

Appendice 1. Metodi –

*Chris Roberts, Jorma Tynjälä,
Dorothy Currie and Matt King*

Gli obiettivi principali di questa appendice tecnica sono i seguenti: fornire informazioni sulle procedure utilizzate nell'indagine HBSC 2001/2002 per il campionamento e la raccolta dei dati ed assistere il lettore nell'interpretazione dei risultati contenuti nel rapporto.

Campionamento

I paesi e le regioni dovrebbero selezionare il proprio campione in modo da assicurarsi la possibilità di fare validi paragoni nel tempo, nell'ambito delle aree considerate e tra le stesse. Tuttavia, in un progetto di ricerca internazionale che mette a confronto i dati tra i vari paesi, il campione deve anche essere selezionato in modo simile in ciascun paese. I confronti transnazionali validi sono particolarmente importanti quando si evidenziano caratteristiche comuni e diversità tra paesi e regioni. Inoltre, i ricercatori e i responsabili delle decisioni politiche in ogni paese o regione dovrebbero considerare il campione come realmente rappresentativo, così da essere sicuri della rilevanza dei risultati quando intraprendono iniziative per la promozione della salute. Molti partecipanti all'indagine sono più interessati ai dati sui cambiamenti che riguardano il loro paese o la loro regione nel corso del tempo, piuttosto che ai confronti tra i diversi paesi; le informazioni su questi cambiamenti permettono loro di stimare l'impatto degli interventi per la promozione della salute a livello nazionale o regionale. Questa parte del rapporto riassume le procedure per il campionamento utilizzate nel 2001/2002. L'attuale protocollo HBSC (1) fornisce tutti i dettagli.

Popolazione target

Il campione selezionato è composto da studenti di 11, 13 e 15 anni, vale a dire, giovani che sono nel loro dodicesimo, quattordicesimo e sedicesimo anno di vita. In alcuni paesi e regioni, ciascuna fascia d'età frequenta il corrispondente anno scolastico, perché i giovani sono promossi ogni anno. In altri paesi e regioni, alcuni giovani sono in ritardo e altri più avanti rispetto al normale iter scolastico, ma anche questi devono essere inclusi nel campione come gli studenti che sono promossi all'anno scolastico successivo secondo l'iter previsto. Il 90% dei soggetti intervistati dovrebbe avere una differenza d'età di 6 mesi al massimo rispetto all'età media della fascia considerata e il restante 10% una differenza d'età di 12 mesi al massimo sempre rispetto all'età media. L'età media auspicata per i tre gruppi è rispettivamente di 11,5, 13,5 e 15,5 anni.

L'ideale sarebbe intervistare tutti i giovani della fascia d'età pertinente, sia che frequentino scuole private, pubbliche o istituti speciali per alunni con esigenze particolari. In realtà, sono pochi gli intervistati che trovano difficile rispondere al questionario o che frequentano istituti speciali difficili da raggiungere. Si presuppone che il 95% circa della popolazione target sia disponibile al campionamento. La maggior parte dei paesi e delle regioni stratificano i campioni per garantire una buona copertura geografica. Nel 2001/2002 in Germania è stato scelto un campione regionale (Berlino, Assia, Vestfalia e Sassonia).

Belgio (Fiandre) e Belgio (francofono) sono trattati separatamente, così come Inghilterra, Scozia e Galles. Poiché la popolazione della Groenlandia è piuttosto esigua, è stato fatto un censimento della popolazione scolastica, dalla quale mancano solo i giovani che erano assenti il giorno della raccolta dei dati.

Selezione del campione

Abbiamo utilizzato un campionamento “a grappolo”, in cui l’unità di campionamento principale è la classe (o la scuola in mancanza di un campione costituito da classi) piuttosto che il singolo studente, come in un campione casuale semplice. Se, da un lato, il campionamento “a grappolo” non è così preciso come il campionamento casuale semplice, dall’altro lato risulta efficace da somministrare e può avere la stessa precisione del campionamento casuale semplice se si aumenta la numerosità del campione.

