di libertà	1995		1996		1997		1998		1999		2000	
		valore del $\chi^2$	p-value	valore del $\chi^2$	p-value	valore del $\chi^2$	p-value	valore del $\chi^2$	p-value	valore del $\chi^2$	p-value	valore del $\chi^2$	p-value
<b>Caratteristiche individuali</b>													
sexo	1	1,6	0,2072	0,2	0,6636	0,7	0,3961	2,5	0,1150	0,5	0,4854	0,2	0,6947
età	4	54,3	<0,0001	36,2	<0,0001	8,6	0,0729	20,1	0,0005	35,2	<0,0001	40,4	<0,0001
stato civile	4	52,3	<0,0001	49,0	<0,0001	25,4	<0,0001	26,1	<0,0001	12,1	0,0164	16,5	0,0024
titolo di studio	3	3,9	0,2766	7,2	0,0644	0,6	0,8850	11,1	0,0111	6,4	0,0952	24,7	<0,0001
condizione professionale	8	14,8	0,0633	14,8	0,0954	25,7	0,0023	7,0	0,6344	27,7	0,0011	13,2	0,1536
stato occupazionale	3	2,7	0,4331	4,0	0,2622	7,2	0,0662	2,1	0,5448	2,1	0,5606	0,7	0,8717
frequenza di conversazione con i vicini	2	71,4	<0,0001	59,7	<0,0001	9,4	0,0092	60,2	<0,0001	26,3	<0,0001	25,3	<0,0001
partecipazione ad associazioni culturali, circoli, etc.	1	0,6	0,4405	2,8	0,0936	1,2	0,2690	2,0	0,1596	2,1	0,1472	0,1	0,7959
frequenza di incontro con amici	2	16,1	0,0003	30,8	<0,0001	0,5	0,7925	17,5	0,0002	10,8	0,0046	10,9	0,0043
frequenza di conversazione con persone al di fuori della famiglia	1	12,9	0,0003	5,0	0,0250	2,8	0,0922	0,7	0,3932	3,4	0,0638	9,4	0,0021
stato di salute	2	40,0	<0,0001	67,5	<0,0001	3,5	0,1769	25,5	<0,0001	23,9	<0,0001	43,8	<0,0001
presenza di malattie croniche/disabilità che ostacolano le attività quotidiane	1	5,0	0,0249	0,2	0,6338	0,0	0,8847	0,0	0,9955	0,3	0,6101	6,7	0,0096
interruzione del lavoro per malattia/incidente	1	18,7	<0,0001	42,6	<0,0001	13,8	0,0002	17,3	<0,0001	34,1	<0,0001	5,3	0,0212
<b>Caratteristiche familiari</b>													
ripartizione geografica	3	160,1	<0,0001	129,4	<0,0001	229,0	<0,0001	187,0	<0,0001	105,3	<0,0001	153,4	<0,0001
dimensione familiare	4	67,1	<0,0001	45,3	<0,0001	25,2	<0,0001	16,5	0,002	36,8	<0,0001	52,4	<0,0001
tipologia familiare	6	71,3	<0,0001	34,6	<0,0001	24,6	4E-04	18,7	0,005	20,1	0,003	33,9	<0,0001
fonte di reddito principale	4	0,7	0,9467	14,3	0,0064	15,0	0,005	1,4	0,837	14,7	0,005	12,3	0,015
classe di reddito familiare equivalente	3	6,2	0,1038	7,9	0,0484	66,6	<0,0001	25,0	<0,0001	14,0	0,003	5,8	0,12
condizione di povertà	1	0,6	0,4582	1,5	0,2247	32,2	<0,0001	5,2	0,022	9,3	0,002	4,1	0,042
tipo di abitazione	2	23,0	<0,0001	9,1	0,0105	0,8	0,685	0,9	0,63	13,9	0,0010	3,2	0,204
titolo di godimento dell'abitazione	2	55,2	<0,0001	29,3	<0,0001	0,1	0,96	14,7	6E-04	5,7	0,058	1,9	0,378

(a) - In grassetto sono evidenziati i valori del test che portano a rifiutare, ad un livello di significatività del 5%, l'ipotesi di indipendenza.

indagine (Tavola 3), dallo stato occupazione e dalla partecipazione ad associazioni culturali, circoli ecc.

