

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**L'approccio controfattuale nei modelli di
mediazione: studio del ruolo delle condizioni di
salute nella relazione fra stato socio – economico e
benessere.**

Relatrice: Prof.ssa Giovanna Boccuzzo
Dipartimento di Scienze Statistiche

Laureanda: Sara Manzella
Matricola N. 1154440

Anno Accademico 2018/2019

*“She was like the sun.
She knew her place in the world.
She would shine again regardless
of all the storms and changeable
weather.
She wouldn’t adjust her purpose
for things that pass.”*

Nikki Rowe

INDICE

Introduzione	1
1. Concetti chiave	5
1.1. Stato socio – economico	5
1.1.1. Misure monodimensionali	7
1.1.2. Misure composite	8
1.1.3. Ricchezza	10
1.2. Benessere	11
1.3. Salute fisica	13
1.4. La relazione tra le variabili in letteratura	19
2. I dati	23
2.1. L'indagine SHARE	23
2.2. Campionamento e criteri di eleggibilità	24
2.3. Le wave	25
2.3.1. Wave 1	27
2.3.2. Wave 2	27
2.3.3. Wave 3 – SHARELIFE	29
2.3.4. Wave 4	30
2.3.5. Wave 5	31
2.3.6. Wave 6	31
2.3.7. Wave 7	31
3. Le variabili	33
3.1. Il campione	34
3.1.1. Demografiche	34
3.1.2. Outcome di salute	37
3.1.3. Misure del benessere	45

3.1.4. Variabili socio – economiche	50
3.2. Creazione delle variabili	53
3.3. Relazioni tra variabili e possibili confondenti	59
3.4. Identificazione dei confondenti	63
4. Analisi di mediazione	67
4.1. Cosa si intende con il termine “mediazione”?	67
4.2. Il controfattuale	68
4.3. Approcci basati sulla regressione	72
4.3.1. Approccio classico	72
4.3.2. Approccio controfattuale	77
4.3.3. Applicazione	84
4.4. Mediatori multipli	90
4.4.1. Metodo basato sulla regressione	90
4.4.2. <i>Inverse Probability Weighting</i>	93
4.5. <i>Time – varying exposure and mediators</i>	100
4.5.1. Il <i>g – computational approach</i> e i <i>marginal structural models</i>	100
4.5.2. Applicazione	110
5. Analisi trasversali	115
5.1. Relazioni tra variabili e <i>selection bias</i>	119
5.2. L’effetto di mediazione senza covariate	123
5.2.1. Risultati	124
5.3. L’effetto di mediazione con covariate	136
5.3.1. Risultati	140
6. Analisi longitudinali	153
6.1. Modello ipotizzato	154
6.1.1. Risultati	161

Conclusioni	167
Bibliografia	173
Appendice A	187
Appendice B	189
Appendice C	194
Appendice D	195
Appendice E	198
Appendice F	201
Appendice G	214

INTRODUZIONE

Lo stato di salute della popolazione è un argomento ampiamente trattato in molti ambiti che variano da quello medico a quello psicologico, a quello politico ed economico.

L'aspettativa di vita alla nascita in Europa (28 paesi) nel 2016¹ era di 84 anni per le donne e 78 anni per gli uomini. In generale, nelle società moderne, si sta assistendo a una decrescita dei tassi di natalità che di conseguenza, con il prolungarsi dell'aspettativa di vita alla nascita, determinano un aumento nella quota di popolazione anziana. L'Organizzazione Mondiale della Sanità afferma che con l'aumento della quota di popolazione anziana, si dovranno affrontare molte nuove sfide sociali, politiche ed economiche (OMS, 2018).

Si prevede che entro il 2025 una percentuale superiore al 20% dei cittadini europei avrà più di 65 anni, con un rapido incremento specialmente nel numero di adulti con più di 80 anni (Hirst, 2015).

Lo sviluppo delle società odierne riflette la trasformazione nella struttura per età della popolazione Europea. I conseguenti cambiamenti sociali e la riduzione della forza lavoro rappresentano un punto focale del discorso che riguarda lo sviluppo futuro della società.

Per poter affrontare al meglio le difficoltà a cui tutte le società dovranno fare fronte, risulta fondamentale lo studio e l'analisi del legame tra la salute dei soggetti e il loro stato socio-economico, in modo da poter valutare in che misura è garantito l'accesso ai servizi in una società nella quale la quota di persone anziane diventa sempre più alta.

L'obiettivo della presente analisi è di valutare il possibile ruolo di mediatore svolto dalla salute fisica nella relazione tra lo stato socio-economico e il benessere degli individui. In questo modo sarà possibile determinare l'importanza di garantire un invecchiamento in buona salute dei soggetti, al fine di favorire un alto stato di benessere. Inoltre, si potrà analizzare in che

¹ Fonte: EUROSTAT (2016)

misura la ricchezza può aiutare a migliorare la condizione di benessere e di salute fisica degli individui.

A tale scopo, sono stati esaminati i dati di persone con più di 50 anni sia in ottica trasversale, sia da un punto di vista longitudinale, considerando l'evoluzione del campione in un periodo temporale che va dal 2004 al 2015.

Nello specifico, nel Capitolo 1 si propone un excursus di come sono stati interpretati i concetti di stato socio-economico, benessere e salute fisica nella storia, di come vengono trattati in letteratura e infine dei risultati degli studi che hanno analizzato le relazioni tra le tre variabili.

Nel Capitolo 2 si presenta la fonte dei dati che verranno utilizzati per le analisi: l'Indagine su Salute, Invecchiamento e Pensioni in Europa (SHARE). Sarà fornita una breve descrizione della struttura dell'indagine, dei criteri di eleggibilità e di raccolta dati utilizzati, insieme ad un riassunto dei temi trattati, delle rilevazioni svolte e dell'aderenza dei paesi alle diverse wave di rilevazione.

Il Capitolo 3 è focalizzato sul campione analizzato: dopo una valutazione descrittiva dei soggetti nelle diverse wave, sono introdotte le modalità seguite per calcolare le variabili relative al SES, al benessere e alla salute fisica degli individui; nell'ultima parte del capitolo sono analizzate le relazioni tra le variabili chiave e i possibili confondenti che possono influenzare le relazioni ipotizzate.

Il Capitolo 4 è il capitolo teorico della tesi e si concentra sulla definizione dei metodi utilizzati per svolgere analisi di mediazione seguendo l'approccio controfattuale: dopo aver fornito le prime definizioni e aver presentato i metodi classici per svolgere analisi di mediazione, sarà possibile ampliare tali metodi all'ambito del controfattuale per comprendere al meglio le potenzialità di tale metodo. Poiché si è interessati ad una valutazione longitudinale del fenomeno, sono introdotti anche i nuovi approcci che permettono di considerare situazioni in cui l'esposizione e il mediatore variano nel tempo, utilizzando come passaggio intermedio l'analisi di mediazione in presenza di più di un mediatore.

Nel Capitolo 5 sono presentate le modalità di applicazione dei metodi e i risultati delle analisi di mediazione svolte sulle singole wave di rilevazione, sia in assenza sia in presenza della correzione dall'effetto dei confondenti. Per ampliare l'analisi e ottenere una valutazione longitudinale, le analisi sono ripetute su due tipologie di campioni: il campione completo di soggetti che partecipa ad ogni wave e il campione relativo al panel, cioè che comprende solamente i soggetti che partecipano a tutte le interviste.

Nel Capitolo 6 si affronta l'applicazione dei metodi per svolgere analisi di mediazione con dati longitudinali e vengono riportati i risultati dei modelli con le relative stime degli effetti.

Infine sono tratte le conclusioni, riassumendo brevemente il lavoro svolto e i risultati ottenuti.

Tutte le analisi sono state svolte mediante l'utilizzo del software statistico SAS Studio.

1. CONCETTI CHIAVE

Le tre dimensioni delle quali si è studiata la natura e la tipologia delle relazioni sono lo stato socio-economico, il benessere e la salute fisica.

Considerando di avere a che fare con una popolazione ultra cinquantenne, i tre campi di studio risultano essere ancora più interessanti da analizzare.

La letteratura in questo campo è molto vasta e copre moltissime situazioni particolari, a seconda delle quali si modifica la visione con cui si interpretano e intendono i concetti chiave. Per tale motivo appare fondamentale presentare un excursus di come lo stato socio-economico, il benessere e la salute fisica vengono trattate in letteratura, i vantaggi e gli svantaggi delle modalità con cui sono intese e come si è deciso di considerarle nella presente analisi.

1.1 Stato socio – economico

Lo stato socio-economico (SES) della popolazione ha acquisito sempre più importanza nello studio dei determinanti di salute degli individui. Si tratta di un concetto molto ampio che può assumere connotazioni molto diverse a seconda dello studio a cui si fa riferimento e del particolare legame che si è interessati a studiare.

Si tratta di una misura globale che, focalizzandosi sul livello di istruzione, sui guadagni e sull'occupazione dei soggetti, combina aspetti economici e sociologici per restituire una rappresentazione della posizione sociale di un individuo e della sua famiglia rispetto a quella degli altri.

Mueller e Parcel (1981) hanno definito lo stato socio-economico come "*the relative position of a family or individual on a hierarchical social structure, based on their access to or control over wealth, prestige and power*". Più recentemente, Miech e Hauser (2001), lo definiscono come "*a broad concept that refers to the placement of persons, families, households and census tracts or other aggregates with respect to the capacity to create or consume goods that are valued in our society*".

Ci sono molti modi e prospettive con cui descrivere il SES di un individuo, ma nell'ambito del suo relazionarsi con gli outcome di salute, si fa sempre riferimento ad esso come al tentativo di individuare le possibilità di accesso alle risorse minime che hanno i singoli individui o gruppi di essi per raggiungere e mantenere una buona condizione di salute.

Come mostrato da Vickie L. Shavers (2007), molto spesso gli studi svolti per studiare il SES come determinante di salute hanno portato a risultati inconsistenti a causa di diverse problematiche: la mancanza di una misura precisa e affidabile globalmente accettata; la difficoltà che spesso si riscontra nel raccogliere i dati poiché gli individui, quando si tratta di quantificare il loro reddito o la propria ricchezza, tendono a non rispondere; il fatto che, a seconda della misura scelta, la dinamica che questa assume nel corso della vita degli individui tende a variare molto.

Il problema principale, come si evince da quanto appena mostrato e come si era già detto in precedenza, è la difficoltà che si ha nell'identificare una misura univoca in grado di cogliere nella maniera più comprensiva possibile un costrutto latente come lo stato socio-economico (Oakes e Kaufman, 2017). Le misure e gli indici utilizzati variano tantissimo a seconda del contesto dello studio, della popolazione di riferimento e della relazione del SES con le altre variabili a cui si è interessati. Con il variare di queste ultime, infatti, si può decidere di considerare il livello di istruzione, piuttosto che una misura di reddito, l'occupazione o ancora la ricchezza come misure proxy. In questo modo si fa riferimento a indici monodimensionali ma, come spiega Oakes (2017), non esistono solo questi. In alternativa, si potrebbe decidere di utilizzare una misura composita che riesce a conciliare più dimensioni singole e quindi a dare una visione più globale del SES di un individuo o popolazione. L'utilizzo di una misura complessa potrebbe tuttavia presentare lo svantaggio di non essere riconosciuta universalmente oppure di non avere una chiara interpretazione.

Di seguito verranno presentate per prime le misure univariate e successivamente gli indici complessi, in modo da dare una presentazione complessiva di entrambi.

1.1.1 Misure monodimensionali

Uno degli indicatori di SES più comunemente usati è il reddito di un individuo, oppure una misura di reddito equivalente. Questo permette di esprimere la possibilità che ogni soggetto ha di reperire le risorse desiderate e mette in evidenza i fattori che possono influenzare la salute di ciascuno di essi. Non si tratta di una misura perfetta, ma può risultare relativamente facile da calcolare se non fosse per gli alti tassi di non risposta che spesso vengono riscontrati quando si pongono domande a riguardo. Inoltre gli individui hanno la tendenza a fornire livelli più alti o più bassi del reddito reale percepito. Si tratta di una misura che presenta anche altre limitazioni, tra cui l'alta dipendenza dall'età del soggetto a cui segue un'instabilità di fondo, e il fatto di non essere una misura omnicomprensiva come può invece essere la ricchezza.

Proprio la ricchezza è una variabile proxy abbastanza utilizzata in letteratura che, a differenza della precedente, riesce a coprire un insieme di risorse che tendono ad accumularsi nel tempo. In generale si tratta tendenzialmente di un'eccellente misura per lo stato socio-economico, nonostante anch'essa sia difficile da misurare correttamente a causa degli errori nel riportare le informazioni e della complessità del concetto; inoltre è altamente dipendente dalle caratteristiche individuali dei soggetti analizzati, come l'etnia o il genere.

Rispetto al reddito, la ricchezza è legata in maniera molto più stretta alla classe sociale e i beni posseduti possono dare una rappresentazione migliore della capacità di un individuo di far fronte ad uno shock economico.

Anche il livello di istruzione, per i soggetti con più di 25 anni, può risultare un'ottima misura monodimensionale. Risulta facile da misurare e in molti studi (Barrow e Rouse, 2005; Carnevale, et al., 2013; Shankar, et al., 2013;

Tamborini, et al., 2015) è stato dimostrato che ad alti livelli di istruzione corrispondono molto spesso lavori migliori, situazioni abitative e lavorative migliori e redditi più alti. Il problema maggiore è la non universalità di questo indicatore, data dal fatto che paesi differenti presentano diversi sistemi scolastici difficilmente confrontabili. Inoltre, il ritorno che si ha dal punto di vista economico può variare molto a seconda del genere e dei gruppi etnici a cui si fa riferimento. Spesso poi non si riscontra un collegamento diretto tra gli anni di istruzione e la crescita dello stato socio-economico.

Il livello di istruzione può essere identificato in due modi: come anni di istruzione oppure come livello conseguito. Il livello di istruzione tende ad essere più problematico da utilizzare nel momento in cui si considerano all'interno della ricerca più paesi messi a confronto.

L'ultimo indicatore unidimensionale che verrà affrontato è l'occupazione dei soggetti. Sicuramente questo indicatore presenta un legame molto forte con il livello di istruzione e risulta essere meno influenzato dai tassi di non risposta rispetto al reddito.

I problemi maggiori legati alla scelta dell'occupazione come indice per valutare il SES riguardano: la mancata precisione quando viene misurata e la difficoltà a classificare i soggetti pensionati e le casalinghe.

In generale, uno dei problemi principali dato dall'utilizzo di una misura proxy unidimensionale è che spesso non riesce a cogliere nella sua interezza un concetto complesso e ampio come quello dello stato socio-economico.

1.1.2 Misure composite

Sono state costruite numerose misure composte per la valutazione dello stato socio-economico e in generale tutte presentano il vantaggio di riuscire a quantificare il SES in maniera più approfondita, integrando misure individuali con misure più generali, a livello di area o paese. Il problema maggiore risiede nel fatto che la loro costruzione richiede assunzioni teoriche forti, che permettano di pesare in modo corretto le dimensioni che entrano a far parte di ogni indice.

Uno degli indici prevalentemente usato nella ricerca negli Stati Uniti è il Duncan Socioeconomic Index (SEI) (Duncan, 1961; Caston, 1989), calcolato con dati provenienti dai censimenti. Si tratta di un indice con valori su scala continua che si basa su informazioni relative all'istruzione, al reddito e all'occupazione dei soggetti. Viene fatta una sorta di classifica sulla base del prestigio dell'occupazione svolta. La logica che sta alla base della costruzione dell'indice è quella di vedere l'istruzione come un prerequisito per la posizione occupazionale di cui il reddito è il premio finale.

Il problema principale dell'utilizzo del SEI è dato dal fatto che non ci sono regole precise sulla modalità con cui creare la classifica e le tipologie di occupazione sono moltissime, inoltre sono presenti gruppi come ad esempio gli studenti, coloro che prestano servizio militare, le casalinghe e i pensionati che risultano difficili da trattare.

Nam e Powers (1982; 1983; 1986) hanno sviluppato una misura simile che prende il nome di Nam-Powers Occupational Status Score (OSS). Questa si distingue dalla precedente in quanto istruzione e reddito sono utilizzati per creare una quantità unica.