Quando si utilizza il campionamento “a grappolo”, le risposte degli studenti non possono essere considerate indipendenti, visto che gli studenti all’interno della stessa classe o scuola, hanno maggiori probabilità di essere simili tra loro che gli studenti in generale. Pertanto il campionamento “a grappolo” causa errori standard che tendono a essere più numerosi di quelli che si otterrebbero se lo stesso campione fosse ottenuto usando un campionamento casuale semplice (2). Se gli errori standard aumentano, anche la numerosità del campione deve essere aumentata per mantenere il livello di precisione della stima dell’indagine. L’effetto disegno è la quantità per cui moltiplicare la numerosità del campione, calcolata per un campione casuale semplice, per tener conto del campionamento complesso ed è definito come il rapporto tra l’errore standard derivato da un’indagine complessa e quello ottenuto presupponendo un campione casuale semplice (3).

Abbiamo stabilito una numerosità minima del campione per le tre fasce d’età, pari a 1.536 studenti. Questo calcolo presuppone, nel 95% dei casi, un intervallo di confidenza del $50\% \pm 3\%$, e un effetto disegno pari a 1,2, basandosi sull’analisi dei dati dell’indagine HBSC del 1993/1994 e del 1997/1998 (4, 5). Gli intervalli di confidenza sono normalmente inclusi per indicare il livello di precisione associato alla stima dell’indagine, perché essi indicano il grado di rappresentatività del campione (6).

Estrazione del campione

Data la diversità dei sistemi scolastici, delle età di ammissione a scuola e dei livelli di ritardo e/o avanzamento degli studenti nei vari paesi, è molto difficile proporre un approccio uniforme per il campionamento. Per superare questa complessità, il fattore prioritario scelto per le procedure di campionamento utilizzate in questo studio è stato l’età: ogni campione è stato selezionato considerando tutti quelli che si collocano nella fascia d’età adatta, secondo le tre fasce d’età. Quando tutti gli studenti con l’età adatta frequentano lo stesso anno, il campione può essere selezionato considerando solo quell’anno, ma si campionano tutti gli anni quando le fasce d’età sono ripartite su più anni scolastici. La situazione si complica ulteriormente quando la popolazione target è presente in due cicli scolastici, per esempio nelle scuole primarie e nelle secondarie.

Nei casi in cui non si conosceva il numero di classi adatte al campionamento, il numero stesso è stato stimato considerando l’intera popolazione scolastica. Se una scuola ha quattro classi adatte al campionamento, allora ciascuna classe deve avere la stessa probabilità di essere rappresentata nel campione di una scuola con una sola classe. Di conseguenza, il valore di ogni scuola è stato ponderato secondo il numero delle classi adatte. Quando è stata selezionata una scuola con due o più classi, quella da campionare è stata scelta casualmente. Questo ha assicurato a ogni classe della popolazione target la stessa probabilità di essere selezionata. Presupponendo una media di 25 studenti per classe, sarebbero state necessarie 62 classi per raggiungere la numerosità consigliata di 1.536 studenti per fascia d’età in ogni paese o regione.

In alcuni paesi o regioni, per ridurre al minimo il numero di scuole necessarie, alcune classi sono state campionate aleatoriamente per una fascia d’età e in alcune scuole, mentre altre classi nelle stesse scuole sono state campionate per le altre due fasce d’età. Ai paesi e alle regioni sono state date istruzioni

affinché considerassero la dimensione della classe, le percentuali di frequenza scolastica e le percentuali di consenso quando si prendeva in considerazione il numero di scuole necessario a raggiungere la numerosità del campione di studio.

Per ottenere l'età media di 11,5, 13,5 e 15,5 anni, l'indagine è stata svolta in periodi dell'anno adatti in ogni paese e regione. Se gli studenti di una particolare fascia d'età frequentano anni scolastici diversi, (paesi in cui gli studenti avanzano o rimangono indietro a seconda del rendimento scolastico), nella maggior parte dei casi si campionano tutti gli anni scolastici. In questi casi, paesi e regioni hanno creato una sorta di "classe" che raggruppa gli studenti distribuiti su più anni scolastici. Il protocollo (1) fornisce ulteriori dettagli sul campionamento.