Influiscono invece sull'*attrition*, in tutti gli anni considerati, lo stato civile, la frequenza con cui l'individuo parla con i vicini e la frequenza con cui interrompe il lavoro per malattia, incidente o disturbi psichici. Altre variabili, sempre relative alle condizioni di salute, esplicano la loro influenza solo in alcuni degli anni presi in esame; lo stesso può dirsi per l'età dell'individuo.

Tra le caratteristiche della famiglia di appartenenza degli individui, la ripartizione geografica, la dimensione della famiglia e la tipologia di quest'ultima risultano avere una influenza costante nella dinamica del processo di perdita di unità campionarie; inoltre, giocano un ruolo importante le caratteristiche legate al reddito familiare (fonte di reddito prevalente, classe di reddito familiare equivalente e condizione di povertà<sup>8</sup>). Passando a considerare la distribuzione condizionata della variabile  $Y_t$  ( $t=1995, 1996, 1997, 1998, 1999, 2000$ ) alle caratteristiche possedute dalla famiglia o dall'individuo all'anno  $t-1$ , la Tavola 6 evidenzia come tra i rispondenti non si rilevano differenze consistenti tra maschi e femmine.

Anche in questa analisi dinamica, si conferma in ogni anno la minore tendenza alla risposta della classe di età più elevata a cui si affianca, a partire dal 1997, il gruppo dei più giovani (16-24 anni).

I possessori di un titolo di studio elevato compaiono in quasi tutti gli anni tra i non rispondenti, ad eccezione dell'ultimo anno, in cui si verifica che è superiore la percentuale dei non rispondenti con licenza elementare o nessun titolo.

Per la condizione professionale, se nel 1995, la quota dei non rispondenti si presenta più marcata tra chi si trova in "altro" (es. militari, economicamente inattivi in condizione diversa da quelle già indicate), nelle successive rilevazioni non si riscontrano spiccate differenziazioni tra le varie tipologie professionali. Chi si è dichiarato in condizione di sottoccupazione nell'anno precedente l'indagine (con un lavoro inferiore alle 15 ore settimanali) è meno propenso a rispondere l'anno successivo: questo fenomeno si rileva nel 1995, 1996 e nel 2000, mentre nei rimanenti anni tende a non partecipare attivamente chi, l'anno precedente, ha dichiarato di svolgere un solo lavoro.

Passando a considerare gli indicatori che denotano l'apertura degli individui a relazioni sociali (partecipazioni ad associazioni, conversazioni o incontri con persone esterne alla famiglia), in generale coloro che frequentano o conversano raramente con persone non della famiglia, si caratterizzano in tutti gli anni da una maggiore percentuale di mancata risposta.

---

<sup>8</sup> La condizione di povertà è stata definita utilizzando il reddito familiare reso equivalente sulla base della scala OCSE modificata e adottando quale soglia di povertà il 60% del reddito mediano equivalente.

**Tavola 6 - Composizione percentuale dei rispondenti secondo alcune caratteristiche possedute dagli individui nell'anno precedente all'indagine**