Oakes (2017) sottolinea però che nella realtà odierna i cambiamenti nell'occupazione sono molto più frequenti e quindi un indice che si basa solamente sul prestigio occupazionale potrebbe non essere tra i migliori da utilizzare per la rappresentazione dello stato socio-economico.

Potrebbe essere interessante considerare il nucleo familiare come unità di primo livello e allora utilizzare un indice come l'*household prestige scale* (HHP) (Rossi, et al., 1974) basato su metodi di regressione.

Esistono anche altri indici tra cui the British Cambridge Social Interaction and Stratification Scale (CAMSIS) (Stewart, et al., 1973; 1980) oppure la National Statistics Socioeconomics classification (NS-SEC) (Rose e Pevalin, 2003) inglese.

La CAMSIS si basa sull'idea che soggetti con posizione sociale simile interagiscono più facilmente tra di loro rispetto a soggetti in altri gruppi sociali e, analizzando l'interazione tra gli individui all'interno di questi

gruppi, si può ottenere una rappresentazione di come le risorse sono distribuite all'interno della società. La scala è costruita sulla base di misure di similarità e differenza tra l'occupazione dei soggetti e si propone di valutare quali strutture sociali emergono, studiando le interazioni tra le persone all'interno di una popolazione.

La NE-SEC, invece, è stata costruita per misurare i rapporti di lavoro e le condizioni di occupazione, con lo scopo di individuare la struttura socio-economica delle società moderne e di spiegare variazioni nel comportamento sociale degli individui.

1.1.3 Ricchezza

La ricchezza è una delle dimensioni chiave dell'analisi, quindi appare opportuno approfondire questo concetto anche da un punto di vista storico, in quanto i dati che si analizzeranno in seguito coprono un periodo di 12 anni all'interno del quale, a causa della crisi economica, sono avvenuti moltissimi cambiamenti dal punto di vista dello stato socio-economico e nello specifico anche della ricchezza globale.

Nei secoli il concetto di ricchezza è cambiato molto, basti pensare a periodi storici come il 1700, quando la ricchezza comprendeva principalmente eredità e possedimenti territoriali. Da quel periodo ad oggi, grazie alla rivoluzione industriale e ai cambiamenti sociali, molte persone hanno potuto accrescere il loro stato socio-economico e sono arrivate a posizioni di prestigio maggiore, riducendo le disuguaglianze che nei secoli erano state mantenute e diminuendo lo stato di povertà in cui versava la maggioranza della popolazione.

Il concetto di ricchezza si è modificato nei secoli, le disuguaglianze sono diminuite, gli standard di benessere sono aumentati rispetto a 300 anni fa, ma la maggior parte della ricchezza mondiale continua a rimanere nelle mani di poche persone. Per questo motivo, analizzando un periodo di crisi, è importante esaminare il legame tra ricchezza e benessere. Molti studiosi hanno infatti evidenziato come la crisi, pur avendo colpito in maniera simile

ogni livello sociale, ha visto le persone con un patrimonio maggiore impiegare molto meno tempo e molta meno fatica a “rialzarsi” o addirittura incrementare il proprio patrimonio, mentre la maggior parte degli individui dei ceti medi o bassi ha dovuto affrontare difficoltà molto maggiori (Lansley, 2012; Story, 2010).

Questo ha fatto sì che l'1% della popolazione detenga più del 50% della ricchezza mondiale, come riporta Jill Treanor in un articolo del 2015 pubblicato sul *The Guardian*, e che le disuguaglianze siano diventate sempre maggiori.

La ricchezza viene misurata in maniera differente a seconda del paese oggetto di analisi, ma in generale si può far riferimento ad essa come al valore complessivo di tutte le risorse di cui una società o un individuo dispongono, cioè il valore complessivo di tutto quello che una persona possiede e che può essere utilizzato per ottenere denaro, beni e servizi.

1.2 Benessere

La seconda dimensione di cui è stato approfondito il significato e i legami con le altre variabili è il benessere. Si tratta di un concetto molto ampio, che comprende diversi aspetti dell'essere umano e ne caratterizza la qualità della vita all'interno di una società. Spesso si è portati a pensare al benessere da un punto di vista prettamente biologico, ma è importante ricordare che in realtà ha a che fare con anche aspetti psichici, mentali e sociali della vita di tutti gli individui.

Il benessere (*wellbeing*) è un aspetto fondamentale dello stato di salute di ogni essere umano. Questo viene confermato anche dalla definizione di salute proposta dall'Organizzazione Mondiale della Sanità: “*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*” (OMS, 1946).

Nei secoli molti filosofi e sociologi hanno dedicato i loro studi alla comprensione del significato di benessere e di cosa comportasse questo per la società e per i soggetti che la compongono. Si tratta di un concetto oggetto di

riflessioni fin dall'antica Grecia, trattato da autori come Aristotele e arrivando alla filosofia moderna, affrontato da autori come Jeremy Bentham e John Stuart Mill (Brey, 2012).

All'interno della filosofia contemporanea sono riconosciute tre teorie principali riguardanti il concetto di benessere (Crisp, 2017):

- **Edonismo.** Questa teoria sostiene che solamente il piacere è intrinsecamente buono e che il dolore è intrinsecamente cattivo. Di conseguenza, la vita di una persona è positiva fino a quando riesce ad accumulare piaceri e a limitare il dolore. Non potendo eliminare completamente il dolore, il raggiungimento del benessere è considerato come il raggiungimento dell'equilibrio tra piacere e dolore nella propria vita.

Le prime linee di questa teoria possono risalire al terzo/quarto secolo a.C. e al pensiero di Epicuro, secondo cui è possibile vivere una buona vita massimizzando il piacere ed evitando le paure prolungate e le sofferenze del corpo.

- **Soddisfacimento del desiderio.** Secondo queste teorie il benessere risiede nel soddisfacimento dei desideri che ogni individuo possiede. Questa branca nasce dal pensiero di economisti che cercano un modo oggettivo per misurare il benessere degli individui e, non potendo quantificare con semplicità il piacere e il dolore, fa ricorso all'identificazione dei desideri individuali. Questo poiché generalmente le persone sono in grado di ordinare i loro desideri e confrontarli con quelli degli altri.
- **Objective list.** Sono così definite le teorie che sostengono che il benessere sia il risultato di una serie di condizioni oggettive delle persone piuttosto che di esperienze personali di piacere o soddisfacimento dei desideri. Si ritiene che ci siano dei beni in grado di aumentare il benessere anche senza creare piacere o desiderio negli individui. Il benessere è vivere una vita in cui i soggetti sono in possesso di tutti o della maggior parte dei beni nella lista.

Sono tre linee di pensiero che non si è ancora riusciti ad unificare, ma in ogni caso fanno emergere diversi aspetti importanti di quello che comunemente viene inteso come benessere: l'aspetto soggettivo che lo caratterizza e che è possibile ritrovare nelle esperienze personali di ogni individuo; la positività dell'emozione che si prova e l'equilibrio che deve essere ricercato tra emozioni negative e positive; il fatto che ricopre diversi ambiti della vita di ogni individuo, anche oggettivi e non solamente soggettivi.

Da un punto di vista psicologico è iniziato ad emergere interesse verso lo stato di salute mentale e di benessere degli individui a partire dal 1950 circa.

Diener e Suh (1997) sostengono che *“subjective well-being consists of three interrelated components: life satisfaction, pleasant affect, and unpleasant affect. Affect refers to pleasant and unpleasant moods and emotions, whereas life satisfaction refers to a cognitive sense of satisfaction with life”*.

L'aspetto psicologico e di salute mentale non può quindi essere escluso dal discorso che riguarda il benessere, come sottolinea anche l'Organizzazione Mondiale della Sanità quando sostiene che *“mental health is defined as a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community”* (OMS, 2006).

La scelta di una misura dello stato di benessere degli individui rimane molto difficile, ma la definizione fornita dall'Organizzazione Mondiale della Sanità racchiude la prospettiva secondo cui verrà valutato e inteso da qui in poi.

Esistono numerosi modi per misurare il benessere e sono state create apposite scale per fare questo e considerare come meglio possibile aspetti soggettivi e psicologici che lo caratterizzano.

1.3 Salute fisica

Come è già stato sottolineato all'inizio del paragrafo precedente, la salute degli individui è da sempre al centro di discorsi molto accesi. Seguendo la definizione di salute fornita dall'Organizzazione Mondiale della Sanità

(1946), emerge subito la complessità del discorso: se in precedenza non si poteva prescindere dal trattare il benessere psicologico degli individui, allo stesso modo adesso non si può escludere il loro stato di salute fisica. I due concetti sono profondamente interconnessi e trattarne uno escludendo l'altro potrebbe portare a conclusioni fuorvianti.

La salute può essere considerata sotto diversi punti di vista: può essere riferita alla struttura e al funzionamento del corpo oppure alla presenza o assenza di patologie; oppure ai sintomi e a quello che può essere fatto per minimizzarli facendo riferimento all'estensione con cui la condizione di salute influenza le attività di vita quotidiana dell'individuo e quindi la qualità della sua vita.

Con la prima parte si fa riferimento allo stato di salute da un punto di vista prevalentemente clinico, che può essere valutato tramite test e misurazioni svolte attraverso la strumentazione medica. La seconda parte fa invece maggiormente riferimento alla percezione degli individui e seppur mantenendo una nota oggettiva, valuta le difficoltà da loro riscontrate nella vita.

Solamente da questa prima e semplice divisione emerge come siano veramente numerosi i modi per valutare lo stato di salute di una popolazione. Appare opportuno sottolineare però che in questa tesi sono analizzati i dati di soggetti prevalentemente ultracinquantenni che vengono seguiti per 12 anni. È importante quindi, al fine di valutare lo stato di salute, considerare alcune variabili che riflettono la salute in età più avanzata degli individui e che permettono di valutarne il processo di invecchiamento.

È altresì importante che nonostante l'invecchiamento sia un processo inevitabile, venga garantita una vita serena e in una condizione di benessere quanto maggiore possibile.

Generalmente la maggior parte delle problematiche e delle situazioni di disagio psicologico in età avanzata sono dovute alla presenza e all'insorgenza di "disabilità" che impediscono ai soggetti di svolgere le normali attività di vita quotidiana.

L'Organizzazione Mondiale della Sanità definisce la disabilità nel seguente modo: *“Disabilities is an umbrella term, covering impairments, activity limitations, and participation restrictions. An impairment is a problem in body function or structure; an activity limitation is a difficulty encountered by an individual in executing a task or action; while a participation restriction is a problem experienced by an individual in involvement in life situations. Disability is thus not just a health problem. It is a complex phenomenon, reflecting the interaction between features of a person’s body and features of the society in which he or she lives.”* (OMS, 2016)

Emerge subito con chiarezza come si tratti di un fenomeno complesso, che comprende numerosi aspetti e che non rappresenta solo un problema di salute ma anche un problema di diritti umani e di sviluppo, che va preso in considerazione con molta attenzione. Si tratta di un evento che colpisce più di un miliardo di persone al mondo e che ha richiesto numerosi anni affinché venisse definito e fossero chiariti i metodi e gli strumenti con cui misurarlo. L'OMS nel 1980 ha stilato la classificazione ICIDH (*International Classification of Impairments Disabilities and Handicaps*) che permette di distinguere tra:

- *Menomazione*, con cui si intende qualsiasi perturbazione mentale o fisica del funzionamento del corpo. Essa è caratterizzata da perdite o anomalia anatomica, psicologica o fisiologica (a livello di un tessuto, di un organo, di un sistema funzionale o di una singola funzione del corpo);
- *Disabilità*, con cui si indica la limitazione o perdita (conseguente alla menomazione) delle capacità di effettuare un'attività nel modo o nei limiti considerati normali per un essere umano;
- *Handicap*, con cui si fa riferimento alla situazione di svantaggio sociale, conseguente a menomazione e/o disabilità che limita o impedisce l'adempimento di un ruolo sociale considerato normale in funzione di età, sesso e fattori culturali e sociali.

Questa classificazione prevede una sequenza che da menomazione passa attraverso la disabilità per poi arrivare a handicap, anche se il passaggio non risulta essere sempre automatico in quanto la menomazione può essere causa diretta di handicap senza passare attraverso il mediatore della disabilità (Nota, et al., 2002). Sono presenti inoltre altre limitazioni in questa classificazione, tra cui il fatto che vengano considerati solamente fattori patologici ma non ambientali, fondamentali per la determinazione delle limitazioni, che non venga considerata la possibilità che la disabilità sia solamente temporanea, oppure le numerose difficoltà che si riscontrano nel definire il livello a cui una persona viene considerata disabile oppure no.

A causa delle numerose problematiche riscontrate nella prima classificazione, nel 1999 l'Organizzazione Mondiale della Sanità propone una nuova versione della classificazione, dandole il nome di *International Classification of Impairments, Activities and Participation* (ICIDH-2). Vengono così ridefiniti due dei tre concetti di base intorno a cui era stata costruita l'ICIDH; in questo modo i termini disabilità e partecipazione sono definitivamente messi da parte e sostituiti con termini che permettono di essere meno discriminatori nel definire i soggetti. La nuova formulazione consente inoltre di vedere la classificazione in una prospettiva più sociale, includendo anche i fattori ambientali e di contesto, cruciali nella definizione della disabilità, considerata come risultato dell'interazione tra questi aspetti e lo stato di salute dei soggetti, rimuovendo il nesso causale implicito presente nella prima classificazione tra le tre dimensioni di interesse. Il risultato è una classificazione molto più completa e articolata, ma anche più complessa.

A quest'ultima classificazione l'OMS ne sostituisce un'altra nel 2001, la ICF (*International Classification of Functioning, Disability and Health*), una classificazione innovativa approvata da tutte le nazioni che fanno parte dell'ONU e che permette di cambiare il concetto di disabilità secondo la definizione precedentemente data. In quest'ultima classificazione vengono presi in considerazione non solo fattori riguardanti lo stato di salute dal punto di vista medico, ma anche dell'interazione sociale dei soggetti, dando così

un'accezione molto più ampia del concetto di disabilità. Alla prospettiva centrata puramente sulla patologia della ICIDH ne subentra una multidimensionale che considera una bipartizione iniziale tra categorie riguardanti il funzionamento e la disabilità, cioè lo stato di salute dell'individuo, e fattori contestuali, cioè ambientali e personali. Cambia l'ottica con cui vengono guardati i soggetti, infatti, prima l'attenzione veniva focalizzata su ciò che caratterizzava la menomazione dell'individuo, mentre adesso su quelle abilità residue che ogni soggetto presenta. La disabilità non è più solamente considerata con una visione unidimensionale, ma ha un'accezione dinamica che comprende stati patologici cronici insieme a fattori ambientali e sociali. In questo modo si propone un modello di disabilità universale, applicabile a qualsiasi persona, normodotata o diversamente abile.

L'ISTAT per rilevare la presenza di disabilità utilizza una batteria di quesiti identificati da un gruppo di lavoro dell'OECD sulla base della classificazione ICIDH dell'OMS, cambiandone tuttavia la denominazione e facendo riferimento alla *rilevazione della presenza di limitazioni funzionali* per seguire la definizione di disabilità data dalla classificazione ICF.

In questo modo è possibile identificare diverse tipologie di limitazioni funzionali:

- *Confinamento*, con cui “si intende costrizione permanente a letto, su una sedia, o nella propria abitazione per motivi fisici o psichici”;
- *Difficoltà nel movimento*, con cui si indicano soggetti che “hanno problemi nel camminare (riescono solo a fare qualche passo senza aver bisogno di fare soste), non sono in grado di salire e scendere da soli una rampa di scale senza fermarsi, non riescono a chinarsi per raccogliere oggetti da terra”;
- *Difficoltà nelle funzioni della vita quotidiana*, cioè che “riguardano la completa assenza di autonomia nello svolgimento delle essenziali attività quotidiane o di cura della persona, quali mettersi a letto o

sedersi da soli, vestirsi da soli, lavarsi o farsi il bagno o la doccia da soli, mangiare da soli anche tagliando il cibo”;

- *Difficoltà nella comunicazione*, cioè “le limitazioni nel sentire (non riuscire a seguire una trasmissione televisiva anche alzando il volume e nonostante l’uso di apparecchi acustici); limitazioni nel vedere (non riconoscere un amico ad un metro di distanza); difficoltà nella parola (non essere in grado di parlare senza difficoltà)”.