Raccolta dei dati e preparazione dell'archivio dati

Quasi tutti i questionari sono stati distribuiti nelle scuole tra ottobre 2001 e maggio 2002. La Tabella 1 indica i periodi della raccolta dei dati in ciascun paese e regione.

Nella maggior parte dei paesi e delle regioni, i questionari sono stati consegnati alle scuole, distribuiti

Tabella 1. Date della raccolta dei dati dell'indagine HBSC 2001/2002, per paese o regione

Paese o regione	Date
Austria	Ottobre–Novembre 2001
Belgio (Fiandre)	Marzo–Aprile 2002
Belgio (francofono)	Gennaio–Febbraio 2002
Canada	Gennaio–Febbraio 2002
Croazia	Febbraio–Marzo 2002
Repubblica Ceca	Maggio 2002
Danimarca	Gennaio–Febbraio 2002
Inghilterra	Marzo 2002
Estonia	Ottobre–Novembre 2001
Finlandia	Marzo–Maggio 2002
Francia	Marzo–Giugno 2002
Germania	Marzo–Maggio 2002
Grecia	Ottobre–Novembre 2002
Groenlandia	Maggio 2002
Ungheria	Marzo–Aprile 2002
Irlanda	Aprile–Giugno 2002
Israele	Maggio–Giugno 2002
Italia	Aprile 2002
Lettonia	Novembre–Dicembre 2001
Lituania	Febbraio–Marzo 2002
Malta	Gennaio 2002
Olanda	Ottobre–Novembre 2001
Norvegia	Dicembre 2001
Polonia	Febbraio–Marzo 2002
Portogallo	Marzo–Aprile 2002
Federazione Russa	Marzo 2002
Scozia	Febbraio–Aprile 2002
Slovenia	Marzo 2002
Spagna	Aprile–Maggio 2002
Svezia	Novembre–Dicembre 2001
Svizzera	Marzo–Maggio 2002
FYROM	Marzo 2002
Ucraina	Febbraio 2002
Stati Uniti	Novembre–Dicembre 2001
Galles	Febbraio–Marzo 2002

dagli insegnanti e restituiti all'istituto di ricerca dopo essere stati compilati. Tuttavia, in alcuni casi, i ricercatori stessi hanno distribuito il questionario per alleggerire l'onere delle scuole. Tutto il personale coinvolto nella raccolta dei dati è stato addestrato e ha seguito le linee guida concordate.

Gli archivi provenienti dai 35 paesi e regioni sono stati preparati e trasferiti nella Banca Dati Internazionale HBSC (HBSC International Data Bank), presso il Norwegian Social Science Data Services, Università di Bergen. I dati sono stati controllati e corretti in base a criteri rigorosi (1). I dati per i giovani al di fuori delle fasce d'età scelte sono stati tolti e tutte le deviazioni dallo standard internazionale sono

Tabella 2. Numero dei soggetti intervistati nell'indagine HBSC del 2001/2002, per paese o regione, sesso e fascia d'età

Paese o regione	Sesso		Fascia d'età (anni)			Totale
	Maschi	Femmine	11	13	15	
Austria	2241	2231	1590	1584	1298	4472
Belgio (Fiandre)	2996	3293	2153	2106	2030	6289
Belgio (francofono)	2069	2254	1439	1503	1381	4323
Canada	1996	2365	1641	1513	1207	4361
Croazia	2180	2217	1451	1500	1446	4397
Repubblica Ceca	2412	2600	1691	1661	1660	5012
Danimarca	2259	2413	1710	1582	1380	4672
Inghilterra	2943	3138	2239	2069	1773	6081
Estonia	1983	1996	1288	1424	1267	3979
Finlandia	2713	2675	1911	1732	1745	5388
Francia	4054	4131	2671	2900	2614	8185
Germania	2786	2864	2100	1801	1749	5650
Grecia	1870	1937	1252	1231	1324	3807
Groenlandia	386	505	295	356	240	891
Ungheria	1848	2316	1371	1463	1330	4164
Irlanda	1302	1573	1012	944	919	2875
Israele	2625	3036	1892	2202	1567	5661
Italia	2125	2261	1524	1633	1229	4386
Lettonia	1633	1848	1195	1169	1117	3481
Lituania	2887	2758	1867	1873	1905	5645
Malta	905	1075	619	694	667	1980
Olanda	2120	2149	1477	1519	1273	4269
Norvegia	2554	2469	1660	1739	1624	5023
Polonia	3204	3179	2100	2131	2152	6383
Portogallo	1419	1521	1174	964	802	2940
Federazione Russa	3752	4285	2522	2940	2575	8037
Scozia	2246	2158	1743	1512	1149	4404
Slovenia	1996	1960	1474	1413	1069	3956
Spagna	2873	2954	2105	1966	1756	5827
Svezia	1978	1948	1499	1201	1226	3926
Svizzera	2309	2370	1468	1671	1540	4679
FYROM	2053	2108	1348	1401	1412	4161
Ucraina	1893	2197	1192	1297	1601	4090
Stati Uniti	2322	2703	1479	1921	1625	5025
Galles	2004	1883	1351	1372	1164	3887
Total	78 936	83 370	55 503	55 987	50 816	162 306