Caratteristiche degli individui nell'anno precedente all'indagine	Anno di indagine					
	1995	1996	1997	1998	1999	2000
<b>Caratteristiche individuali</b>						
<b>Sesso</b>						
Femmina	94,0	94,7	90,3	91,2	92,4	91,4
Maschio	93,6	94,8	89,9	90,5	92,1	91,2
<b>Età</b>						
16-24	94,5	95,5	89,4	90,0	91,9	91,5
25-39	94,0	94,5	91,0	91,1	92,4	92,0
40-54	94,5	95,6	89,9	91,4	93,3	92,0
55-64	94,7	95,0	90,2	92,6	93,7	92,4
65 e più	90,7	92,4	89,1	88,9	89,1	87,2
<b>Titolo di studio</b>						
Titolo universitario	92,5	93,4	89,6	88,5	91,1	94,3
Diploma scuola media superiore	93,8	95,3	90,3	90,1	92,8	92,4
Licenza media o avviamento professionale	94,1	94,6	89,9	91,4	92,7	91,1
Licenza elementare, nessun titolo	93,7	94,6	90,2	91,4	91,7	89,9
<b>Condizione professionale</b>						
Lavoratore dipendente	93,8	95,1	89,4	90,7	92,5	91,6
Apprendista, Contr. Formaz. Lavoro	93,9	96,0	86,4	92,1	92,9	90,1
Lavoratore autonomo	93,8	94,5	90,4	90,6	90,4	91,0
Collaboratore familiare	.	95,3	91,9	94,7	89,2	91,5
Studente	95,3	95,6	91,6	91,0	91,3	94,4
Disoccupato, in cerca di prima occupazione	94,5	95,7	91,8	91,0	93,8	92,0
Ritirato dal lavoro	93,1	93,8	88,8	90,6	91,3	90,1
Casalinga	93,6	94,4	91,0	91,5	94,0	91,9
Altro	88,7	95,9	89,5	88,8	90,8	90,5
<b>Stato occupazionale</b>						
Un lavoro	93,8	95,0	89,7	90,7	91,9	91,4
Due o più lavori	96,6	93,2	90,1	94,2	92,1	93,9
Lavora meno di 15 ore a settimana	92,4	92,7	98,5	94,7	94,6	90,7
Non lavora	93,8	94,6	90,3	90,9	92,5	91,3
<b>Partecipazione ad associazioni culturali, circoli, etc.</b>						
Si	94,1	95,4	89,5	91,6	93,0	91,2
No	93,7	94,6	90,2	90,7	92,1	91,4
<b>Frequenza di conversazione con i vicini</b>						
Sempre	95,1	95,8	90,7	92,5	93,1	92,3
Qualche volta	92,8	93,9	89,4	89,3	92,0	91,0
Raramente o mai	90,9	91,9	88,8	87,2	89,0	87,7
<b>Frequenza di incontro con amici</b>						
Spesso	94,4	95,7	90,1	91,8	92,0	91,5
Qualche volta	93,7	94,6	90,2	90,2	92,9	91,6
Raramente o mai	92,3	92,8	89,5	88,2	89,8	87,9



**Tavola 6 continua - Composizione percentuale dei rispondenti secondo alcune caratteristiche possedute dagli individui nell'anno precedente all'indagine**

Caratteristiche degli individui nell'anno precedente all'indagine	Anno di indagine					
	1995	1996	1997	1998	1999	2000
<b>Caratteristiche individuali (continua)</b>						
<b>Frequenza di conversazione con persone al di fuori della famiglia</b>						
Si	93,7	93,8	90,8	88,9	91,2	90,3
No	90,9	91,5	87,2	86,9	86,9	81,5
<b>Stato di salute</b>						
Ottima o buona	94,4	95,1	90,0	91,7	92,7	91,9
Normale	93,7	95,5	90,5	90,4	92,7	92,1
Cattiva o pessima	90,8	90,8	89,0	87,8	89,2	86,8
<b>Interruzione del lavoro per malattia/incidente</b>						
Si	90,2	88,8	85,4	85,7	86,0	88,7
No	94,0	94,9	90,2	91,1	92,6	91,5
<b>Ricovero per malattia in una struttura sanitaria</b>						
Si	91,7	90,8	88,2	88,8	89,7	87,6
No	94,0	95,1	90,2	91,1	92,6	91,7
<b>Caratteristiche familiari</b>						
<b>Ripartizione geografica</b>						
Nord-ovest	91,2	91,7	85,6	88,7	90,1	89,3
Nord-Est	90,7	92,8	85,8	89,1	90,3	85,5
Centro	95,4	96,1	90,1	86,2	89,6	91,1
Sud	95,7	96,3	93,7	94,3	94,8	94,2
<b>Dimensione familiare</b>						
Un componente	89,1	91,9	87,7	89,2	90,0	88,2
Due componenti	92,5	93,3	90,1	89,7	91,4	90,5
Tre componenti	94,1	94,5	88,9	91,0	90,9	89,8
Quattro componenti	94,0	95,9	90,3	90,8	93,7	93,9
Cinque o più componenti	95,4	95,3	91,9	92,5	93,8	91,6
<b>Tipologia familiare</b>						
Single <30 anni	87,7	89,9	82,0	80,8	91,0	89,2
Single 30-64 anni	90,3	92,8	89,2	90,5	90,2	87,3
Single di età >= 65 anni	88,4	91,5	87,6	89,2	89,8	88,2
Coppia senza figli	92,6	93,9	90,9	90,0	90,9	90,8
Coppia con figli	94,9	95,4	90,3	91,5	93,0	92,4
Monogenitore	93,5	93,5	86,8	88,8	90,6	91,1
Altro	93,4	94,1	91,0	90,9	92,8	89,0
<b>Fonte di reddito principale</b>						
Salari e stipendi	94,1	95,3	89,9	90,7	93,0	92,3
Redditi da lavoro indipendente o autonomo	94,2	94,5	91,4	91,2	90,9	90,8
Pensioni	94,1	94,1	89,2	90,9	92,2	90,3
Assegni disoccupazione, cassa integrazione guadagni e altre indennità	93,0	90,5	95,8	91,9	88,9	92,1
Investimenti/proprietà e altro	95,0	94,4	91,8	89,9	91,8	91,4