Emerge subito la difficoltà nell’identificare la maniera più opportuna per definire e valutare lo stato di salute caratterizzato dalla presenza di disabilità.

Alcuni autori hanno deciso di valutare lo stato di salute degli individui facendo riferimento a degli elenchi di attività che ognuno si trova a dover svolgere nella quotidianità della vita e la disabilità è identificata sulla base del numero di limitazioni che vengono riscontrate oppure semplicemente dalla presenza di almeno una limitazione (Jagger, et al., 2009).

Esistono molti altri modi per valutare le limitazioni di un soggetto, ad esempio lo United States Census Bureau identifica quattro aspetti e in presenza di almeno uno di questi è assegnato lo stato di disabilità (U.S. Census Bureau, 2005):

- Utilizzo di un supporto per la mobilità (bastone, stampelle o sedia a rotelle);
- Difficoltà nello svolgere attività di vita quotidiana (ADL), attività strumentali di vita quotidiana (IADL) oppure attività funzionali specifiche;
- Presenza di una o più menomazioni;
- Impossibilità di svolgere lavori domestici o impegni lavorativi.

In altri studi, invece, viene valutato lo stato di salute sulla base della forza di presa della mano, ma si tratta di un indicatore che riporta numerose problematiche tra cui la discrepanza tra le modalità e gli strumenti di misura utilizzati dai diversi paesi (Hairi, et al., 2009; Andersen-Ranberg, et al., 2009).

La scelta di come considerare e misurare lo stato di salute resta comunque strettamente legato alla tipologia di analisi che si intende svolgere.

1.4 La relazione tra le variabili in letteratura

Molti paesi sviluppati hanno generalmente accettato i 65 anni di età come definizione di una persona anziana (OMS, 2009). Oltre all'età anagrafica, l'età degli individui può essere determinata esaminando processi biologici, psicologici e socio-culturali (Cohen, 2002) e questa molteplicità di punti di vista rende ancora più evidente il legame tra l'età di una persona e il benessere della stessa, sia da un punto di vista psicologico che fisico, soprattutto quest'ultimo legato all'inevitabile processo di invecchiamento a cui tutti sono sottoposti. Nelle società moderne la speranza di vita alla nascita² è molto alta, mentre il tasso di natalità si abbassa, determinando un aumento nella quota di popolazione anziana. L'Organizzazione Mondiale della Sanità afferma che con la crescita della popolazione anziana, si devono affrontare molte nuove sfide sociali, politiche ed economiche (OMS, 2002; 2018).

Nell'analisi dell'impatto che le crisi economiche succedutesi nella storia hanno avuto sulla salute degli individui si identificano *pattern* differenti (Bacigalupe e Escolar-Pujolar, 2014). Emerge che con l'aggravarsi delle condizioni economiche, crescono le problematiche di salute fisica, quelle di salute mentale e la mortalità (Catalano, et al., 2011). Tuttavia è stato mostrato da altri studi (Edwards, 2008) che la mortalità tende ad aumentare anche in periodi di espansione economica e a diminuire durante quelli di recessione. Questo potrebbe essere giustificato dal fatto che durante le recessioni migliora lo stato di salute dei soggetti che rimangono nel mercato del lavoro, poiché tendono ad esporsi meno a comportamenti a rischio rispetto a coloro che non hanno un impiego. Inoltre alcune cause di morte potrebbero

² L'aspettativa di vita alla nascita in Europa (28 paesi) è nel 2016 di 84 anni per le donne e 78 anni per gli uomini. (Fonte: EUROSTAT)

lentamente sparire, e questo in linea teorica compenserebbe il peggioramento delle condizioni di salute negli altri gruppi sociali (Catalano e Bellows, 2005). In ogni caso il ruolo delle politiche di welfare durante i periodi di crisi economica risulta essere cruciale nella gestione e nel contenimento della crescita delle disuguaglianze (Lahelma, et al., 1997).

Secondo l'Organizzazione Mondiale della Sanità (2013), in tutti i paesi la crisi economica del 2008 ha avuto un effetto negativo sui sistemi sanitari, provocando un calo nella disponibilità di risorse e un aumento della richiesta di tali servizi. È stato mostrato (Toffolutti e Suhrcke, 2014) che in seguito al rapido deterioramento delle condizioni economiche in numerosi paesi sono aumentati anche i tassi di suicidio (USA: Reeves, et al., 2012; UK: Barr, et al., 2012; Italia e Grecia: De Vogli, et al., 2013).

Quanto appena detto assume ancora più valore nell'analisi che viene qui riportata, poiché in seguito sarà analizzato un periodo temporale che va dal 2004 al 2015, che quindi comprende tutte le difficoltà che derivano dalla crisi economica che ha colpito l'Italia e l'Europa nel 2008.

Analizzando la relazione tra stato socio-economico e benessere in letteratura appaiono pareri contrastanti: da un lato è possibile individuare una relazione positiva per cui a livelli socio-economici più alti corrispondono condizioni di benessere psicologico migliori (Kaplan, et al., 2008; Blustein, 2008); dall'altro invece, viene posta maggior enfasi sul fatto che coloro con livelli di SES più alti ricoprono spesso ruoli lavorativi con responsabilità maggiore che portano con sé alti livelli di stress e un conseguente benessere psicologico inferiore (Wadsworth, et al., 2010; La Torre, et al., 2018).

Anche la relazione tra lo stato socio-economico e la salute fisica risulta essere complicata. Generalmente vengono presentate le società moderne come società del benessere in cui la salute fisica è praticamente garantita per tutti ma, nonostante questo, confrontando paesi differenti e sistemi sanitari differenti, possono emergere delle realtà nelle quali il benessere economico risulta essere associato alla condizione fisica dei soggetti (Newell, 2000; Whitehead, et al., 2001; Barbeau, et al., 2004; Adler, et al., 2016).

Studiando questa relazione è possibile avere una visione globale della realtà dei paesi che si stanno analizzando, valutando se soggetti con uno stato socio-economico più alto sono in grado non solo di ricorrere a cure migliori ma anche a sistemi di supporto psicologico che possono incrementare il loro benessere. Appare quindi sempre più importante la distinzione tra salute fisica e benessere psicologico poiché, seppur parte di un benessere complessivo individuale, riflettono aspetti distinti della vita degli individui soprattutto se analizzati in popolazioni con età più avanzate nelle quali le due variabili risultano essere fortemente interconnesse (Steptoe, et al., 2014).

In questo scenario risulta quindi cruciale indagare la relazione tra stato socio-economico, salute fisica e benessere psicologico degli individui per capire meglio in che modo entrano in relazione le tre dimensioni in un'analisi a livello europeo, nella quale il campione analizzato, composto prevalentemente da soggetti ultracinquantenni, ricopre un ruolo fondamentale.

2. I DATI

I dati utilizzati per svolgere tutte le analisi provengono da SHARE, un acronimo per indicare l'Indagine sulla Salute, l'Invecchiamento e il Pensionamento in Europa (*Survey of Health, Ageing and Retirement in Europe*). Si tratta di un database multidisciplinare e multinazionale che raccoglie dati di tipo panel riguardanti la salute, lo status socio-economico, le famiglie e gli stili di vita dei soggetti che ne prendono parte.

Di seguito si andrà ad analizzare nel dettaglio l'indagine, i criteri di eleggibilità necessari per prenderne parte e le caratteristiche delle diverse wave che sono state svolte.

2.1 L'indagine SHARE

SHARE è in gran parte finanziata dalla Commissione Europea ed è stata fondata nel 2002 con lo scopo di “esaminare la possibilità di stabilire, in collaborazione con gli stati membri, un'indagine longitudinale sull'invecchiamento a livello europeo”.

La rilevanza che ha acquisito a livello europeo viene confermata dal fatto che nel 2006 venne selezionato da parte del European Strategy Forum on Research Infrastructures (ESFRI) come uno dei progetti da implementare e nel 2011 viene riconosciuto come il primo European Research Infrastructure Consortium (SHARE – ERIC).

Il coordinamento è svolto centralmente da parte di Axel Börsch-Supan al Munich Center for the Economics of Aging (MEA), Max Planck Institute for Social Law and Social Policy. Si tratta di una collaborazione che riguarda più di 150 ricercatori in tutto il mondo, organizzati in team multidisciplinari a livello nazionale e gruppi di lavoro multinazionali.

Seppure armonizzandosi e prendendo ispirazione dai due pilastri U.S. Health and Retirement Study (HRS) e English Longitudinal Study of Ageing (ELSA), il primo americano e il secondo inglese, SHARE è riuscito a distinguersi e a

diventare un modello di riferimento per molte altre indagini grazie al suo approccio multidisciplinare e alla sua capacità di cogliere i caratteri dinamici del processo di invecchiamento della popolazione.

2.2 Campionamento e criteri di eleggibilità

La popolazione target dell'indagine sono i soggetti ultra cinquantenni residenti in 27 paesi europei e Israele. Dalla prima intervista nel 2004, fino alla sesta nel 2015 sono stati considerati eleggibili più di 297000 soggetti e più di 120000 rispondenti.

Tra i criteri di esclusione ci sono l'incarcerazione, l'ospedalizzazione, la permanenza fuori dal paese per tutta la durata dell'indagine, l'incapacità di parlare la lingua del paese e il trasferimento ad un indirizzo non noto.

Nella prima wave sono considerati eleggibili tutti i membri delle diverse famiglie nati nel 1954 o prima. A partire dalla seconda wave, per i nuovi paesi che prendono parte all'analisi o per il *refreshment sample*, inserito per far fronte all'*attrition*, cioè alla perdita dei soggetti al follow-up, viene selezionato un solo rispondente con più di 50 anni. Quest'ultimo campione presenta anche l'utilità di arricchire le analisi, rendendo disponibili i dati di soggetti che precedentemente non risultavano eleggibili in quanto troppo giovani. Insieme al *refreshment sample* è possibile trovare il *longitudinal sample* che comprende invece i soggetti che sono già stati intervistati in una wave precedente. In tutte le interviste, oltre che al rispondente viene sottoposto il questionario anche al partner che vive nella stessa abitazione, indipendentemente dalla sua età.

La raccolta dei dati viene svolta mediante l'utilizzo dello strumento CAPI (Computer – Assisted Personal Interviewing), insieme allo svolgimento di altre misurazioni e alla compilazione di questionari cartacei con il supporto di un intervistatore.

Le interviste vengono svolte di persona utilizzando un computer portatile sul quale lo strumento CAPI è installato; questa modalità di rilevazione risulta fondamentale in quanto rende possibile l'esecuzione di test a livello fisico.

2.3 Le wave

Prima di affrontare in modo più dettagliato le caratteristiche delle diverse wave è importante sottolineare come uno degli aspetti fondamentali dell'indagine SHARE sia la sua caratteristica di onnicomprensività. Infatti essa riesce a raccogliere le informazioni su un largo numero di paesi e di conseguenza le caratteristiche dei diversi sistemi di welfare che si caratterizzano per le differenze a livello di assistenza sanitaria, di pubblica istruzione, del mondo lavorativo, della previdenza sociale e dell'accesso alle risorse culturali.

Di seguito viene riportata una tabella di tutti i 27 paesi europei e Israele che hanno preso parte dal 2004 al 2015 alle diverse wave, specificando per ognuno di essi a quali indagini hanno preso parte.

Nelle analisi svolte successivamente verranno considerati solamente i 9 paesi che hanno partecipato a tutte le sei wave, riportati in grassetto nella *Tabella 2.1*.

Paese	Wave					
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
	2004	2007	(SHARELIFE) 2009	2011	2013	2015
Austria	✓	✓	✓	✓	✓	✓
Belgio	✓	✓	✓	✓	✓	✓
Croazia	-	-	-	-	-	✓
Danimarca	✓	✓	✓	✓	✓	✓
Estonia	-	-	-	✓	✓	✓
Francia	✓	✓	✓	✓	✓	✓
Germania	✓	✓	✓	✓	✓	✓
Grecia	✓	✓	✓	-	-	✓
Irlanda	-	✓	✓	-	-	-
Israele	✓	✓	-	-	✓	✓
Italia	✓	✓	✓	✓	✓	✓
Lussemburgo	-	-	-	-	✓	✓
Paesi Bassi	✓	✓	✓	✓	✓	✓
Polonia	-	✓	✓	✓	-	✓
Portogallo	-	-	-	✓	-	✓
Rep. Ceca	-	✓	✓	✓	✓	✓
Slovenia	-	-	-	✓	✓	✓
Spagna	✓	✓	✓	✓	✓	✓
Svezia	✓	✓	✓	✓	✓	✓
Svizzera	✓	✓	✓	✓	✓	✓
Ungheria	-	-	-	✓	-	-

Tabella 2.1 – Aderenza dei paesi alle wave dell'indagine SHARE

2.3.1 Wave 1

La prima wave di SHARE risale al 2004 e ne presero parte 11 paesi.

Le nazioni costituiscono una buona rappresentazione delle varie regioni europee, a partire dalla Scandinavia, con la partecipazione di Danimarca e Svezia, passando poi attraverso l'Europa Centrale, con Austria, Francia, Svizzera, Belgio e Paesi Bassi, arrivando fino all'area mediterranea includendo Spagna, Italia e Grecia.

Israele prende parte a SHARE alla fine del 2004, diventando il primo paese del Medio Oriente ad avviare uno studio sistematico sull'invecchiamento della sua popolazione.

La popolazione intervistata è di 22119 soggetti, ai quali viene sottoposto un questionario principale composto da 20 moduli (*Tabella 2.2*) che coprono numerosi argomenti riguardanti la salute, le condizioni socio-economiche e le relazioni sociali degli individui.

2.3.2 Wave 2

A partire dal 2007, con l'uscita della seconda wave, SHARE inizia ad aprirsi alla possibilità di svolgere studi non solo trasversali ma anche a livello longitudinale, avendo la possibilità di seguire le stesse persone nel tempo e analizzare i loro cambiamenti.

Nella raccolta dei dati, si è cercato di ricontattare tutti i rispondenti della prima wave e in più è stato selezionato anche un *refreshment sample* in tutti i paesi che avevano già partecipato, esclusi Austria e Belgio Fiammingo. I metodi di campionamento rimangono gli stessi, in modo da mantenere lo stesso livello di rappresentatività della popolazione di 50 anni o più.

La Repubblica Ceca e la Polonia entrano a far parte dell'indagine a partire dal 2006/2007 e Israele prenderà parte a questa wave solo nel 2009/2010.

Rispetto alla wave precedente, il questionario consta di 22 moduli (*Tabella 2.3*) e in aggiunta viene svolta un'intervista sulla "fine della vita" per i soggetti deceduti.

Moduli	Contenuti
Modulo iniziale	Informazioni demografiche di base sugli individui conviventi come genere ed età
Informazioni demografiche	Domande riguardanti lo stato civile, il paese di nascita e il livello di istruzione
Figli	Domande riferite ai figli del rispondente, come stato civile, frequenza delle visite e occupazione
Salute fisica	Gli aspetti generali della salute degli individui: la salute percepita, patologie diagnosticate e difficoltà
Stili di vita	Comportamenti legati ad alcol e fumo insieme ad alimentazione e attività fisica svolta
Funzioni cognitive	Misurazioni riguardanti la memoria, la concentrazione, la capacità di fare conti e la fluidità verbale degli individui
Salute mentale	Informazioni riguardanti la salute mentale ed emotiva dell'individuo utilizzate poi per il calcolo delle scale di depressione
Assistenza sanitaria	Si tiene traccia delle visite mediche, dei ricoveri e degli interventi svolti
Occupazione e pensionamento	Attività lavorative svolte, soddisfazione, fonti di guadagno e risorse economiche
Forza della presa	Test fisico per la valutazione della forza di entrambe le mani
Capacità motoria	Test per valutare la velocità della camminata degli individui (valutata solo nelle wave 1 e 2)
Supporto sociale	L'insieme degli aiuti che il rispondente riceve oppure è in grado di fornire a familiari, amici o conoscenti
Trasferimenti finanziari	Tutti i trasferimenti e i pagamenti in entrata e in uscita, comprese le eredità
Situazione abitativa	Informazioni riguardanti la situazione abitativa attuale del rispondente, con anche la grandezza e la qualità della stessa
Reddito familiare	Misure riassuntive di varie tipologie di reddito familiare
Consumi	Spese fatte a livello familiare
Beni e possedimenti	Quantità di beni posseduti, finanziari e non finanziari
Attività	Attività svolte e percezione dell'individuo su vari aspetti della sua vita
Aspettative	Aspettative sul futuro con relativo grado di fiducia che l'individuo ha su di esse, insieme alle modalità con cui vengono prese le decisioni a livello familiare
Osservazioni dell'intervistatore	Valutazione del livello di comprensione delle domande, del luogo in cui è stata svolta l'intervista e della disponibilità del rispondente

Tabella 2.2 – Moduli e contenuti comuni a tutte le wave

Moduli	Contenuti
Forza nelle gambe	Test per valutare la forza nelle gambe mediante il tempo che si riesce a stare in piedi (valutata solo nelle wave 2 e 4)
Test spirometrico	Test per valutare la forza espiratoria dei polmoni (valutata solo nelle wave 2, 4 e 6)
Fine della vita	Cause e circostanze del decesso

Tabella 2.3 – Moduli e contenuti che si aggiungono nella seconda wave

2.3.3 Wave 3 – SHARELIFE

La terza wave è resa disponibile a partire dal 2009 e ha la particolarità di focalizzarsi sulla storia della vita degli individui, intervistando quasi 30000 persone in 13 paesi europei.