Tabella 3. Età medie dei soggetti intervistati nell'indagine HBSC del 2001/2002, per paese o regione e fascia d'età

Paese o regione	Età del soggetto intervistato (anni)		
	11 anni	13 anni	15 anni
Austria	10.8	12.8	14.8
Belgio (Fiandre)	11.5	13.5	15.5
Belgio (francofono)	11.5	13.4	15.5
Canada	11.8	13.8	15.7
Croazia	11.4	13.4	15.4
Repubblica Ceca	11.5	13.5	15.4
Danimarca	11.8	13.8	15.8
Inghilterra	11.9	13.9	15.9
Estonia	11.4	13.3	15.3
Finlandia	11.8	13.8	15.8
Francia	11.2	13.1	15.1
Germania	11.6	13.6	15.7
Grecia	11.4	13.3	15.3
Groenlandia	12.3	14.3	16.4
Ungheria	11.5	13.5	15.5
Irlanda	11.7	13.5	15.4
Israele	12.0	14.1	16.0
Italia	11.8	13.8	15.9
Lettonia	11.6	13.6	15.5
Lituania	11.7	13.7	15.7
Malta	11.7	13.7	15.6
Olanda	11.5	13.5	15.5
Norvegia	11.5	13.5	15.5
Polonia	11.7	13.7	15.7
Portogallo	12.1	14.2	16.1
Russia	11.6	13.6	15.6
Scozia	11.5	13.6	15.5
Slovenia	11.7	13.7	15.8
Spagna	11.5	13.5	15.5
Svezia	11.4	13.5	15.5
Svizzera	11.7	13.8	15.8
FYROM	11.5	13.5	15.5
Ucraina	11.9	13.9	16.0
Stati Uniti	11.6	13.5	15.5
Galles	12.0	14.0	16.0
Totale	11.6	13.6	15.6

state documentate. Il protocollo di ricerca (1) fornisce una serie completa di istruzioni per la correzione dei dati.

Le Tabelle 2 e 3 contengono informazioni sui soggetti intervistati presenti nell'archivio-dati internazionale. Nell'indagine del 2001/2002 sono stati raccolti i dati di più di 160.000 giovani. La ripartizione dei soggetti intervistati per quanto riguarda sesso e fascia d'età è abbastanza uniforme. Considerando l'intero campione, l'età media è pari a 11,6; 13,6; e 15,6 anni, rispettivamente per le fasce d'età di 11, 13 e 15 anni. Sono tuttavia presenti delle deviazioni, dai 10,8 anni in Austria ai 12,3 anni in Groenlandia, per quanto riguarda la fascia d'età più giovane, con un andamento simile per i tredicenni e i quindicenni.

Analisi e interpretazione dei dati

Quando si interpretano i risultati contenuti in questo rapporto è necessario considerare molti fattori importanti. In questo paragrafo ne analizziamo due: l'impatto del disegno di campionamento e l'analisi e la presentazione appropriate dei dati.

Impatto del disegno di campionamento sull'interpretazione dei risultati

Le fonti dei potenziali errori nei dati dell'indagine HBSC, in particolare per quanto riguarda gli errori di campionamento, sono già state analizzate (4). Questo paragrafo fornisce comunque degli orientamenti sui fattori principali da prendere in considerazione per facilitare l'interpretazione dei dati contenuti in questo rapporto.