**Tavola 6 continua - Composizione percentuale dei rispondenti secondo alcune caratteristiche possedute dagli individui nell'anno precedente all'indagine**

Caratteristiche degli individui nell'anno precedente all'indagine	Anno di indagine					
	1995	1996	1997	1998	1999	2000
<b>Caratteristiche familiari (continua)</b>						
<b>Classe di reddito familiare equivalente</b>						
Fino al 1° quartile	94,2	95,3	92,6	91,8	93,7	91,9
Fino al 2° quartile	94,3	95,1	91,0	92,0	92,2	92,1
Fino al 3° quartile	94,1	94,3	87,6	90,7	92,1	91,0
Oltre il 3° quartile	93,2	94,2	88,9	88,9	91,3	90,5
<b>Condizione di povertà</b>						
Non povero	93,9	94,7	89,4	90,6	92,0	91,2
Povero	94,2	95,2	92,8	92,0	93,8	92,5
<b>Tipo di abitazione</b>						
Villa, villino e casa rurale	95,0	97,4	87,6	92,0	96,1	88,8
Appartamento	93,1	92,6	85,0	91,4	87,1	82,7
Altro	93,9	85,3	86,8	85,7	93,8	90,9
<b>Titolo di godimento dell'abitazione</b>						
Proprietà o usufrutto	94,4	95,1	90,0	91,3	92,6	91,2
Affitto o subaffitto	91,0	92,7	90,2	88,8	91,3	91,4
Altro titolo	94,8	95,9	90,2	91,1	90,7	92,8
<b>Totale</b>	<b>93,8</b>	<b>94,7</b>	<b>90,1</b>	<b>90,9</b>	<b>92,3</b>	<b>91,3</b>

Anche lo stato di salute incide sul comportamento di risposta nei vari anni: chi valuta come pessime le proprie condizioni di salute o chi ha oggettivamente problemi di salute perché manifesta la presenza di malattie croniche o disabilità che ostacolano le attività quotidiane, ha interrotto il lavoro per malattia, incidenti, disturbi psichici, ha subito un ricovero in una struttura sanitaria, tende tutti gli anni a una minore partecipazione all'indagine.

Analizzando le caratteristiche della famiglia di appartenenza nell'indagine dell'anno precedente, gli individui che risiedono nelle regioni meridionali sono contraddistinti da una maggior partecipazione in tutti gli anni.

Le persone sole sono caratterizzate da una minore presenza attiva, in particolare gli anziani e i giovani single, mentre in generale all'aumentare della dimensione familiare le percentuali di partecipazione crescono, con una tendenza più accentuata per le coppie con figli.

Se si considerano le variabili familiari di tipo reddituale, mentre l'ammontare di reddito a disposizione incide sul comportamento di mancata risposta (in generale maggiore è l'ammontare, minore la tendenza alla risposta), l'insieme dei non rispondenti non sembra essere caratterizzato dalla stessa fonte di reddito familiare prevalente nel corso del tempo.

Per quanto riguarda le variabili inerenti all'abitazione, le persone che vivono in ville, villini o case rurali o chi ha dichiarato di essere proprietario o usufruttuario dell'abitazione in cui vive partecipano di più all'indagine.

I risultati dell'analisi condotta sui sei insiemi di dati (intervistati e non nel 95 con le caratteristiche del 94, intervistati e non nel 96 con le caratteristiche del 95, intervistati e non nel 97 con le caratteristiche del 96, etc.) attraverso le tecniche di segmentazione binaria evidenziano che le variabili esplicative della permanenza in ciascuna occasione di indagine sono sostanzialmente quelle che spiegano il fenomeno dell'*attrition* cumulato.