Il questionario si modifica rispetto a quelli precedentemente mostrati, cercando di coprire tutte le aree più importanti della vita degli individui in merito a partner e bambini, situazione abitativa e storia lavorativa, con domande dettagliate riguardanti la salute e l'assistenza sanitaria ricevuta. In questo modo è stato possibile costruire un dataset multidisciplinare che copra aspetti sociologici, economici, gerontologici e demografici.

Nella *Tabella 2.4* vengono riportate le informazioni retrospettive che vengono raccolte in questa wave.

La strategia seguita per la raccolta dei dati a livello retrospettivo si basa su quello che viene chiamato *life history calendar* (LHC). Con questo termine si indica il fatto che la vita del rispondente viene rappresentata graficamente mediante l'utilizzo di una griglia che viene riempita durante l'intervista.

L'idea che sta alla base di questa tecnica è il cercare di aiutare il rispondente a rammentare tutti gli eventi, partendo da quelli che è molto probabile vengano ricordati in maniera più accurata (es. nomi e date di nascita dei figli del rispondente).

Moduli	Contenuti
Informazioni demografiche	Informazioni demografiche di base sugli individui conviventi come genere ed età
Infanzia	Domande riguardanti la qualità dell'infanzia del rispondente, l'abitazione e il percorso educativo
Disabilità	Invalità, diminuzione dell'attività lavorativa, congedi concessi e pensioni di invalidità
Storia finanziaria	Fondi, beni e azioni nella durata della vita
Forza della presa	Test fisico per la valutazione della forza di entrambe le mani
Salute	Informazioni riguardanti la salute passata, i ricoveri e la percezione della propria salute
Assistenza sanitaria	Visite mediche, vaccini e trattamenti ricevuti
Storia dei figli	Domande riguardanti le gravidanze, le nascite e le caratteristiche dei figli
Storia lavorativa	Attività lavorative svolte, caratteristiche e fonti di guadagno
Partner	Storia dei partner quindi matrimoni, divorzi e relazioni precedenti
Qualità del lavoro	Condizioni lavorative e impegno richiesto
Storia della vita generale	Percezione della vita con specificazione dei periodi di stress, di oppressione e di difficoltà
Situazione abitativa	Informazioni riguardanti la situazione abitativa passata includendo trasferimenti e spostamenti
Fine della vita	Cause e circostanze del decesso
Osservazioni dell'intervistatore	Valutazione del livello di comprensione delle domande, del luogo in cui è stata svolta l'intervista e della disponibilità del rispondente

Tabella 2.4. Moduli e contenuti della terza wave

2.3.4 Wave 4

Con questa wave si ritorna alle modalità precedenti rispetto alla wave 3.

Si aggiungono all'indagine nel 2010 Estonia, Ungheria e Slovenia.

La particolarità del questionario risiede nell'aggiunta di un modulo dedicato alle reti sociali degli individui, rendendo SHARE la prima indagine internazionale in grado di fornire informazioni dettagliate relative a questo aspetto della vita dei singoli.

2.3.5 Wave 5

A partire dal 2013, con il rilascio della quinta wave, vengono introdotte nuove informazioni mediante l'aggiunta dei moduli riguardanti l'esclusione sociale, le condizioni durante l'infanzia e l'utilizzo di internet.

All'indagine aderiscono Austria, Belgio, Svizzera, Repubblica Ceca, Germania, Danimarca, Estonia, Spagna, Francia, Israele, Italia, Lussemburgo, Paesi Bassi, Svezia e Slovenia.

2.3.6 Wave 6

Nel 2015 vengono resi disponibili i dati relativi alla sesta wave di SHARE, contenenti per la prima volta dopo la wave 4 informazioni relative alla rete sociale degli individui che prendono parte all'analisi.

Per la prima volta entra a far parte dei paesi che partecipano all'indagine anche la Croazia, arrivando così a 18 paesi partecipanti.

2.3.7 Wave 7

Nel 2017 è avvenuta la raccolta delle informazioni per la settima wave in 28 paesi, raggiungendo così una copertura completa dell'Unione Europea a seguito della partecipazione di Finlandia, Lituania, Lettonia, Slovacchia, Romania, Bulgaria, Malta e Cipro.

All'interno del questionario che viene sottoposto ai soggetti intervistati è stato incluso un modulo relativo a SHARELIFE in modo da raccogliere le informazioni retrospettive per tutti i soggetti che non hanno preso parte alla wave 3.

Il rilascio dei dati è previsto per la primavera del 2019, per tale motivo la wave non sarà parte delle analisi che verranno esposte in questa tesi.

3. LE VARIABILI

In questo capitolo sono presentate le caratteristiche dei soggetti che prendono parte all'analisi nelle diverse wave. Inizialmente vengono valutate le distribuzioni di alcune delle variabili principali; successivamente verrà trattata la definizione e la creazione delle variabili relative allo stato socio-economico, alla salute fisica e al benessere degli individui; e infine saranno analizzate le variabili nel loro insieme per avere una comprensione migliore di come queste si relazionano tra di loro.

Si ricorda inoltre che SHARE, oltre ai dati originali ottenuti dalle interviste, mette a disposizione un apposito dataset contenente i dati imputati in modo da risolvere il problema dei valori mancanti. Si tratta di un problema rilevabile specialmente nelle domande relative il reddito e le ricchezze degli individui nelle quali i rispondenti spesso si rifiutano di rispondere e per tale motivo sono presenti alti tassi di valori mancanti.

SHARE utilizza delle apposite tecniche per ottenere una stima generale o un range di valori all'interno del quale collocare ciascun individuo quando gli vengono poste domande a cui non vuole rispondere. Sulla base di queste informazioni sono generati dei dataset con i valori imputati, per cui ai soggetti da cui non è stato possibile avere una risposta, è possibile comunque associare una stima del valore mancante.

Nelle variabili in cui erano disponibili sono stati utilizzati i dati imputati.

3.1 Il campione

Vengono qui riportate le distribuzioni delle variabili coinvolte nell'analisi e il modo in cui sono state rilevate nel questionario utilizzato da SHARE.

3.1.1 Caratteristiche socio – demografiche

Sono considerate eleggibili le persone con più di 50 anni con un domicilio regolare nel paese SHARE corrispondente; in tutte le wave, indipendentemente dalla data di nascita, sono intervistati anche i partner che hanno la stessa residenza.

Nella *Tabella 3.1* è possibile osservare la distribuzione dei soggetti nelle diverse wave di rilevazione. Si ricorda che nelle wave successive alla prima, al campione iniziale si aggiungono degli altri soggetti, *refreshment sample*, per mitigare la perdita di soggetti al follow-up.

Di questi soggetti solamente 6707 hanno partecipato a tutte le interviste.

Wave	Anno	Numerosità
1	2004	22119
2	2007	22381
3	2009	18513
4	2011	33418
5	2013	43491
6	2015	38979

Tabella 3.1 – Numero di soggetti nelle diverse wave

All'indagine SHARE hanno aderito nel corso degli anni 27 paesi europei ed Israele. Nell'analisi sono stati selezionati solamente i nove paesi che partecipano a tutte le wave di rilevazione, per avere una visione più uniforme possibile in tutto il periodo analizzato. Nella *Tabella 3.2* è possibile osservare il numero di soggetti in ciascuno dei paesi considerato.

Wave	Paese									Tutti i paesi
	Austria	Germania	Svezia	Spagna	Italia	Francia	Danimarca	Svizzera	Belgio	
1	1569	2997	3049	2316	2553	3122	1706	997	3810	22119
2	1200	2628	2796	2427	2986	2989	2630	1498	3227	22381
3	999	1921	1961	2271	2528	2500	2144	1324	2865	18513
4	5255	1621	1969	3728	3595	5851	2287	3788	5324	33418
5	4382	5752	4556	6708	4750	4506	4146	3051	5640	43491
6	3402	4412	3906	5636	5313	3948	3733	2806	5823	38979

Tabella 3.2 – Numero di soggetti partecipanti alle wave per paese di provenienza

In tutte la wave vi è una percentuale maggiore di donne rispetto agli uomini (*Tabella 3.3*), legato alla struttura per età della popolazione. Infatti, considerando che le donne hanno una speranza di vita superiore a quella degli uomini³ e che sono inclusi prevalentemente soggetti con più di 50 anni, tra le età più anziane vi è una maggiore presenza di donne.

Wave	Genere				
	Maschi		Femmine		Maschi e femmine
	Frequenza	%	Frequenza	%	Frequenza
1	9843	44.51	12276	55.49	22119
2	9976	44.57	12405	55.43	22381
3	8180	44.19	10333	55.81	18513
4	14750	44.14	18668	55.86	33418
5	19573	45.01	23918	54.99	43491
6	17430	44.72	21549	55.28	38979

Tabella 3.3 – Numero e percentuali di soggetti partecipanti alle wave per genere

I soggetti che partecipano alla rilevazione sono gli ultracinquantenni, ma oltre a loro vengono intervistati anche i partner e, per tale motivo, quando si

³ L'aspettativa di vita alla nascita in Europa (28 paesi) è nel 2016 di 84 anni per le donne e 78 anni per gli uomini. (Fonte: EUROSTAT)

analizza la distribuzione per età delle persone sono presenti soggetti di età inferiore ai 50 anni.

L'età dei soggetti è particolarmente importante anche a fronte della relazione che si vuole analizzare: volendo comprendere in che modo stato socio-economico, salute fisica e benessere si relazionano tra di loro, prendere in considerazione dei soggetti prevalentemente con più di 50 anni e osservandoli per un periodo di 10 anni, implica considerare delle situazioni in cui le condizioni di salute e di benessere, strettamente legate al processo di invecchiamento, possono cambiare con l'età degli individui.

Valore medio, minimo e massimo della distribuzione della variabile relativa all'età in ogni wave sono riportati nella *Tabella 3.4*.

Strettamente legata all'età e al genere delle persone, è possibile osservare la distribuzione della variabile relativa allo stato civile. Nonostante la maggior parte degli individui in tutte le wave risulti essere sposata, la seconda percentuale più alta riguarda i soggetti vedovi, che comprende circa il 13-14% dei soggetti in tutte le wave di rilevazione (*Tabella 3.5*).

Wave	Età		
	Media	Minimo	Massimo
1	64	26	103
2	65	26	105
4	66	28	104
5	67	22	104
6	68	24	106

Tabella 3.4 – Valore medio, minimo e massimo delle distribuzioni della variabile relativa all'età nelle diverse wave

Stato civile	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Sposato/a	15460	69.89	15769	70.46	22936	68.63	30419	69.94	26680	68.45
Unione registrata	366	1.65	399	1.78	537	1.61	653	1.50	627	1.61
Separato/a o divorziato/a	1750	7.91	1795	8.02	3307	9.90	4147	9.54	3890	9.98
Celibe/nubile	1309	5.92	1252	5.59	2156	6.45	2712	6.24	2452	6.29
Vedovo/a	3234	14.62	3166	14.15	4482	13.41	5560	12.78	5330	13.67
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.5 – Numero e percentuali di soggetti partecipanti alle wave per stato civile

3.1.2 Outcome di salute

Tra gli outcome di salute rientrano sia gli aspetti legati alla salute fisica degli individui sia a quella psicologica alla quale nella presente verrà fatto generalmente riferimento come benessere.

Prima di scegliere la misura con cui definire la salute fisica degli individui sono stati valutati diversi outcome di salute.

Uno dei primi indicatori della condizione fisica è legato al numero di limitazioni che ciascuno individua nello svolgimento di attività di vita quotidiana (ADL – *Activities of Daily Living*). Si tratta di un indice proposto nel 1963 da Sidney Katz e viene utilizzato per identificare il numero di limitazioni nelle attività della vita quotidiana. Si fa riferimento alle attività di tutti i giorni che gli individui devono svolgere per prendersi cura di sé stessi e che gli permettono di mantenere la propria indipendenza, come ad esempio vestirsi, camminare e mangiare. In seguito tuttavia verrà fatto riferimento ad una versione modificata dell'indice (Nicholas et al., 2003) che include solamente sei attività.

All'interno del questionario ai soggetti veniva posta la seguente domanda:

“Please tell me if you have any difficulty with these because of a physical, mental, emotional or memory problem. Exclude any difficulties you expect to last less than three months.

1. *Dressing, including putting on shoes and socks*
2. *Walking across a room*

3. *Bathing or showering*
4. *Eating, such as cutting up your food*
5. *Getting in or out of bed*
6. *Using the toilet, including getting up or down”*

Ad ogni categoria è assegnato valore 0 oppure 1, rispettivamente se non sono presenti oppure se sono presenti difficoltà nello svolgere una specifica attività. I valori così ottenuti per ogni soggetto sono sommati in modo da ottenere un punteggio finale che varia da 0 a 6. Più alto è il valore dell'indice, maggiori difficoltà sono riscontrate e minore la mobilità del rispondente.

Nelle *Tabella 3.6* è possibile osservare le distribuzioni della variabile relativa al numero di ADL che ciascun individuo presenta.

In generale circa l'88% degli intervistati in tutte le wave non presenta limitazioni.

Numero di limitazioni nelle ADL	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	19829	89.65	20080	89.72	29568	88.48	38601	88.76	34476	88.45
1	1214	5.49	1193	5.33	2036	6.09	2405	5.53	2266	5.81
2	507	2.29	458	2.05	698	2.09	915	2.10	889	2.28
3	207	0.94	217	0.97	398	1.19	510	1.17	420	1.08
4	130	0.59	129	0.58	200	0.60	317	0.73	289	0.74
5	97	0.44	136	0.61	217	0.65	277	0.64	267	0.68
6	135	0.61	168	0.75	301	0.90	466	1.07	372	0.95
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.6 – Numero e percentuali di soggetti partecipanti alle wave per numero di difficoltà nelle ADL

Tra i dati raccolti da SHARE è possibile trovare il numero di limitazioni che gli individui presentano nella mobilità. Anche questo è un ottimo indicatore dello stato di salute fisica degli individui ed è una delle componenti che viene considerata quando si cerca di valutare la disabilità di un individuo secondo la classificazione ICF, presentata nel Capitolo 1.

All'interno del questionario ai soggetti veniva posta la seguente domanda:

“We need to understand difficulties people may have with various activities because of a health or physical problem. Please tell me whether you have any difficulty doing each of the everyday activities on card 11. Exclude any difficulties that you expect to last less than three months. (Because of a health problem, do you have difficulty doing any of the activities on this card?)

- *Walking 100 metres*
- *Sitting for about two hours*
- *Getting up from a chair after sitting for long periods*
- *Climbing several flights of stairs without resting*
- *Climbing one flight of stairs without resting*
- *Stooping, kneeling, or crouching*
- *Reaching or extending your arms above shoulder level*
- *Pulling or pushing large objects like a living room chair*
- *Lifting or carrying weights over 10 pounds/ 5 kilos, like a heavy bag of groceries*
- *Picking up a small coin from a table”*

Nella *Tabella 3.7* è possibile osservare le distribuzioni della variabile relativa al numero di limitazioni nella mobilità che ciascun individuo ritiene di avere. Anche in questo caso, come nelle ADL, la maggior parte dei soggetti non presenta limitazioni. Tuttavia si tratta di percentuali che si aggirano intorno al 50%, molto inferiori rispetto a prima. Questo indica che quasi metà della popolazione analizzata in tutte le wave presenta almeno una limitazione nella mobilità e si tratta di un aspetto molto importante da valutare quando si cerca di individuare il livello di salute fisica dei rispondenti.