Lo studio HBSC si basa su un campione di soggetti intervistati e non su un censimento dell'intera popolazione (eccetto la Groenlandia), come la maggior parte delle indagini sociali. Gli errori di campionamento e altre fonti di errore casuale (per esempio errori nell'interpretazione delle domande) possono essere stimati calcolando la varianza o l'errore standard della stima dell'indagine. Molti fra i software statistici più diffusi presuppongono l'uso del campionamento casuale semplice quando calcolano la varianza della stima dell'indagine. Se così fosse, il campione sarebbe selezionato scegliendo casualmente dei soggetti da un campione base che elenca tutti i giovani in età scolare in ogni paese e regione. Con un disegno di questo genere, l'errore standard ($se = \text{standard error, errore standard}$) di una percentuale può essere calcolato usando la percentuale del campione selezionato e inserendo le cifre nella seguente equazione:

$$se(p) = \sqrt{\frac{pq}{n}} \quad \text{dove} \quad \begin{array}{l} q = 1 - p \\ n = \text{numero dei soggetti intervistati} \\ p = \text{percentuale dei soggetti intervistati con il profilo ricercato} \end{array}$$

Per esempio, nel campione gallese ci sono 1.351 undicenni ($n = 1.351$), di cui il 36% dichiara di avere subito episodi di bullismo almeno una volta nei due mesi precedenti ($p = 0,36$), quindi:

$$se(p) = \sqrt{\frac{(0,36 \times 0,64)}{1.351}} = 0,013 \text{ or } 1,3\%$$

L'intervallo di confidenza al 95% della stima dell'indagine è dato da:

$$P \pm 1,96 \times se(p)$$

che, in questo esempio, dà come risultato intervalli di confidenza del $36\% \pm 2,5\%$ (o 33,5–38,5%). Questi risultati, tradotti in parole semplici, dimostrano che esiste una probabilità del 95% che la popolazione reale si trovi all'interno degli intervalli calcolati.

Tuttavia, come osservato in precedenza, lo studio HBSC utilizza un campionamento “a grappolo”, in cui l'unità di campionamento primaria è la classe (o scuola) piuttosto che il singolo studente, come in un campione casuale semplice. Dato tale disegno, le risposte degli studenti non possono essere considerate indipendenti, visto che gli studenti all'interno della stessa classe o scuola, hanno maggiori probabilità di essere simili tra loro che gli studenti in generale. Il campionamento “a grappolo”, quindi, causa errori standard che tendono a essere più numerosi di quelli che si otterrebbero se lo stesso campione fosse ottenuto usando un campionamento casuale semplice. Di conseguenza, gli errori standard devono essere calcolati utilizzando un metodo appropriato che tenga conto delle relazioni tra gli studenti all'interno delle scuole e delle classi.

Inoltre molti paesi e regioni stratificano i loro campioni, suddividendo il campione principale in unità

Tabella 4. Valori dell'effetto disegno per variabili selezionate nell'indagine HBSC del 2001/2002, 11 anni

Variable	Finlandia	Ungheria	Lettonia	Norvegia	Portogallo	Scozia	FYROM	Galles
Consuma frutta quotidianamente	1.01	1.15	1.09	1.16	1.18	1.36	1.43	1.34
In conformità con le linee guida sull'attività fisica ^a	1.61	1.18	1.26	1.58	1.24	1.46	1.24	1.39
Salute discreta o scadente	0.95	1.19	1.22	1.07	1.14	1.20	1.25	1.15
Ha subito episodi di bullismo negli ultimi mesi	1.17	1.55	1.21	1.34	1.31	1.24	1.52	1.21
Ha subito infortuni nell'ultimo anno	1.13	1.19	1.02	1.21	1.07	1.24	1.54	1.27
È interessato alla scuola	1.18	1.46	1.41	1.20	1.12	1.44	1.83	1.55
Risultati scolastici ^b	1.33	1.29	1.40	1.09	1.15	1.32	1.18	1.10
Trascorre quattro o più sere con gli amici ogni settimana	1.46	1.46	1.31	1.40	1.16	1.64	1.47	1.84
Comunicazione elettronica giornaliera	1.06	1.24	1.54	1.12	1.07	1.24	1.59	1.05
Comunica facilmente con la madre	1.05	1.27	1.06	1.10	0.98	1.14	1.35	1.14

^a punteggio MVPA pari o superiore a 5 (Capitolo 3, pagg. 94–101).