Si è quindi proceduto alla stima dei seguenti modelli:

$$P(Y_t=1|Y_{t-1}=1, \dots, Y_{t-i}=1, X_{t-1}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_q X_q)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_q X_q)} \quad (t=1995, \dots, 2000)$$

(i=1, ..., 6)  
(t-i) ≥ 1994

dove le variabili esplicative sono quelle considerate nel modello (1). I risultati sono riportati nella Tavola 7.

Nell'esaminare la dinamica del processo di perdita delle unità statistiche nel corso delle diverse occasioni di indagine, si nota che le variabili che esercitano un impatto sulla probabilità di risposta in tutti gli anni o quasi possono essere riassunte nelle seguenti:

- ripartizione geografica: ogni anno i residenti nel Nord-Est e nel Nord-Ovest sono caratterizzati da minori probabilità di risposta e, a partire dal 1997 in poi, si aggiungono ai residenti del Nord anche quelli delle regioni del Centro;
- l'età: in generale le classi intermedie (40-64 anni) manifestano una maggiore propensione a rispondere rispetto agli anziani. Per i giovani nella classe tra i 16 e i 24 anni si segnala nel 1997 una minore probabilità di risposta rispetto agli anziani.
- salute: fatta eccezione per l'anno 1997, sono maggiormente propensi a continuare a collaborare coloro che risultano in buone o normali condizioni di salute.
- socievolezza (valutata dalla frequenza con cui l'individuo parla con persone esterne alla famiglia): gli individui che mantengono un maggiore livello di relazioni sociali, mostrano una maggiore propensione a essere parte attiva del campione.

Anche in questo caso l'analisi della devianza porta a concludere che l'*attrition* che ha colpito il campione nei diversi anni non è di tipo casuale bensì ha interessato in modo differenziato segmenti diversi della popolazione.

**Tavola 7 - Risultati di modelli di regressione logistica sulla probabilità di permanenza nel panel nei diversi anni condizionata alle caratteristiche individuali e familiari nell'anno precedente all'indagine**

Variabili individuali e familiari	Probabilità di permanenza nel panel fino all'anno 1995   variabili dell'anno 1994			Probabilità di permanenza nel panel fino all'anno 1996   variabili dell'anno 1995			Probabilità di permanenza nel panel fino all'anno 1997   variabili dell'anno 1996						
	Stima del parametro	valore del $\chi^2$	p-value	odds ratio	Stima del parametro	valore del $\chi^2$	p-value	odds ratio	Stima del parametro	valore del $\chi^2$	p-value	odds ratio	
<b>Intercetta</b>	2,1084	34,3	<0,0001	1,8896	34,5	<0,0001	2,6025	91,9	<0,0001				
<b>età</b>													
16-24	<b>0,6445</b>	36,7	<0,0001	0,2718	11,9	0,0178	<b>-0,2262</b>	12,4	0,0147	1,3123	4,0	0,0462	0,7976
25-39	<b>0,5694</b>	19,3	<0,0001	1,9050	3,3	0,0691	0,0444	0,2	0,6794	1,1182	0,2	0,6794	1,0454
40-54	<b>0,6587</b>	19,8	<0,0001	1,7673	0,7	0,3965	-0,1117	1,2	0,2716	1,4326	1,2	0,2716	0,8943
55-64	<b>0,6795</b>	27,1	<0,0001	1,9324	7,5	0,0063	0,0130	0,0	0,8982	1,3146	0,0	0,8982	1,0131
65 e più	<b>0,0000</b>	26,6	<0,0001	1,9728	4,5	0,0333	<b>0,0000</b>	.	.	1,0000	.	.	1,0000
<b>freq. di conversazione con i vicini</b>													
Sempre	<b>0,6980</b>	58,3	<0,0001	0,6828	51,2	<0,0001	<b>0,1863</b>	7,7	0,0216	1,9795	4,0	0,0448	1,2048
Qualche volta	<b>0,2992</b>	52,0	<0,0001	2,0097	42,4	<0,0001	0,0468	0,2	0,6245	1,3206	0,2	0,6245	1,0479
Raramente o mai	<b>0,0000</b>	9,3	0,0023	1,3488	6,9	0,0085	<b>0,0000</b>	.	.	1,0000	.	.	1,0000
<b>stato di salute</b>													
Ottima/buona	<b>0,4706</b>	17,4	0,0002	0,6417	44,2	<0,0001	0,1481	3,0	0,2219	1,8996	2,4	0,1251	1,1596
Normale	<b>0,3733</b>	17,0	<0,0001	1,6009	32,1	<0,0001	0,1613	2,9	0,0911	2,0629	2,9	0,0911	1,1750
Cattiva o pessima	<b>0,0000</b>	11,3	0,0008	1,4526	40,7	<0,0001	<b>0,0000</b>	.	.	1,0000	.	.	1,0000
<b>Ripartizione geografica</b>													
Nord-Ovest	<b>-0,8623</b>	131,7	<0,0001	-0,8924	110,6	<0,0001	<b>-0,8749</b>	161,0	<0,0001	0,4097	127,2	<0,0001	0,4169
Nord-Est	<b>-0,9331</b>	78,0	<0,0001	-0,6648	84,6	<0,0001	<b>-0,8418</b>	114,4	<0,0001	0,5144	114,4	<0,0001	0,4309
Centro	-0,1635	93,8	<0,0001	-0,0272	41,9	<0,0001	<b>-0,4522</b>	30,5	<0,0001	0,9732	30,5	<0,0001	0,6362
Sud	<b>0,0000</b>	2,0	0,1564	0,8491	0,1	0,8172	<b>0,0000</b>	.	.	1,0000	.	.	1,0000
<b>Devianza (b)</b>	677 (18)			318			345						
<b>p-value</b>	<0,0001			<0,0001			<0,0001						
<b>Numerosità campionaria</b>	17729			16628			15751						