Limitazioni nella mobilità	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	11439	51.72	12130	54.20	17128	51.25	23431	53.88	20956	53.76
1	3472	15.70	3217	14.37	5254	15.72	6172	14.19	5470	14.03
2	2287	10.34	2069	9.24	3184	9.53	3998	9.19	3614	9.27
3	1542	6.97	1447	6.47	2142	6.41	2693	6.19	2505	6.43
4	1068	4.83	988	4.41	1624	4.86	2002	4.60	1843	4.73
5	728	3.29	777	3.47	1203	3.60	1434	3.30	1298	3.33
6	570	2.58	567	2.53	907	2.71	1141	2.62	990	2.54
7	427	1.93	470	2.10	761	2.28	907	2.09	853	2.19
8	316	1.43	341	1.52	558	1.67	760	1.75	671	1.72
9	183	0.83	223	1.00	388	1.16	544	1.25	463	1.19
10	87	0.39	152	0.68	269	0.80	409	0.94	316	0.81
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.7 – Numero e percentuali di soggetti partecipanti alle wave per numero limitazioni nella mobilità

Spesso le limitazioni nello svolgere attività quotidiane e nella mobilità sono il risultato di un processo di invecchiamento strettamente legato all'insorgenza di malattie croniche.

All'interno del questionario ai soggetti veniva posta la seguente domanda:

“Has a doctor ever told you that you had / Do you currently have any of the conditions on card 8? (With this we mean that a doctor has told you that you have this condition, and that you are either currently being treated for or bothered by this condition).

Please tell me the number or numbers of the conditions.

- 1. A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure*
- 2. High blood pressure or hypertension*
- 3. High blood cholesterol*
- 4. A stroke or cerebral vascular disease*
- 5. Diabetes or high blood sugar*
- 6. Chronic lung disease such as chronic bronchitis or emphysema*
- 7. Cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers*

8. *Stomach or duodenal ulcer, peptic ulcer*
9. *Parkinson disease*
10. *Cataracts*
11. *Hip fracture*
12. *Other fractures*
13. *Alzheimer's disease, dementia, organic brain syndrome, senility or any other serious memory impairment*
14. *Other affective or emotional disorders, including anxiety, nervous or psychiatric problems*
15. *Rheumatoid Arthritis*
16. *Osteoarthritis, or other rheumatism*
17. *Chronic kidney disease*
18. *None*
19. *Other conditions, not yet mentioned*

Per semplicità si considera solamente il numero di malattie croniche riportate da ciascun individuo (*Tabella 3.8*). È una variabile importante da considerare, poiché generalmente il concetto di buona salute è associato all'assenza di malattie. Inoltre, considerando dei soggetti ultracinquantenni, il numero di patologie individuate è superiore rispetto al numero di limitazioni che presentano.

Numero di malattie croniche	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	5671	25.64	5789	25.87	8472	25.35	10227	23.52	9583	24.59
1	7162	32.38	6882	30.75	9888	29.59	12959	29.80	11344	29.10
2	4500	20.34	4567	20.41	7042	21.07	9254	21.28	8133	20.87
3	2581	11.67	2731	12.20	4281	12.81	5681	13.06	5055	12.97
4	1243	5.62	1346	6.01	2116	6.33	2869	6.60	2635	6.76
5	612	2.77	604	2.70	950	2.84	1409	3.24	1260	3.23
6 +	350	1.58	462	2.06	669	2.00	1092	2.51	969	2.49
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.8 – Numero e percentuali di soggetti partecipanti alle wave per numero malattie croniche

SHARE rileva anche diversi fattori di rischio e fattori di protezione legati agli outcome di salute, tra cui l'indice di massa corporea (BMI)⁴ e l'attività fisica svolta dagli individui.

Queste variabili possono rappresentare dei possibili confondenti nella relazione tra SES, salute fisica e benessere, fulcro centrale dell'analisi che si intende svolgere.

Per quanto riguarda il BMI, i soggetti coprono quasi completamente in tutte le wave le due classi centrali (*Tabella 3.9*), cioè quelle relative all'essere normopeso oppure sovrappeso. Questo è importante poiché si tratta di una variabile che potrebbe influenzare non solo la salute fisica ma anche il benessere dei soggetti.

Anche l'attività fisica può essere una variabile che influisce non solo sulla salute fisica ma anche sul benessere degli individui. SHARE permette di rilevare due livelli di attività fisica: quella vigorosa e quella moderata.

All'interno del questionario ai soggetti venivano poste le seguenti due domande:

“We would like to know about the type and amount of physical activity you do in your daily life. How often do you engage in vigorous physical activity, such as sports, heavy housework, or a job that involves physical labour?”

1. *More than once a week*
2. *Once a week*
3. *One to three times a month*
4. *Hardly ever, or never”*

e

“How often do you engage in activities that require a moderate level of energy such as gardening, cleaning the car, or doing a walk?”

1. *More than once a week*
2. *Once a week*

⁴ L'indice di massa corporea (BMI) è una misura per valutare il peso corporeo in relazione all'altezza di un individuo. È definito (Quetelet, 1832) come la massa di un soggetto, divisa per il quadrato dell'altezza ed è generalmente espresso in kg/m².

L'indice così ottenuto può essere categorizzato in quattro classi, definite dall'OMS (1995):

- Sottopeso (< 18.5)
- Normopeso (18.5 – 24.9)
- Sovrappeso (25 – 29.9)
- Obeso (> 30)

3. *One to three times a month*
4. *Hardly ever, or never*”

Anche per queste variabili si riscontra un andamento simile in tutte le wave. La differenza principale è dovuta al fatto che circa il 70% degli individui in tutte le wave svolge attività fisiche di tipo moderato più di una volta a settimana (*Tabella 3.10*), mentre quando si osservano le attività che richiedono un alto livello di energia, ci si trova in una situazione intermedia per cui il 44% circa dei soggetti in tutte le wave non svolge mai o quasi mai attività di questo tipo, mentre il 33% circa le svolge più di una volta a settimana (*Tabella 3.11*). Probabilmente questa differenza è legata alla differente occupazione che hanno i soggetti. Molti rispondenti sono in pensione, ma quelli che non lo sono rientrano in una classe di età inferiore e quindi, a causa della loro occupazione oppure poiché svolgono attività fisica con costanza, rientrano nel livello con frequenza più alta.

Quando vengono trattate variabili relative all'attività fisica svolta in un'analisi che coinvolge la salute fisica degli individui e i nessi causali in cui essa è coinvolta, tuttavia, si corre il rischio di incorrere nel problema della *reverse causality*. Trattando variabili relative alla salute fisica e alle attività svolte, infatti, non si ha sempre la certezza dell'ordinamento temporale con cui avvengono: i soggetti stanno bene perché fanno attività fisica oppure non fanno attività fisica perché stanno male? Questo problema verrà affrontato più nel dettaglio nel momento in cui vengono applicati i metodi per svolgere l'analisi di mediazione nei Capitoli 5 e 6.

BMI	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Sottopeso	315	1.42	362	1.62	547	1.64	639	1.47	558	1.43
Normopeso	8854	40.03	8729	39.00	13170	39.41	17137	39.40	15217	39.04
Sovrappeso	9159	41.41	9262	41.38	13473	40.32	17512	40.27	15746	40.40
Obeso	3791	17.14	4028	18.00	6228	18.64	8203	18.86	7458	19.13
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.9 – Numero e percentuali di soggetti partecipanti alle wave per BMI

Attività vigorose	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Più di una volta a settimana	7570	34.39	7388	33.40	10788	32.76	14644	33.75	13083	33.63
Una volta a settimana	2913	13.23	2897	13.10	4379	13.30	5842	13.46	5262	13.52
Da una a tre volte al mese	1901	8.64	1897	8.58	2629	7.98	3500	8.07	2955	7.59
Quasi mai o mai	9618	43.69	9920	44.85	15107	45.87	19397	44.70	17591	45.21
Totali	22002	99.47	22102	98.75	32903	98.46	43383	99.75	38891	99.77
Valori mancanti	117	0.53	279	1.25	515	1.54	108	0.25	88	0.23

Tabella 3.10 – Numero e percentuali di soggetti partecipanti alle wave per livelli di attività fisica vigorosa svolta

Attività moderate	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Più di una volta a settimana	15359	69.77	15599	70.53	22502	68.32	29963	69.05	26620	68.42
Una volta a settimana	2830	12.85	2717	12.28	4401	13.36	5587	12.87	4949	12.72
Da una a tre volte al mese	1158	5.26	1135	5.13	1783	5.41	2366	5.45	2048	5.26
Quasi mai o mai	2657	12.07	2654	12.00	4220	12.81	5466	12.60	5274	13.56
Totali	22004	99.49	22105	98.75	32906	98.47	43382	99.75	38891	99.77
Valori mancanti	115	0.51	276	1.25	512	1.53	109	0.25	88	0.23

Tabella 3.11 – Numero e percentuali di soggetti partecipanti alle wave per livelli di attività fisica moderata svolta

Tra i possibili indicatori dello stato di salute fisica, dalla rilevazione svolta da SHARE, viene anche raccolta la forza massima della presa degli individui. Questo indicatore era già stato presentato nel Capitolo 1 in tutta la sua problematicità e per questo motivo si è deciso di non includerlo tra le possibili variabili che rappresentano la salute fisica.

3.1.3 Misure del benessere

Esistono numerosi modi per misurare il benessere e in letteratura sono proposte apposite scale per considerare come meglio possibile aspetti soggettivi, oggettivi e psicologici che lo caratterizzano.

Invece di focalizzarsi su variabili che semplificano molto il concetto di benessere, come la salute percepita, si preferisce utilizzarne altre due: la scala EURO – D e il CASP – 12. Queste due variabili permettono di riassumere in maniera più completa il livello di benessere dei soggetti.

La scala EURO – D è stata originariamente creata (Prince et al., 1999) per identificare una scala che considerasse i sintomi della depressione più frequenti. La scala che è stata ottenuta è composta da diversi item: depressione, pessimismo, suicidalità, senso di colpa, sonno, interesse, irritabilità, appetito, fatica, concentrazione (o lettura o intrattenimento), appagamento e tristezza. Agli item viene assegnato valore 0 se il sintomo non è presente, oppure 1 se è presente. I valori così ricavati sono sommati in modo da ottenere un punteggio assegnato ad ogni individuo che indica il grado di depressione da 0, soggetto non depresso, a 12 soggetto con massimo grado di depressione. Tale punteggio può essere eventualmente dicotomizzato per identificare un soggetto depresso scegliendo un livello soglia tra il 3 e il 4.

SHARE genera automaticamente all'interno dei dataset una variabile per la scala EURO-D. Si tratta di una variabile generata utilizzando le risposte date a 16 domande nel modulo relativo alla salute mentale, riportate nella *Tabella 3.12*. Al momento della creazione dell'indice sono state combinate le domande 4 e 5, 7 e 8, 10 e 11, 13 e 14 per permettere che i 16 item siano ridotti a 12.

È stato rilevato anche il CASP – 12 (*Tabella 1.13*). Si tratta di una rivisitazione della scala CASP – 19 (Hyde et al., 2003; Von dem Knesebeck et al., 2005), che veniva utilizzata per valutare la qualità della vita in età anziana. La scala è composta da quattro sottoscale che ne compongono l'acronimo: *control*, *autonomy*, *self-realization* e *pleasure*. I 12 item, che solitamente sono inseriti all'interno dei questionari di rilevazione, sono valutati con una scala Likert a quattro punti: spesso, qualche volta, raramente, mai. A queste risposte è assegnato rispettivamente valore 1, 2, 3 e 4. Il punteggio finale è una somma dei punteggi ottenuti in ogni item e va da un minimo di 12 ad un massimo di 48. Quanto più alto è il punteggio, tanto migliore è la qualità della vita. In letteratura tuttavia non è stata indicata nessuna soglia per stabilire dei possibili livelli in cui categorizzare i punteggi.

Domanda	Risposte
1. <i>In the last month, have you been sad or depressed?</i>	1. <i>Yes</i> 2. <i>No</i>
2. <i>What are your hopes for the future?</i>	1. <i>Any hopes mentioned</i> 2. <i>No hopes mentioned</i>
3. <i>In the last month, have you felt that you would rather be dead?</i>	1. <i>Any mention of suicidal feelings or wishing to be dead</i> 2. <i>No such feelings</i>
4. <i>Do you tend to blame yourself or feel guilty about anything?</i>	1. <i>Obvious excessive guilt or self-blame</i> 2. <i>No such feelings</i> 3. <i>Mentions guilt or self-blame, but it is unclear if these constitute obvious or excessive guilt or self-blame</i>
Se alla domanda 4 veniva data una risposta incerta (3.) allora era posta la seguente domanda:	1. <i>Example(s) given constitute obvious excessive guilt or self-blame</i> 2. <i>Example(s) do not constitute obvious excessive guilt or self-blame, or it remains unclear if these constitute obvious or excessive guilt or self-blame</i>
5. <i>So, for what do you blame yourself?</i>	1. <i>Example(s) given constitute obvious excessive guilt or self-blame</i>

6. <i>Have you had trouble sleeping recently?</i>	1. <i>Trouble with sleep or recent change in pattern</i> 2. <i>No trouble sleeping</i>
7. <i>In the last month, what is your interest in things?</i>	1. <i>Less interest than usual mentioned</i> 2. <i>No mention of loss of interest</i> 3. <i>Non-specific or uncodeable response</i>
Se alla domanda 7 veniva data una risposta incerta (3.) allora era posta la seguente domanda:	1. <i>Yes</i> 2. <i>No</i>
8. <i>So, do you keep up your interests?</i>	
9. <i>Have you been irritable recently?</i>	1. <i>Yes</i> 2. <i>No</i>
10. <i>What has your appetite been like?</i>	1. <i>Diminution in desire for food</i> 2. <i>No diminution in desire for food</i> 3. <i>Non-specific or uncodeable response</i>
Se alla domanda 10 veniva data una risposta incerta (3.) allora era posta la seguente domanda:	1. <i>Less</i> 2. <i>More</i> 3. <i>Neither more nor less</i>
11. <i>So, have you been eating more or less than usual?</i>	
12. <i>In the last month, have you had too little energy to do the things you wanted to do?</i>	1. <i>Yes</i> 2. <i>No</i>
13. <i>How is your concentration? For example, can you concentrate on a television program, film or radio program?</i>	1. <i>Difficulty in concentrating on entertainment</i> 2. <i>No such difficulty mentioned</i>
14. <i>Can you concentrate on something you read?</i>	1. <i>Difficulty in concentrating on reading</i> 2. <i>No such difficulty mentioned</i>
15. <i>What have you enjoyed doing recently?</i>	1. <i>Fails to mention any enjoyable activity</i> 2. <i>Mentions ANY enjoyment from activity</i>
16. <i>In the last month, have you cried at all?</i>	1. <i>Yes</i> 2. <i>No</i>

Tabella 3.12 – Lista di variabili che rientrano nel calcolo dell'indice EURO – D

Domanda	Sottoscala
<i>How often do you think your age prevents you from doing the things you would like to do?</i>	<i>Control</i>
<i>How often do you feel that what happens to you is out of your control?</i>	<i>Control</i>
<i>How often do you feel left out of things?</i>	<i>Control</i>
<i>How often do you think that you can do the things that you want to do?</i>	<i>Autonomy</i>
<i>How often do you think that family responsibilities prevent you from doing what you want to do?</i>	<i>Autonomy</i>
<i>How often do you think that shortage of money stops you from doing the things you want to do?</i>	<i>Autonomy</i>
<i>How often do you look forward to each day?</i>	<i>Pleasure</i>
<i>How often do you feel that your life has meaning?</i>	<i>Pleasure</i>
<i>How often, on balance, do you look back on your life with a sense of happiness?</i>	<i>Pleasure</i>
<i>How often do you feel full of energy these days?</i>	<i>Self-Realization</i>
<i>How often do you feel that life is full of opportunities?</i>	<i>Self-Realization</i>
<i>How often do you feel that the future looks good for you?</i>	<i>Self-Realization</i>

Tabella 3.13 – Lista di variabili che rientrano nel calcolo dell'indice CASP – 12

Nella *Tabella 3.14* sono riportati il valore medio, massimo, minimo e la percentuale di osservazioni mancanti relativi al CASP-12, mentre nella *Tabella 3.15* è riportata la distribuzioni delle variabile relativa alla scala EURO – D.