^b Coloro che pensano di avere un rendimento scolastico buono o molto buono.

Tabella 5. Valori dell'effetto disegno per variabili selezionate nell'indagine HBSC del 2001/2002, 13 anni

Variabile	Finlandia	Ungheria	Lettonia	Norvegia	Portogallo	Scozia	FYROM	Galles
Fuma tutti i giorni	1.42	1.46	1.30	1.25	1.43	1.12	1.01	1.36
Si è ubriacato due o più volte	1.29	1.17	1.22	1.20	1.22	1.18	1.29	1.30
Consuma frutta quotidianamente	1.18	1.25	1.14	1.17	1.04	1.14	1.27	1.18
In conformità con le linee guida sull'attività fisica ^a	1.15	1.41	1.15	1.25	1.13	1.21	1.23	1.22
Salute discreta o scadente	1.00	1.38	1.28	1.10	1.25	1.26	1.34	1.12
Ha subito episodi di bullismo negli ultimi mesi	1.17	1.14	1.18	1.28	1.26	1.03	1.43	1.21
Ha subito infortuni nell'ultimo anno	1.09	1.11	1.15	1.14	1.11	1.17	1.50	1.30
È interessato alla scuola	1.43	2.48	1.61	1.36	1.29	1.17	2.31	1.32
Risultati scolastici ^b	1.18	1.24	1.21	1.27	1.06	1.12	1.25	1.21
Trascorre quattro o più sere con gli amici ogni settimana	1.53	1.35	1.27	1.51	1.09	1.46	1.50	2.16
Comunicazione elettronica giornaliera	1.13	1.13	1.61	1.09	1.33	1.14	1.79	1.16
Comunica facilmente con la madre	1.23	1.09	1.16	1.11	1.20	1.17	1.23	1.00

^a punteggio MVPA pari o superiore a 5 (Capitolo 3, pagg. 94–101).^b Coloro che pensano di avere un rendimento scolastico buono o molto buono.

Tabella 6. Valori dell'effetto disegno per variabili selezionate nell'indagine HBSC del 2001/2002, 15 anni

Variable	Finlandia	Ungheria	Lettonia	Norvegia	Portogallo	Scozia	FYROM	Galles
Fuma tutti i giorni	1.11	1.79	1.29	1.29	1.11	1.26	1.57	1.62
Si è ubriacato due o più volte	1.26	1.95	1.35	1.56	1.14	1.15	1.37	1.01
Consuma frutta quotidianamente	1.18	1.24	1.04	1.24	1.08	1.08	1.35	1.11
In conformità con le linee guida sull'attività fisica ^a	1.26	1.50	1.12	1.25	1.36	1.30	1.07	1.16
Salute discreta o scadente	1.11	1.41	1.19	0.95	1.00	1.23	1.30	1.34
Ha subito episodi di bullismo negli ultimi mesi	1.16	1.12	1.26	1.42	1.03	0.98	1.67	1.23
Ha subito infortuni nell'ultimo anno	1.08	1.20	1.26	0.98	1.01	1.12	1.40	1.25
È interessato alla scuola	1.09	1.88	1.37	1.21	1.30	1.06	2.42	1.20
Risultati scolastici ^b	1.08	1.85	1.06	0.96	1.10	0.95	1.36	1.29
Trascorre quattro o più sere con gli amici ogni settimana	1.44	1.56	1.29	1.30	1.05	1.35	1.68	1.54
Comunicazione elettronica giornaliera	1.27	1.41	1.67	1.26	1.22	1.19	2.10	0.98
Comunica facilmente con la madre	1.00	1.10	1.13	1.02	0.93	0.94	1.18	1.12

^a punteggio MVPA pari o superiore a 5 (Capitolo 3, pagg. 94–101).