(a) I parametri significativamente diversi da zero, a livello di significatività del 5%, sono in grassetto.-

(b) La devianza è pari a  $-2\log(\text{verosimiglianza del modello con la sola intercetta}/\text{verosimiglianza del modello stimato})$ . Tra parentesi sono riportati i gradi di libertà, uguali per tutti i modelli.

**Tavola 7 continua - Risultati di modelli di regressione logistica sulla probabilità di permanenza nel panel nei diversi anni condizionate alle caratteristiche individuali e familiari nell'anno precedente all'indagine**

Variabili individuali e familiari	Probabilità di permanenza nel panel fino all'anno 1998   variabili dell'anno 1997		Probabilità di permanenza nel panel fino all'anno 1999   variabili dell'anno 1998		Probabilità di permanenza nel panel fino all'anno 2000   variabili dell'anno 1999	
	Stima del parametro	valore del $\chi^2$	Stima del parametro	valore del $\chi^2$	Stima del parametro	valore del $\chi^2$
<b>Intercetta</b>	1,8021	66,1 <0,0001	1,8749	47,7 <0,0001	1,8091	39,9 <0,0001
<b>età</b>						
16-24	-0,1867	19,8 0,0006	0,1369	22,3 0,0002	0,2111	14,7 0,0053
25-39	0,0374	2,3 0,1321	<b>0,2582</b>	1,0 0,3257	<b>0,2664</b>	2,3 0,1292
40-54	0,1198	0,1 0,7470	<b>0,4121</b>	3,9 0,0472	<b>0,3287</b>	4,2 0,0403
55-64	<b>0,2968</b>	0,1 0,2816	<b>0,4864</b>	11,0 0,0009	<b>0,4440</b>	7,3 0,0070
65 e più	<b>0,0000</b>	6,6 0,0102	<b>0,0000</b>	14,8 0,0001	<b>0,0000</b>	12,8 0,0003
						1,5589
						1,0000
						1,0000
<b>frequenza di conversazione con i vicini</b>						
Sempre	<b>0,5708</b>	56,0 <0,0001	<b>0,4752</b>	21,3 <0,0001	<b>0,4898</b>	21,1 <0,0001
Qualche volta	0,1683	34,4 <0,0001	<b>0,3020</b>	20,5 <0,0001	<b>0,3061</b>	19,8 <0,0001
Raramente o mai	<b>0,0000</b>	3,0 0,0848	<b>0,0000</b>	8,1 0,0044	<b>0,0000</b>	7,6 0,0059
						1,3581
						1,0000
						1,0000
<b>stato di salute</b>						
Ottima/buona	<b>0,5494</b>	32,2 <0,0001	<b>0,4098</b>	14,0 0,0009	<b>0,4397</b>	23,9 <0,0001
Normale	<b>0,2647</b>	28,2 <0,0001	<b>0,3563</b>	13,2 0,0003	<b>0,5192</b>	16,0 <0,0001
Cattiva o pessima	<b>0,0000</b>	6,9 0,0085	<b>0,0000</b>	10,5 0,0012	<b>0,0000</b>	23,2 <0,0001
						1,6806
						1,0000
<b>Ripartizione geografica</b>						
Nord-Ovest	<b>-0,8046</b>	169,6 <0,0001	<b>-0,6926</b>	87,5 <0,0001	<b>-0,7260</b>	136,2 <0,0001
Nord-Est	<b>-0,7595</b>	81,9 <0,0001	<b>-0,6624</b>	50,3 <0,0001	<b>-1,0756</b>	52,4 <0,0001
Centro	<b>-1,0244</b>	70,3 <0,0001	<b>-0,7592</b>	45,2 <0,0001	<b>-0,5471</b>	132,4 <0,0001
Sud	<b>0,0000</b>	156,5 <0,0001	<b>0,0000</b>	67,1 <0,0001	<b>0,0000</b>	29,9 <0,0001
						0,5786
						1,0000
<b>Devianza (b)</b>	354		248		422	
<b>p-value</b>	<0,0001		<0,0001		<0,0001	
<b>Numerosità campionaria</b>	14184		12891		11895	