È possibile osservare che se viene seguita la classificazione di depressione indicata, circa il 24% dei soggetti nelle varie wave risulterebbe depresso. Poiché dicotomizzando la variabile si perde il livello di depressione sulla scala originale a cui si è interessati, si è deciso di mantenere la variabile EURO-D su scala discreta, ricordando che con valori bassi della scala si fa riferimento a soggetti che sono in situazioni di benessere psicologico migliore. Generalmente solo il 5% degli individui raggiunge livelli della scala superiori al 7, quindi si può affermare che il campione è composto da soggetti che si trovano in condizioni di benessere abbastanza buone. Questo viene

confermato anche dalla distribuzione del CASP – 12 nelle diverse wave, che presenta una media che si aggira sempre intorno a punteggi di 37-38, con una coda a sinistra che indica come ci sia un numero basso di soggetti che presenta un livello di benessere basso.

Wave	Età			
	Media	Minimo	Massimo	% di valori mancanti
1	37.599	12	48	40.6
2	37.506	12	48	7.3
4	37.933	12	48	5.87
5	38.296	12	48	5.04
6	38.321	12	48	6.22

Tabella 3.14 – Valore medio, minimo, massimo e percentuale di osservazioni mancanti relativi al CASP-12 in ogni wave

EURO – D	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6		
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	
0	5074	22.94	5328	23.81	6969	20.85	10019	23.04	8897	22.83	
1	Non depresso	4835	21.86	4975	22.23	7253	21.70	9396	21.60	8486	21.77
2		3777	17.08	3787	16.92	5652	16.91	7222	16.61	6651	17.06
3		2731	12.35	2813	12.57	4495	13.45	5559	12.78	4934	12.66
4		2021	9.14	1907	8.52	3253	9.73	3983	9.16	3586	9.20
5		1385	6.26	1305	5.83	2208	6.61	2781	6.39	2429	6.23
6		922	4.17	911	4.07	1443	4.32	1847	4.25	1648	4.23
7		579	2.62	556	2.48	962	2.88	1151	2.65	1035	2.66
8	Depresso	389	1.76	381	1.70	564	1.69	748	1.72	636	1.63
9		221	1.00	232	1.04	357	1.07	441	1.01	396	1.02
10		131	0.59	135	0.60	178	0.53	233	0.54	195	0.50
11		41	0.19	42	0.19	70	0.21	91	0.21	71	0.18
12		13	0.06	9	0.04	14	0.04	20	0.05	15	0.04
Totali		22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.15 – Numero e percentuali di soggetti partecipanti alle wave per livelli della scala EURO – D

3.1.4 Variabili socio – economiche

Una delle prime variabili rilevate riguarda il livello di istruzione raggiunto da ciascun individuo. SHARE a tale scopo utilizza l'*International Standard Classification of Education* (ISCED). Si tratta di una codifica standard dell'istruzione creata dall'UNESCO, che permette confronti internazionali. È una classifica organizzata secondo livelli di istruzione gerarchici. SHARE utilizza le codifiche specifiche per ogni paese sulla base della versione ISCED-97 (UNESCO, 2006) e nelle ultime due wave è disponibile anche la versione ISCED-11 (UNESCO, 2012) della classifica. Tuttavia, per garantire maggiore uniformità nella lettura dei risultati, nella tesi sarà considerata solamente la classificazione ISCED-97.

I soggetti vengono quindi collocati al livello di istruzione corrispondente, secondo le categorie globalmente riconosciute della classificazione ISCED-97:

- *None - Pre-primary level of education*
- *Primary level of education*
- *Lower secondary level of education*
- *Upper secondary level of education*
- *Post-secondary non tertiary education*
- *First stage of tertiary education*
- *Second stage of tertiary education.*

Per comodità si è deciso di raggruppare ulteriormente le categorie in modo da ottenere tre gradi di istruzione:

- Basso, comprende i primi due livelli della classificazione ISCED-97;
- Medio, comprende il terzo e il quarto livello della classificazione ISCED-97;
- Alto, comprende il quinto, il sesto e il settimo livello della classificazione ISCED-97.

Nella *Tabella 3.16* è possibile trovare le distribuzioni relative al livello di istruzione degli individui nelle wave analizzate. La maggior parte dei soggetti in tutte le wave ha un livello intermedio di istruzione. Appare interessante osservare come nelle prime wave la seconda percentuale più alta era nei livelli

più bassi di istruzione, mentre nelle ultime due wave si osserva la situazione opposta. Potrebbe essere un segnale di cambiamenti sociali che ci sono stati nella seconda metà del XX secolo e il fatto che, con l'ingresso di coorti più giovani, molti più soggetti hanno iniziato a raggiungere livelli più alti di istruzione.

Livello di istruzione	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Basso	7469	33.77	6872	30.70	9003	26.94	11116	25.56	9679	24.83
Medio	9938	44.93	10265	45.86	15984	47.83	20411	46.93	18378	47.15
Alto	4712	21.30	5244	23.43	8431	25.23	11964	27.51	10922	28.02
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.16 – Numero e percentuali di soggetti partecipanti alle wave per livello di istruzione

Nella *Tabella 3.17* è possibile osservare l'occupazione dei soggetti al momento della rilevazione. Il 50% del campione in tutte le wave si trova in pensione e le principali categorie successive, in ordine decrescente di frequenza, sono i soggetti occupati e le casalinghe.

Un'altra variabile che può essere utilizzata per riflettere le disponibilità economiche di un individuo è la sua situazione abitativa (*Tabella 3.18*).

All'interno del questionario ai soggetti veniva posta la seguente domanda:

“Is your household occupying the dwelling you live in as

- 1. Owner*
- 2. Member of a cooperative*
- 3. Tenant*
- 4. Subtenant*
- 5. Rent free”*

La variabile è stata successivamente resa dicotomica in modo da indicare se un soggetto è proprietario o meno dell'abitazione in cui abita. È così possibile osservare che circa il 72% dei soggetti in tutte le wave è proprietario della propria abitazione.

Occupazione	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
In pensione	11245	50.84	11266	50.34	17620	52.73	23298	53.57	21841	56.03
Occupato o lavoratore autonomo	6374	28.82	6440	28.77	9345	27.96	12521	28.79	10292	26.40
Disoccupato	768	3.47	639	2.86	1000	2.99	1351	3.11	979	2.51
Permanentemente malato o disabile	649	2.93	768	3.43	1027	3.07	1362	3.13	1172	3.01
Casalinga	3001	13.57	2845	12.71	3762	11.26	4038	9.28	3577	9.18
Altro	82	0.37	244	1.09	380	1.14	472	1.09	658	1.69
Missing by design	-	-	179	0.80	284	0.85	449	1.03	460	1.18
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.17 – Numero e percentuale di soggetti partecipanti alle wave per occupazione

Situazione abitativa	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
Proprietario	15713	71.04	16214	72.45	24187	72.38	31690	72.87	28487	73.08
Non proprietario	6406	28.96	5988	26.75	8947	26.77	11352	26.10	10032	25.74
Mancanti	-	-	179	0.8	284	0.85	449	1.03	460	1.18
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella 3.18 – Numero e percentuale di soggetti che partecipano ad ogni wave per possesso dell'abitazione

Quando si parla di stato socio-economico si fa riferimento ad un concetto molto vasto e complesso, composto da diverse dimensioni. Nella scelta del miglior modo per valutare il SES dei soggetti in analisi sono state valutate numerose variabili, tra cui alcune misure di reddito individuale e reddito equivalente. Queste variabili sono state escluse a causa della presenza di casalinghe nel campione analizzato che, non percependo reddito, potevano risultare problematiche da includere correttamente nelle analisi.

Si è deciso di utilizzare la ricchezza come variabile rappresentativa dello stato socio-economico. Non essendo già presente una variabile relativa alla

ricchezza resa disponibile da SHARE, sono state considerate le seguenti variabili per rappresentarla:

- Valore della residenza principale. All'interno del questionario ai soggetti veniva posta la domanda *“In your opinion, how much would you receive if you sold your property today?”*
- Valore di altri possedimenti. All'interno del questionario ai soggetti veniva posta la domanda *“Do you [or your husband/wife/partner] own secondary homes, holiday homes, other real estate, land or forestry? In your opinion, how much would this or these properties be worth now if you sold it?”*
- Conti bancari. All'interno del questionario ai soggetti veniva posta la domanda *“About how much do you [and your husband/wife/partner] currently have in bank accounts, transaction accounts, saving accounts or postal accounts?”*
- Bond, azioni e fondi comuni. All'interno del questionario ai soggetti veniva posta la domanda *“About how much do you and your partner currently have in government or corporate bonds, stocks or shares that are listed or unlisted on stock market, mutual funds or managed investment accounts?”*

Le distribuzioni delle variabili che costituiscono le dimensioni della ricchezza sono riportate nell'Appendice A.

3.2 Creazione delle variabili

L'obiettivo della tesi è quello di valutare la relazione tra stato socio-economico, benessere e salute fisica degli individui. È necessario quindi costruire delle variabili che siano in grado di rappresentare al meglio questi tre concetti.

Sono presenti numerose variabili che possono essere utilizzate come misure monodimensionali per rappresentare lo stato socio-economico, il benessere e la salute fisica degli individui intervistati. Tuttavia, invece di scegliere solamente una variabile, si è deciso di ricorrere a delle analisi fattoriali che

permettono di considerare congiuntamente diverse misure monodimensionali. L'analisi fattoriale è un metodo che mediante un certo numero di fattori non osservabili cerca di spiegare le correlazioni tra le variabili osservate. Si ritiene un approccio particolarmente adatto poiché tutte e tre le variabili possono essere considerate delle variabili latenti, cioè dei costrutti non osservabili e misurabili, che possono però essere basati su una serie di altre variabili che li definiscono.

Per ottenere una variabile che rappresenti lo stato socio-economico degli individui, inteso come ricchezza di coppia, si è deciso di utilizzare le seguenti sotto-dimensioni: valore della residenza principale; valore di altri possedimenti; conti bancari; bond, azioni e fondi comuni.

Per definire la condizione di salute fisica degli individui, si è deciso di utilizzare il numero di limitazioni nelle attività della vita quotidiana e di difficoltà legate alla mobilità riportate dagli stessi.

Si tratta di eccellenti indicatori per valutare la salute fisica, specialmente poiché in parte delle analisi successive (Capitolo 5 e 6) il campione verrà ristretto ai 6707 soggetti che partecipano a tutte le wave e con questa variabile si può valutare il processo di invecchiamento in buona salute dei soggetti. Specialmente in età più avanzate, infatti, non è la presenza della malattia in quanto tale il problema, ma l'effetto che questa ha nella vita quotidiana dei soggetti affetti.

Inoltre, per rappresentare il concetto di benessere, si è deciso di considerare il CASP-12, misura riconosciuta per valutare il benessere degli individui, congiuntamente all'EURO-D, che permette di considerare anche aspetti prettamente psicologici della vita dei soggetti in analisi.

La scala della variabile relativa all'EURO-D è stata invertita in modo che a livelli alti corrisponda una migliore condizione di benessere.

Come prima cosa è stata valutata la correlazione delle variabili appena elencate, in modo da avere un'idea più chiara delle relazioni che hanno l'una con l'altra. È stato utilizzato il coefficiente di correlazione Tau-b di Kendall

(1938; 1948) e tutte le correlazioni stimate in tutte le wave sono risultate significative.

Per semplicità vengono riportate solamente le correlazioni tra le variabili nella sesta wave (*Tabella 3.19*), mentre le altre sono riportate nell'Appendice B (*Tabelle B.1 – B.4*), poiché sono stati rilevati dei valori simili in tutte le wave.

È possibile osservare che le limitazioni nelle ADL sono fortemente associate alle difficoltà nella mobilità e che entrambe sono associate negativamente alle variabili relative al benessere degli individui e quelle relative alla ricchezza. Si può quindi ipotizzare che maggiori ricchezze possano portare a condizioni di salute migliore e che un livello più alto di salute porti ad un benessere maggiore. Le variabili relative al CASP-12 e all'EURO-D sono correlati positivamente ed entrambe presentano una correlazione positiva con le variabili relative alla ricchezza. Quindi in ottica causale si può ipotizzare che uno stato socio-economico migliore possa portare ad un benessere più alto. Le variabili relative alla ricchezza presentano tutte un'associazione positiva tra di loro.

Le associazioni più forti che si evidenziano sono tra limitazioni nelle ADL e nella mobilità, tra CASP-12 e EURO-D e tra le variabili relative alla ricchezza. Oltre a queste si evidenziano delle forti correlazioni considerando mobilità e CASP-12 o EURO-D, e anche tra le variabili relative alla ricchezza e il CASP-12. Si può quindi avere una conferma del fatto che effettivamente le variabili considerate tendono a raggrupparsi tra di loro, ma anche della complessità del fenomeno e del fatto che le variabili siano altamente interconnesse.

Coefficiente di correlazione Tau b di Kendall								
	Limitazioni ADL	Limitazioni mobilità	CASP -12 ⁵	EURO-D invertita ⁶	Valore della residenza principale	Valore di altri possedimenti	Conti bancari	Bond, azioni e fondi comuni
Limitazioni ADL	1.00000	0.44319	-0.20324	-0.20222	-0.06596	-0.10708	-0.07936	-0.08082
Limitazioni mobilità	0.44319	1.00000	-0.29989	-0.31125	-0.09203	-0.14010	-0.10318	-0.10627
CASP-12	-0.20324	-0.29989	1.00000	0.37223	0.11460	0.13728	0.19757	0.16044
EURO-D invertita	-0.20222	-0.31125	0.37223	1.00000	0.05307	0.08238	0.09449	0.08314
Valore della residenza principale	-0.06596	-0.09203	0.11460	0.05307	1.00000	0.17473	0.15456	0.21107
Valore di altri possedimenti	-0.10708	-0.14010	0.13728	0.08238	0.17473	1.00000	0.19218	0.19907
Conti bancari	-0.07936	-0.10318	0.19757	0.09449	0.15456	0.19218	1.00000	0.26278
Bond, azioni e fondi comuni	-0.08082	-0.10627	0.16044	0.08314	0.21107	0.19907	0.26278	1.00000

Tabella 3.19 – Coefficienti di correlazione Tau b di Kendall delle variabili che rappresentano lo stato socio-economico, il benessere e la salute fisica degli individui nella sesta wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Dopo queste prime osservazioni, si è deciso di procedere con lo svolgimento di un'analisi fattoriale esplorativa, in modo da avere un'idea più chiara delle

⁵ A valori alti dell'indice CASP-12 corrispondono condizioni di benessere migliori.

⁶ La scala EURO-D è stata invertita prima del calcolo delle correlazioni in modo che avesse la stessa direzione del CASP-12: a valori alti della scala corrispondono livelli di depressione più bassi e quindi condizioni di benessere migliori.

relazioni tra le variabili e comprendere se gli accorpamenti ipotizzati possono effettivamente essere plausibili nell'analisi.

Emerge così la presenza chiara di due distinti fattori: uno relativo alle variabili socio-economiche e l'altro a quelle di salute. Il numero di fattori è stato scelto automaticamente dall'algoritmo che conserva solamente quelli che presentano autovalore maggiore di 1. Tuttavia, analizzando il *pattern* dei fattori prima della rotazione è possibile osservare tre cluster distinti di variabili (per la sesta wave *Figura 3.1*).

I risultati dell'analisi fattoriale esplorativa per la sesta wave sono riportati nell'Appendice B (*Tabella B.5 – B.7*)

Si ritiene che le variabili relative alla salute fisica e al benessere siano mantenuti insieme a causa del forte legame che vi è tra le due componenti. Questo è confermato anche nel momento in cui viene imposto all'algoritmo di creare tre fattori; in questo modo le tre dimensioni appaiono separate in maniera più evidente (per la sesta wave *Tabella 3.20*).