^b Coloro che pensano di avere un rendimento scolastico buono o molto buono.

più piccole, spesso aree geografiche, per assicurare rappresentatività a tutte le regioni. Grazie a questa stratificazione è possibile ridurre gli errori standard e pertanto dovrebbe essere considerata al momento in cui si calcolano gli errori.

Vari software statistici sono ora disponibili per calcolare gli errori standard nei disegni di campionamento complessi. Come alternativa alla rappresentazione degli errori standard veri (considerando il complesso disegno di campionamento) per tutte le percentuali d'interesse in un rapporto come questo, si propone una selezione di effetti disegno. In questo caso, l'effetto disegno o design factor (deft) è il rapporto tra l'errore standard derivato dal campionamento "a grappolo" e quello ottenuto presupponendo un campione casuale semplice (3). Utilizzando l'esempio degli episodi di bullismo tra gli undicenni gallesi, l'effettivo errore standard ottenuto per questa stima è pari all'1,6%, con intervalli di confidenza, nel 95% dei casi, vicino alla stima di 32,9–39,1%. Questo regge il confronto con un intervallo di confidenza del 33,5–38,5% nell'ipotesi di campionamento casuale semplice. Il valore dell'effetto disegno (o valore deft) per questa stima è quindi pari a 1,6/1,3 or 1,23.

I valori dell'effetto disegno per le variabili selezionate sono stati calcolati per un numero ristretto di paesi e regioni e sono indicati nelle Tabelle 4–6 per undicenni, tredicenni e quindicenni. Fra questi non ci sono i valori per gli undicenni riguardanti le domande "fuma tutti i giorni" e "si è ubriacato due o più volte", vista l'esigua prevalenza di queste variabili: meno dell'1% in molti paesi e regioni. Gli effettivi errori standard sono stati calcolati usando il pacchetto software Stata (7). Le Tabelle 4–6 evidenziano variazioni sostanziali nei valori del disegno tra le variabili selezionate per ciascun paese o regione e fascia d'età, anche se emergono alcuni andamenti. Per esempio, i valori dell'effetto disegno tendono a essere più alti in diversi paesi o regioni per quanto riguarda le variabili relative alla scuola e alle amicizie (come "è interessato alla scuola" e "trascorre quattro o più sere con gli amici ogni settimana") e in riferimento ad alcuni comportamenti a rischio (come "fuma quotidianamente"). Al contrario, si registrano valori dell'effetto disegno più bassi per variabili quali "risultati scolastici", "comunica facilmente con i genitori" e "ha subito infortuni nell'ultimo anno". Valori dell'effetto disegno più bassi suggeriscono che le opinioni o i comportamenti di studenti della stessa classe o scuola hanno la stessa probabilità di essere simili tra loro e simili a opinioni e comportamenti di altri studenti selezionati su base puramente casuale.

Utilizzando un valore dell'effetto disegno adeguato, come indicato nelle Tabelle 4–6, l'effettivo errore standard (e l'intervallo di confidenza) di una variabile, che spiega il disegno complesso dell'indagine, può essere stimato moltiplicando l'errore standard (presupponendo un campionamento casuale semplice) per il corrispondente valore dell'effetto disegno.

Analisi dei dati e presentazione dei risultati

La maggior parte delle stime contenute in questo rapporto sono costituite da percentuali presentate

Tabella 7. Intervalli di confidenza (con probabilità del 95%) approssimativi nell'indagine HBSC del 2001/2002

Percentuale d'interesse (%)	Intervallo di confidenza (%)
5	± 1.9
10	± 2.6
15	± 3.1
20	± 3.4
25	± 3.7
30	± 3.9
35	± 4.1
40	± 4.2
45	± 4.3
50	± 4.3