## 6. Conclusioni e riflessioni per ulteriori approfondimenti

L'analisi del fenomeno dell'*attrition* per la componente italiana dell'ECHP era già stata svolta (Baldassarri, Regoli, 1998) restringendo il periodo temporale alle prime tre onde e utilizzando strumenti di analisi statistica differenti da quelli impiegati in questo lavoro. I risultati ottenuti in passato spingevano però ad approfondire l'analisi.

Il presente lavoro, estendendo il periodo di osservazione e includendo un numero maggiore di variabili a cui poter imputare la selettività osservata, si è posto l'obiettivo di comprendere meglio la natura del fenomeno dell'*attrition*.

Dall'analisi svolta emerge che l'entità dell'*attrition* per la componente italiana del Panel Europeo sulle famiglie, è trascurabile se confrontata con quella riscontrata negli altri paesi partecipanti all'ECHP (Eurostat, 1999a). Con riferimento solamente alle prime tre onde di indagine, l'Italia non pare fortemente interessata al fenomeno. Per il nostro Paese il tasso di *attrition* è stato del 4,46% tra la prima e la seconda onda, e del 3,94% dalla seconda alla terza, mentre la media europea si è attestata sul 10% in entrambi i casi, con valori massimi nel Regno Unito (23% e 18% rispettivamente tra la prima e la seconda onda, e tra la seconda e la terza) e minimi in Portogallo (4% tra la prima e la seconda onda) e Germania (3% tra la seconda e la terza onda). Anche l'esame dei tassi di *attrition* riportati nella letteratura relativa ad altri panel extraeuropei (ad esempio PSID e SIPP) porta a concludere che le entità rilevate nel caso italiano non destano preoccupazione.

Cionondimeno l'analisi condotta in questo lavoro ha evidenziato come particolari segmenti della popolazione, definiti da variabili territoriali, demografiche e individuali, siano interessati in maniera differenziata dal fenomeno dell'*attrition*. Un risultato analogo era già stato evidenziato da Eurostat relativamente all'analisi dell'*attrition* nel passaggio dalla prima alla seconda onda (Eurostat, 1999b). L'approfondimento condotto in questo lavoro con riferimento ad un arco temporale di 7 onde rafforza la necessità di mettere a punto un sistema correttivo di ponderazione che tenga conto della differente probabilità di risposta di gruppi con le caratteristiche trovate e nutre il dibattito attualmente in corso presso Eurostat per il computo dei pesi campionari.