Analizzando la matrice di correlazione dei tre fattori così ottenuti (*Tabella 3.21*) si riesce ad avere un'ulteriore immagine delle relazioni che entrano in gioco tra le variabili: ci si aspetta un miglioramento delle condizioni di salute e del benessere al crescere dello stato socio-economico degli individui e un aumento del benessere insieme al miglioramento delle condizioni fisiche dei soggetti.

Le due correlazioni tra il primo e il terzo fattore e tra il secondo e il terzo sono negative poiché la salute fisica è rappresentata dalle difficoltà dei soggetti nello svolgere attività quotidiane e nella mobilità, e al crescere del valore assunto dalla variabile, le condizioni di salute peggiorano.

In seguito a questa valutazione delle relazioni tra le variabili, sono state svolte tre diverse analisi fattoriali per ogni wave, ognuna con la finalità di creare una sola variabile che rappresenti ciascun aspetto della vita degli individui: lo stato socio-economico, la salute fisica e il benessere.

Alla crescita dei valori della prima e della terza variabile corrispondono livelli più alti di stato socio-economico e benessere, al contrario, a valori più alti del secondo fattore corrisponde un livello di salute fisica peggiore.

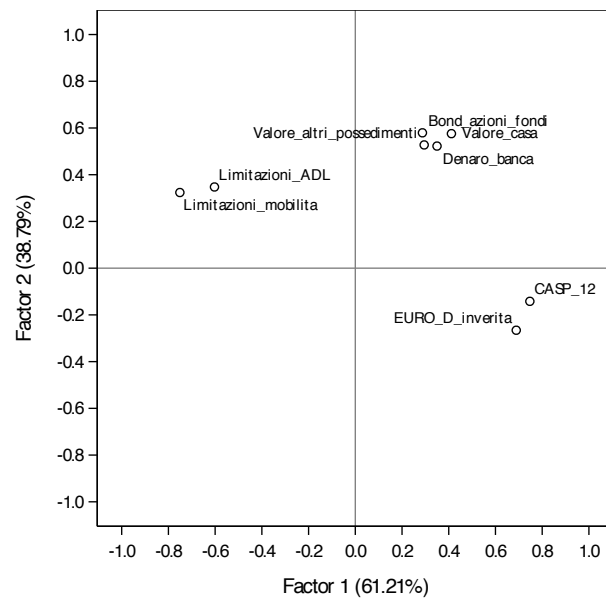


Figura 3.1 – Rappresentazione grafica dei pesi fattoriali non ruotati nella sesta wave (default)

	Factor 1	Factor 2	Factor 3
Valore della residenza principale	0.69770	0.04337	-0.04119
Valore di altri possedimenti	0.65855	-0.22745	-0.21615
Conti bancari	0.65264	-0.01426	0.03805
Bond, azioni e fondi comuni	0.56345	0.29926	0.24855
EURO-D	-0.08364	0.84619	-0.06708
CASP	0.06133	0.81131	-0.08751
Limitazioni ADL	0.00296	0.02437	0.89687
Limitazioni mobilità	-0.01963	-0.29315	0.72476

Tabella 3.20 – Matrice dei fattori selezionati ruotati con trasformazione Promax nella sesta wave

	Factor1	Factor2	Factor3
Factor1	1.00000	0.17461	-0.07417
Factor2	0.17461	1.00000	-0.30503
Factor3	-0.07417	-0.30503	1.00000

Tabella 3.21 – Matrice di correlazione dei fattori selezionati ruotati con trasformazione Promax nella sesta wave

3.3 Relazioni tra variabili e possibili confondenti

Dopo aver definito tutte le variabili che entrano a far parte dell'analisi, è possibile approfondire le relazioni tra stato socio-economico, benessere, salute e confondenti. Poiché i risultati delle diverse wave risultano essere simili, sono riportati di seguito solamente quelli ottenuti per la sesta wave.

Sono state svolte inizialmente della analisi di correlazione e associazione mantenendo la scala continua delle variabili relative a stato socio-economico, benessere e salute fisica. Successivamente, per valutare quali variabili costituiscono effettivamente dei confondenti e per approfondire le relazioni tra essi e i tre fattori, si è deciso di categorizzate questi ultimi in tre livelli, dividendone la distribuzione al 33-esimo e al 66-esimo percentile.

Considerando i fattori ottenuti nel Paragrafo 3.2, relativi allo stato socio-economico, al benessere e alla salute degli individui e mantenendoli continui, è possibile valutare la correlazione con l'età, anch'essa mantenuta continua (*Tabella 3.22*).

Si ricorda che la salute fisica è costituita dalle limitazioni nelle ADL e nella mobilità, quindi a valori alti del fattore corrispondono condizioni di salute peggiori.

È possibile quindi affermare che in età più avanzate le condizioni di salute fisica sono peggiori, prevalgono soggetti con livello socio-economico più basso e con un grado di benessere inferiore.

Coefficiente di correlazione Rho di Pearson				
	SES	Benessere	Salute fisica	Età
SES	1.00000	0.18501	-0.12999	-0.02523
Benessere	0.18501	1.00000	-0.44718	-0.12646
Salute fisica	-0.12999	-0.44718	1.00000	0.33822
Età	-0.02523	-0.12646	0.33822	1.00000

Tabella 3.22 – Coefficienti di correlazione Rho di Pearson delle variabili relative allo stato socio-economico, al benessere, alla salute fisica e all'età degli individui nella sesta wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Per studiare la relazione dei fattori con le variabili relative al genere, allo stato civile, al paese di residenza, al livello di istruzione, al BMI, alla situazione abitativa e all'attività fisica dei soggetti analizzati, invece, trattandosi di relazioni asimmetriche tra variabili continue e variabili categoriali, si è deciso di procedere con l'utilizzo dell'ANOVA non parametrica e il test di Kruskal – Wallis. Tutti i χ^2 dei test risultano significativi e indicano la presenza di un'associazione tra le variabili. Tramite l'analisi delle tabelle di contingenza ottenute categorizzando i fattori è possibile svolgere un'ulteriore valutazione delle caratteristiche del campione.

Il processo di invecchiamento gioca un ruolo fondamentale nel benessere dei soggetti, poiché nelle classi tra i 50 e i 70 anni prevalgono livelli medio-alti di benessere, che tendono ad abbassarsi oltre i 70 e gli 80 anni. Questo può essere una risposta del fatto che anche per la salute fisica si osserva un comportamento simile: in età più avanzate le difficoltà dei soggetti risultano essere molto superiori rispetto alle classi d'età intermedie.

La distribuzione dell'età tra uomini e donne appare leggermente differente (*Figura 3.2*) poiché gli uomini sono prevalentemente concentrati nelle classi che vanno dai 55 ai 75 anni mentre le donne si distribuiscono in maniera più

omogenea e rappresentano la maggior parte dei soggetti ultra 80enni e con meno di 50 anni. La prevalenza di donne anziane porta ad osservare come queste si trovino non solo in condizioni fisiche peggiori, ma anche a livelli di benessere più bassi.

In merito allo stato civile dei soggetti, si osserva che la maggior parte di essi risulta sposata e si osserva una percentuale maggiore di donne vedove, causata sempre dalla differente distribuzione per età.

Anche valutando la relazione tra benessere e stato civile è possibile trovare un'associazione significativa tra le variabili. Il 70% circa dei soggetti sposati si trova in situazioni di benessere medio-alte mentre i soggetti separati/divorziati e vedovi si trovano rispettivamente per il 37% e per il 50% a livelli di benessere bassi.

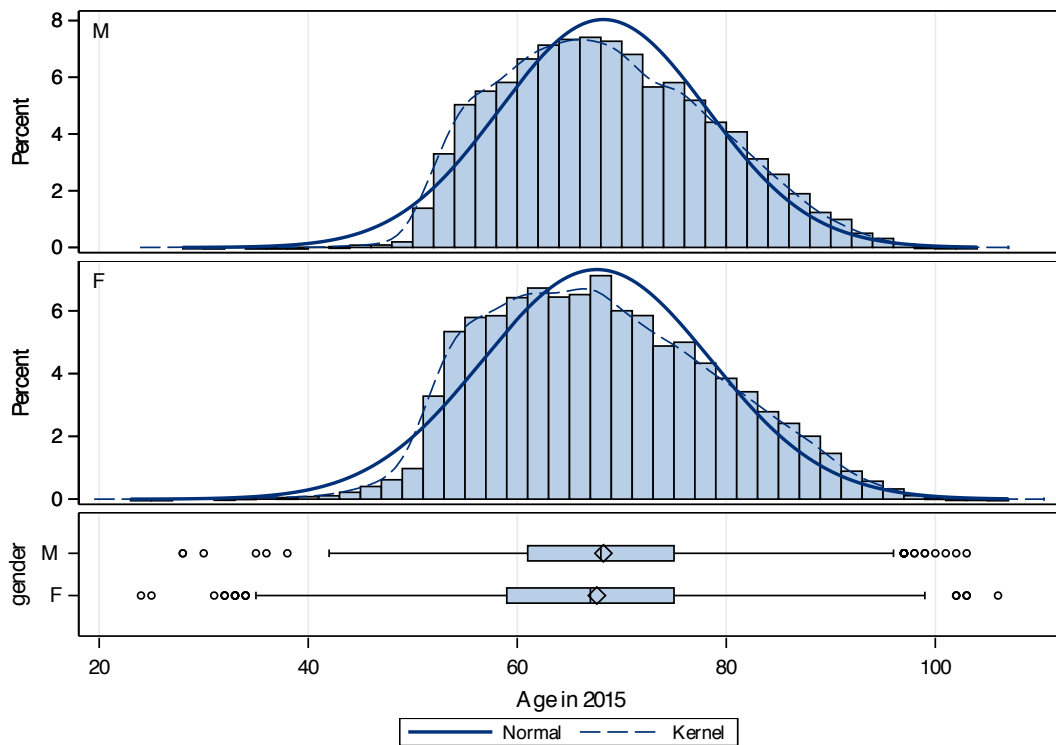


Figura 3.2 – Distribuzione della variabile relativa all'età secondo il genere nella sesta wave

Risultano interessanti anche le differenze relative al paese di residenza degli individui. Si tratta di paesi europei, ma che presentano comunque delle differenze tra di loro.

Nel campione, i paesi con la popolazione in media più giovane sono la Danimarca, la Germania e il Belgio (età media rispettivamente di 65, 66 e 66 anni), mentre quelli con la popolazione mediamente più anziana sono la Spagna e la Svezia (età media di 70 anni). Questo risulta ancora più interessante, considerando il benessere nei diversi paesi: insieme ad Austria, Danimarca e Svizzera, anche la Svezia, con la popolazione più anziana, si trova in condizioni di benessere migliori.

Questa differenza è legata alla ricchezza degli individui nei diversi paesi analizzati, infatti, è possibile osservare che i soggetti mediamente più ricchi risiedono in Svezia, Danimarca, Svizzera e Belgio (*Tabella 3.23*). Si tratta di un ulteriore segnale del legame tra stato socio-economico e benessere.

I paesi con i soggetti mediamente in condizioni di benessere inferiore sono la Spagna, l'Italia e la Francia che costituiscono, insieme all'Austria, il gruppo di paesi con livelli di ricchezza mediamente più bassi.

Si osserva inoltre un legame di associazione significativo tra ricchezza e livello di istruzione: a livelli più alti di istruzione corrispondono soggetti con livelli socio-economici più alti (*Tabella 3.24*).

Non appare quindi sorprendente il fatto che i paesi come Italia e Spagna abbiano SES più basso ma anche livelli di istruzione inferiori, rispetto a Belgio, Danimarca, Svezia e Germania in cui la situazione è invertita e si hanno soggetti con livelli di istruzione più alti a cui corrisponde un SES maggiore.

Paese	SES			
	Media	Minimo	Massimo	Deviazione standard
Austria	-0.2965676	-0.6962360	3.1085545	0.5524856
Germania	-0.2088151	-0.6962360	7.2133188	0.6478003
Svezia	0.1083474	-0.6962360	6.6971004	0.8207600
Spagna	-0.3204583	-0.6962360	2.6353982	0.3463985
Italia	-0.2733789	-0.6962360	3.1642767	0.3898047
Francia	-0.2152017	-0.6962360	3.1355183	0.4176731
Danimarca	0.0372164	-0.6962360	4.8467098	0.6755942
Svizzera	1.5328884	-0.6962360	20.1033083	2.4196336
Belgio	0.2017822	-0.6962360	6.5344807	0.8916738

Tabella 3.23 – Valore medio, minimo e massimo della variabile relativa allo stato socio-economico per paese di residenza

Livello di istruzione	SES			
	Media	Minimo	Massimo	Deviazione standard
Basso	-0.3138409	-0.6962360	8.6914022	0.4604765
Medio	-0.0465085	-0.6962360	18.9473329	0.9287823
Alto	0.3563816	-0.6962360	20.1033083	1.3083009

Tabella 3.24 – Valore medio, minimo e massimo della variabile relativa allo stato socio-economico per livello di istruzione

3.3 Identificazione dei confondenti

Si ricorda che con il termine confondimento si fa riferimento alla distorsione e inaccuratezza delle misure di associazione che avviene quando sia l'esposizione di interesse, sia l'outcome risultano associati ad una terza variabile.

Nello svolgere analisi di mediazione, identificare e depurare le stime degli effetti da questo fenomeno risulta fondamentale. Il metodo utilizzato per il riconoscimento delle variabili che fungono da confondenti prende il nome di

change – in – estimate criterion e si basa sulla valutazione della quota secondo cui le stime delle misure di associazione di modificano. Una volta controllato per il presunto confondente, se la quota supera una soglia prefissata allora la variabile è ritenuta confondente. Mediante simulazione è stato dimostrato che la soglia a cui si fa riferimento più adatta è quella del 10 – 20% (Maldonado e Greenland, 1987).

A livello pratico, considerando le variabili relative allo stato socio-economico, al benessere e alla salute fisica categorizzate, sono stati costruiti per ogni wave tre modelli logistici cumulati ad odds non proporzionali (Agresti, 2013), considerando rispettivamente come risposta ed esplicativa:

- Salute fisica e SES;
- Benessere e SES;
- Benessere e salute fisica;

e in questo modo è stato possibile ottenere le stime degli *Odds Ratio* marginali.

Successivamente sono state aggiunte al modello una per volta le variabili relative a età, genere, BMI, situazione abitativa, paese, stato civile e istruzione, che si pensava potessero confondere l'effetto di interesse, in modo da valutare separatamente per ciascuna se il suo ingresso nel modello comportasse modifiche sufficienti per definirla effettivamente confondente.

I risultati sono stati riassunti nella *Tabella 3.25*.

Confondenti		
SES → Salute fisica	SES → Benessere	Salute fisica → Benessere
Età	Paese di residenza	Età
Genere	Stato civile	Genere
	Livello di istruzione	BMI
		Situazione abitativa

Tabella 3.25 – Confondenti nelle relazioni tra stato socio-economico, benessere e salute fisica

Queste brevi osservazioni sono fondamentali per comprendere la complessità del fenomeno che si intende studiare e il forte legame tra le variabili coinvolte. In uno scenario simile, la correzione dall'effetto di variabili che svolgono il ruolo di confondenti appare fondamentale, non solo per il rispetto delle assunzioni dei metodi che sono presentati nel Capitolo 4, ma anche per avere una visione più completa e approfondita della relazione tra stato socio-economico degli individui e benessere, e per valutare il ruolo della salute fisica in questa relazione.

4. ANALISI DI MEDIAZIONE

4.1 Cosa si intende con il termine “mediazione”?

Il termine mediazione viene utilizzato in statistica e nelle scienze sociali per indicare un meccanismo o un processo caratterizzato dalla presenza di una variabile indipendente che causa, influenza una variabile dipendente, o outcome, attraverso una terza variabile chiamata mediatore, intermediario o interveniente.

I metodi che permettono di svolgere analisi di mediazione sono utili per riuscire a determinare non solamente la presenza di un effetto diretto che una variabile ha su un'altra ma, tramite la presenza di un mediatore, per comprendere la reale natura della relazione tra le variabili di interesse.

Bisogna quindi prendere in considerazione due aspetti: il procedimento con cui una variabile influenza un'altra e il modo in cui ciò accade. L'analisi di mediazione aiuta a comprendere situazioni in cui le variabili non sembrano avere un legame diretto immediato.