sotto forma di semplice istogramma, suddivise per paese o regione, età e sesso. Talvolta, queste stime possono differire leggermente da quelle presentate in altri rapporti (per esempio, in relazioni nazionali o articoli su giornali e riviste). È più probabile che ciò accada quando si approfondiscono dei temi particolari e quando si uniscono due o più domande per creare una nuova misura, per esempio valutare gli attuali andamenti sul consumo di bevande alcoliche da parte dei giovani utilizzando i dati disponibili sulle frequenze con cui si assume alcol o non se ne fa alcun uso. I dati di una categoria di risposta (o combinazioni di categorie di risposta) sono indicati in modo rappresentativo. In teoria si dovrebbero dare i vari intervalli di confidenza per ciascuna stima dell'indagine, a condizione che nella popolazione target sia stata riscontrata la gamma di valori adatti. Questo non è fattibile per un rapporto di questa portata, ma la Tabella 7 fornisce intervalli di confidenza approssimativi per una serie di percentuali. Calcolando questi intervalli si presuppone un campione di 750 soggetti e i dati sono suddivisi per età e sesso in ogni paese o regione. Inoltre si presuppone un valore dell'effetto disegno pari a 1,2 per tener conto della natura "a grappolo" dei dati. I vari intervalli di confidenza sono simmetrici vicino al 50%: per esempio l'intervallo di confidenza sia per il 40% che per il 60% è $\pm 4,2\%$, per il 70% e 30%, è $\pm 3,9\%$ e così via. Per esempio, se la percentuale stimata delle femmine gallesi di 15 anni che fumano una volta la settimana fosse del 27%, l'intervallo di confidenza, nel 95% dei casi, sarebbe di circa $\pm 3,9\%$ e il dato sulla popolazione vera sarebbe compreso tra il 23% e il 31%.

In alcuni paragrafi gli autori hanno anche riportato l'associazione tra variabili d'interesse e fattori ad esse collegati in forma di semplici associazioni bivariate: per esempio nel Capitolo 2, (pagg. 42–51) sono riportati i dati relativi al sostegno per gli studenti e all'interesse per la scuola.

Data la natura ordinale e nominale di molte variabili dell'indagine HBSC, l'analisi è stata condotta utilizzando soprattutto statistiche non parametriche come i coefficienti di correlazione tra ranghi di Spearman (8). Nella maggior parte dei casi, queste associazioni sono state calcolate aggregando dati per tutti i paesi e regioni, poiché lo scopo principale è di fornire modelli generali.

Il rilievo statistico di ogni associazione non è stato riportato vista la dimensione del campione quando si esamina l'intero archivio di dati. Con una tale dimensione del campione ci si aspetta che la maggioranza dei coefficienti, anche se bassi, sia statisticamente importante e la presentazione dei valori P insignificante.

Bibliografia

1. Roberts C, François Y, King A. Sampling. In: Currie C, Samdal O, Boyce W et al., eds. *Health Behaviour in School-aged Children: a World Health Organization cross-national study (HBSC). Research protocol for the 2001/2002 survey*. Edinburgh, Child and Adolescent Health Research Unit, University of Edinburgh, 2001:28–33 (http://www.hbsc.org/downloads/Protocol_Section1.pdf, accessed 27 January 2004).
2. Levy SP, Lemeshow S. *Sampling of populations. Methods and applications*. New York, Wiley, 1991.
3. Kish L. *Survey sampling*. New York, Wiley, 1965.
4. Roberts C. Sample design and sampling error. In: King A, et al. *The health of youth. A cross-national survey*. Copenhagen, WHO Regional Office for Europe, 1996:211–216 (WHO Regional Publications, European Series, No .69). http://www.euro.who.int/InformationSources/Publications/Catalogue/20010911_53; accessed 20 January 2004).
5. Currie C et al., eds. *Health and health behaviour among young people. International report from the HBSC 1997/98 survey*. Copenhagen, WHO Regional Office for Europe, 2000 (Health Policy for Children and Adolescents, No. 1; http://www.hbsc.org/downloads/Int_Report_00.pdf, accessed 27 January 2004).
6. Gardner MJ, Altman DG. *Statistics with confidence: confidence intervals and statistical guidelines*, 2nd ed. London, BMJ Publishing, 2000.
7. Stata statistical software: release 8.0. College Station, TX, Stata Corporation, 2003.
8. Siegel S, Castellan N. *Nonparametric statistics for the behavioural sciences*, 2nd ed. New York, McGraw-Hill, 1988.