Da questo lavoro emerge che una quota di *attrition* è imputabile a fenomeni naturali ed è quindi fisiologica: questo tipo di *attrition* ovviamente non incide sulla qualità dell'indagine, anzi riflette nel campione, attraverso i decessi, la dinamica demografica della popolazione. Una possibile linea di approfondimento è, quindi, la distinzione tra i veri *attritors* e quelli che risultano *attritors* perché non più obiettivo di indagine (morti, istituzionalizzati, trasferiti in un paese fuori UE). Altro spunto di analisi potrebbe essere l'implementazione delle analisi svolte anche sul gruppo dei 'rientrati' al fine di verificare se sussistano delle differenze comportamentali rispetto agli *attritors* qui esaminati.

Si fa presente che l'analisi dinamica dell'*attrition* considera la “dinamica nella risposta” nel senso che i modelli implementati considerano come variabile dipendente la probabilità di rispondere continuativamente fino all'anno in esame contro la probabilità di rispondere fino all'anno precedente ma le variabili esplicative sono di tipo statico, generalmente riferite all'ultima onda disponibile (anno precedente a quello in cui si verifica l'eventuale uscita). Sarebbe utile estendere lo studio anche con l'inclusione di variabili dinamiche, che misurano quindi i cambiamenti di stato o le persistenze in una stessa condizione nel corso dell'indagine, per isolare eventi che potrebbero eventualmente avere un impatto sul fenomeno dell'*attrition*.

### Riferimenti bibliografici

- Baldassarri E., Regoli A., 1998, *La qualità dei dati nell'indagine Panel Europeo sulle famiglie: una valutazione dell'attrito nelle prime tre waves (1994-1996)*, Atti della Riunione scientifica della Società Italiana di Statistica, Sorrento.
- Ernst L.R., 1989, *Weighting Issues for Longitudinal Household and Family Estimates*, in Kasprzyk D., Duncan G., Kalton G., Singh M.P. (Eds.), *Panel Surveys*, John Wiley and Sons, United States, pp. 139-159.
- Eurostat, 1994, *Tracing for the Second Wave of the ECHP*, ECHP Doc. Pan. 27/94, Lussemburgo.
- Eurostat, 1999a, *ECHP Data Quality - Second Report*, ECHP Doc. Pan. 108/99, Lussemburgo.
- Eurostat, 1999b, *Sample attrition between Waves 1 and 2 in the European Community Household Panel*, ECHP Doc. Pan. 118/99, Lussemburgo.
- Fitzgerald J., Gottschalk P., Moffitt R., 1998, *An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics*, *The Journal of Human Resources* 33, N. 2, pp 255-299.
- Folsom R., La Vange L., Williams R.L., 1989, *A Probability Sampling Perspective on Panel Data Analysis*, in Kasprzyk D., Duncan G., Kalton G., Singh M.P. (Eds.), *Panel Surveys*, John Wiley and Sons, United States, pp. 108-138.
- Istat, 2001, *La situazione economica e finanziaria delle famiglie in Italia e in Europa. Panel europeo sulle famiglie - Anni 1994-1996*, Collana Informazioni, n. 11.
- Kasprzyk D., Duncan G.J., Kalton G., Singh M.P. (Eds.), 1989, *Panel Surveys*, John Wiley and Sons, United States.
- Little R.J.A., Rubin D.B., 1987 *Statistical analysis with missing data*, John Wiley and Sons, New York, pp. 218-230.

Pannenberg M., 2000, *Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (GSOEP) (1984 until 1999)*, Discussion Papers n. 232, Deutsches Institut für Wirtschaftsforschung, Berlino.

Trivellato U., 1999, *Issues in the Design and Analysis of Panel Studies: A Cursory Review*, *Quality & Quantity*, n. 33, pp. 339-352.

### ***An analysis of attrition in the European Community Household Panel***

**Summary:** *The attrition represents one of the most important problem of panel surveys and the consequences of this phenomenon are related more to the nature than to the extent. The aim of this paper is to analyse the attrition process affecting the sample of the European Community Household Panel, a longitudinal survey annually conducted in the European Union Member States since 1994. The paper first examines the extent of the Italian sample attrition in the 1994-2000 period; it then highlights what kind of respondents are more likely to definitely leave the sample in comparison to those who keep on taking part to the survey. The combined use of binary segmentation techniques and logit models make it possible to identify the respondents' characteristics affecting the attrition.*

**Keywords:** *Panel surveys, attrition, logistic regression.*