Per semplificare la comprensione delle relazioni di cui si sta parlando, spesso si fa ricorso alla parte grafica della *path analysis* per la descrizione di un processo causale. Si tratta di un metodo sviluppato tra il 1918 e il 1934 dal genetista Sewall Wright (Pedhazur, 1982), nel quale ogni freccia unidirezionale viene utilizzata per indicare una relazione causale, mentre le frecce bidirezionali indicano la presenza di un'associazione tra le variabili coinvolte.

In questo modo è possibile visualizzare graficamente lo scenario della mediazione a cui si è interessati.

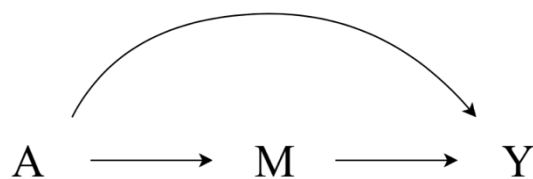


Figura 4.1 – Modello di mediazione semplice

Si veda a tale proposito la *Figura 4.1*, dove A indica la variabile indipendente o esposizione, Y la variabile dipendente o outcome e M il mediatore. Per semplicità non viene per ora considerata la possibile presenza di covariate.

4.2 Il controfattuale

Il metodo della mediazione verrà affrontato seguendo l'approccio controfattuale degli outcome potenziali⁷. Tale approccio si basa tutto sul cercare di capire cosa sarebbe successo se l'esposizione, o semplicemente più in generale la variabile indipendente, avesse assunto un valore diverso da quello che effettivamente ha assunto. L'idea di partenza è che se questo cambiamento modifica l'outcome di interesse, allora si può pensare che l'esposizione causi, o comunque abbia un effetto, sull'outcome (Menzies, 2017).

Sia Y_a il valore assunto dalla risposta quando l'esposizione A è fissata al valore a ; per semplicità si può considerare un'esposizione dicotomica con $a = 0$ oppure $a = 1$, ottenendo così che gli outcome potenziali o controfattuali possano essere identificati come Y_0 e Y_1 .

Il problema principale di quest'approccio è che in generale si conosce solamente quello che è effettivamente accaduto, quindi solo Y_0 o Y_1 e non è possibile valutare direttamente cosa sarebbe successo altrimenti. Questo solitamente prende il nome di *consistency assumption* (VanderWeele, *Explanation in Causal Inference: Methods for Mediation and Interaction.*, 2015). Il valore controfattuale dell'outcome è quindi solitamente qualcosa che non si conosce e il quesito può essere ridotto ad un problema di valori mancanti talvolta difficile da affrontare in un contesto empirico.

⁷ L'idea di considerare la causazione da un punto di vista controfattuale risale a Hume (1748), ma all'interno del mondo statistico viene introdotta da Neyman (1923) per la prima volta una notazione precisa dell'approccio controfattuale. Tutto questo viene poi approfondito e portato avanti da Rubin (1974; 1978), fino ad essere ampliato ulteriormente con l'aggiunta di più esposizioni ed esposizioni che variano nel tempo da Robins (1986). Viene spesso affiancato da una rappresentazione grafica di causalità introdotta da Spirtes, et al. (1993) e Pearl (1995; 2001).

L'impossibilità di conoscere entrambi i valori della risposta per lo stesso individuo rappresenta il problema fondamentale dell'inferenza causale (Angrist & Pischke, 2009).

L'effetto causale può essere semplicemente definito come la differenza tra gli outcome ottenuti nel caso in cui fossero state attuate le azioni potenziali, cioè $Y_1 - Y_0$, tuttavia questa differenza solitamente non può essere identificata a livello individuale.

Lavorando a livello di popolazione si può però sperare di riuscire a fare inferenza sugli effetti medi: $E(Y_1 - Y_0)$. La randomizzazione, quando può essere attuata, permette di comparare a livello di media di popolazione il risultato di due trattamenti o esposizioni differenti, consentendo di ottenere che $Y_a \perp A$. Nonostante possa sembrare scorretto un legame di indipendenza come quello appena definito, in realtà al suo interno è contenuto il significato centrale del controfattuale: non è l'outcome osservato ad essere indipendente dall'esposizione, ma lo sono gli outcome potenziali, Y_a , in quanto rappresentano gli outcome che si potrebbero verificare a priori, indipendentemente dal valore che poi effettivamente ha assunto l'esposizione. Purtroppo molto spesso non si ha la possibilità di ottenere dati provenienti da studi sperimentali, ma sono disponibili solo dati osservazionali e quindi nonostante possano essere svolti dei controlli per assicurarsi che sia possibile il confronto fra le risposte con due stati di esposizioni differenti, non si ha mai la certezza che essi siano sufficienti. Si cercherà pertanto di garantire che risulti almeno possibile confrontare gli outcome potenziali all'interno di gruppi definiti sulla base dei valori delle covariate C , assumendo $Y_a \perp A|C$. Questo prende il nome di *no-unmeasured-confounding assumption* o *exogeneity assumption* (VanderWeele, 2015).

Sotto queste assunzioni risulta possibile ottenere gli effetti causali dai dati come

$$\begin{aligned} E(Y_1 - Y_0|c) &= E(Y_1|c) - E(Y_0|c) = E(Y_1|A = 1, c) - E(Y_0|A = 0, c) \\ &= E(Y|A = 1, c) - E(Y|A = 0, c) \end{aligned} \tag{4.1}$$

e quindi, eliminando il condizionamento,

$$\begin{aligned} E(Y_1 - Y_0) &= \sum_c E(Y_1 - Y_0|c)P(c) \\ &= \sum_c \{E(Y|A = 1, c) - E(Y|A = 0, c)\}P(c) \end{aligned} \quad (4.2)$$

Tramite le Formule 4.1 e 4.2 è possibile osservare come l'effetto causale medio dell'esposizione sull'outcome possa essere ottenuto semplicemente come una differenza di valori attesi, stimabili dai dati osservati. La Formula 4.2 rappresenta l'effetto causale medio dell'esposizione sull'outcome sull'intera popolazione invece che considerarlo all'interno di strati di covariate $C = c$ come veniva fatto nella Formula 4.1.

È importante sottolineare che non tutte le forme di causazione possono essere ricondotte ad un ambito controfattuale, ad esempio nel caso in cui siano presenti più cause che influiscono sull'avvenimento di un effetto.

Si può considerare il caso semplice con due cause che agiscono contemporaneamente su un outcome. Nel caso in cui solamente una delle due venisse modificata, l'outcome potrebbe rimanere invariato non perché la variabile non rappresenti un'effettiva causa, ma semplicemente perché è presente qualche altro fenomeno, cioè la seconda causa, che agisce contemporaneamente al primo e modifica il risultato che ci si attende. In questa situazione, analizzando il fenomeno dal punto di vista controfattuale, si potrebbe giungere alla conclusione errata che la prima variabile non sia una causa della risposta.

I metodi descritti in seguito si concentrano su situazioni in cui si ha ragione di ritenere che il mediatore abbia un ruolo sufficientemente importante nell'effetto che la causa ha sulla risposta. Si cercano degli strumenti che permettano di capire con quale entità l'effetto complessivo possa essere

dovuto alla presenza dell'intermediario. Per questo motivo si fa la distinzione tra tre tipologie di effetto:

- Effetto diretto, cioè l'effetto della causa sull'outcome non dovuto alla presenza del mediatore.

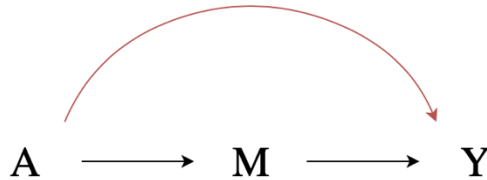


Figura 4.2 – Modello di mediazione semplice con evidenziato in rosso l'effetto diretto

- Effetto indiretto, cioè l'effetto della causa sull'outcome dovuto alla presenza del mediatore.

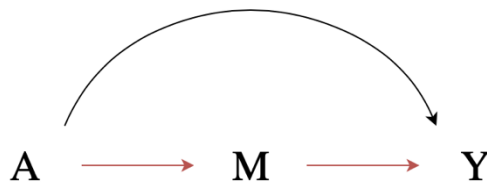


Figura 4.3 – Modello di mediazione semplice con evidenziato in rosso l'effetto indiretto

- Effetto totale, cioè l'effetto complessivo che la causa ha sull'outcome.

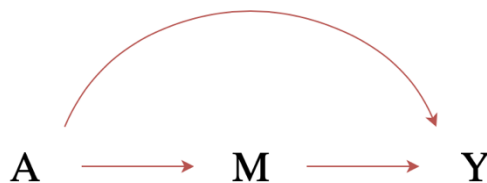


Figura 4.4 – Modello di mediazione semplice con evidenziato in rosso l'effetto totale

Le definizioni appena fornite servono per iniziare ad avvicinarsi alla mediazione e al modo in cui vengono valutati i legami causali tra le variabili, verranno tuttavia fornite in seguito delle definizioni più formali e corrette di questi effetti utilizzando la notazione propria dell'approccio controfattuale.

Come sarà chiarito meglio successivamente, serviranno assunzioni molto più forti per valutare empiricamente la presenza di mediazione rispetto a quelle necessarie all'identificazione di una causazione complessiva.

4.3 Approcci basati sulla regressione

Di seguito verranno presentati degli approcci basati sulla regressione per l'analisi di mediazione seguendo l'approccio controfattuale. Questo permetterà di generalizzare i metodi per la mediazione che sono usati in campo epidemiologico, sociale ed economico, consentendo di superare alcune delle loro difficoltà. Verranno inizialmente ripresi i metodi più classici per poi ampliarli, fornendo definizioni formali degli effetti seguendo l'approccio controfattuale. La loro definizione e stima verrà presentata in uno scenario di regressione.

4.3.1 Approccio classico

Nella letteratura sono presenti metodi tradizionali che permettono di svolgere analisi di mediazione che prendono ispirazione dalla *path analysis* di Wright (1921), adesso vista come caso particolare dei modelli ad equazioni strutturali (SEM). Questi modelli permettono di stimare degli effetti diretti e indiretti modellando matrici di varianza e correlazione. Tali metodi sono spesso criticati in quanto si ritiene che non diano sufficiente importanza e considerazione ai possibili confondenti, mentre l'approccio controfattuale pone grande enfasi su di loro e sulle ipotesi che li riguardano (Robins e Greenland, 1992; Pearl, 2001). Per questo motivo molti autori hanno cercato di tradurre i modelli ad equazioni strutturali e riportare anch'essi all'interno dello scenario del controfattuale (VanderWeele e Vansteelandt, 2009; Pearl, 2011).

Il primo metodo tradizionale che verrà affrontato e poi nel paragrafo successivo trasposto nell'ambito del controfattuale, prende il nome di *product method* o *product of coefficients method*. Si tratta di un approccio che è stato

portato avanti da Baron e Kenny (1986) e rappresenta il metodo più classico per l'analisi di mediazione.

I due autori, facendo riferimento ad un modello di mediazione semplice come quello della *Figura 4.1*, descrivono dei criteri necessari a definire un mediatore: una variazione nei livelli dell'esposizione provoca una variazione significativa nei livelli del presunto mediatore; una variazione nei livelli del mediatore provoca una variazione significativa nei livelli della risposta; dopo aver controllato per i due precedenti *pathway*, la relazione tra esposizione e risposta che prima risultava essere significativa non dovrebbe più esserlo, con il più alto grado di mediazione che si verifica quando l'effetto è nullo.

Per verificare questi criteri, Judd e Kenny (1981) suggeriscono di stimare una serie di modelli di regressione: il modello che regredisce il mediatore sull'esposizione (Formula 4.3); quello che regredisce la risposta sull'esposizione (Formula 4.4) e infine quello che regredisce la risposta sull'esposizione e sul mediatore contemporaneamente (Formula 4.5).

$$M = \beta_0 + \beta_1 A + \varepsilon \quad (4.3)$$

$$Y = \gamma_0 + \gamma_1 A + \delta \quad (4.4)$$

$$Y = \theta_0 + \theta_1 A + \theta_2 M + \xi \quad (4.5)$$

Per stabilire la presenza di mediazione una volta stimati e testati i coefficienti dei modelli devono essere verificate quattro condizioni: nel primo modello (Formula 4.3) l'esposizione deve influire sul mediatore; nel secondo modello (Formula 4.4) l'esposizione deve avere un effetto sulla risposta; nel terzo modello (Formula 4.5), anche controllando per l'esposizione, il mediatore deve influire sulla risposta e se queste condizioni sono rispettate allora l'effetto dell'esposizione sull'outcome nel terzo modello deve essere inferiore all'effetto stimato nel secondo modello, trovandosi nella situazione di mediazione

perfetta quando l'effetto stimato dopo aver controllato per il mediatore è nullo.

Tra queste condizioni, le prime due solitamente si ritiene siano abbastanza importanti per stabilire la presenza di un mediatore, mentre le ultime due sono state spesso criticate (MacKinnon, Introduction to Statistical Mediation Analysis, 2008). La terza condizione viene spesso criticata poiché si ritiene che non sia necessario che la relazione tra esposizione e outcome risulti significativa affinché sia presente una mediazione. Fare così non prende in considerazione casi di mediazione inconsistente, cioè quando l'effetto diretto e quello mediato hanno segno opposto e quindi l'associazione tra A e Y può sembrare quasi nulla. Anche l'ultima viene spesso criticata poiché non distingue situazioni di mediazione completa da situazioni di mediazione parziale.

Inoltre, nell'analisi pratica, questi criteri vengono valutati tramite l'applicazione di test di significatività nonostante sia possibile avere un'associazione anche nel caso in cui non si ottenga la significatività, risultando così in un approccio ancora più problematico (VanderWeele, 2015). Quello proposto da Baron e Kenny è un approccio parametrico che permette di calcolare effetti diretti e indiretti⁸.

Considerando come prima A il trattamento, Y l'outcome, M il mediatore e C le covariate e considerando per comodità M e Y continui, possono essere specificati i seguenti modelli di regressione,

$$E(M|A = a, C = c) = \beta_0 + \beta_1 a + \beta_2' c \quad (4.6)$$

$$E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c \quad (4.7)$$

nei quali i coefficienti contrassegnati con un apice possono anche non essere presenti.

⁸ Altri autori avevano precedentemente proposto metodi simili (Alwin e Hausen, 1975; Hyman, 1955; Judd e Kenny, 1981; Sobel, 1982).

Secondo Baron e Kenny l'effetto diretto poteva essere quantificato tramite il valore assunto da θ_1 mentre quello indiretto dal prodotto $\beta_1\theta_1$. L'effetto totale invece rappresenta semplicemente l'effetto complessivo dell'esposizione o trattamento sull'outcome, ottenuto come somma dei due effetti: $\theta_1 + \beta_1\theta_1$.

Considerando per semplicità i modelli delle Formule 4.3 – 4.5, le espressioni appena utilizzate per esprimere le varie tipologie di effetto possono essere esplicitate facilmente. Se si ignorano i *pathway* casuali della *Figura 4.1* e si regredisce Y su A con un modello lineare come $Y = \gamma_0 + \gamma_1A + \delta$ (Formula 4.4), il coefficiente γ_1 rappresenta l'effetto totale di A su Y , cioè l'effetto complessivo dell'esposizione sulla risposta senza fare riferimento ad alcun percorso particolare.

Tuttavia, una volta ipotizzato un diagramma causale come quello della *Figura 4.1*, le relazioni tra le variabili A , Y e M possono essere espresse tramite i modelli $M = \beta_0 + \beta_1A + \varepsilon$ (Formula 4.3) e $Y = \theta_0 + \theta_1A + \theta_2M + \xi$ (Formula 4.5). In queste equazioni, infatti, β_1 rappresenta l'effetto del percorso $A \rightarrow M$, θ_1 rappresenta l'effetto del percorso $A \rightarrow Y$ e θ_2 rappresenta l'effetto del percorso $M \rightarrow Y$.

Sostituendo la Formula 4.3 all'interno della 4.5 si ottiene

$$Y = (\theta_0 + \theta_2\beta_0) + (\theta_1 + \theta_2\beta_1)A + (\theta_2\varepsilon + \xi)$$

e comparando questa con il modello della Formula 4.4 si ottiene l'uguaglianza $\gamma_1 = \theta_1 + \theta_2\beta_1$ per l'effetto complessivo di A su Y .

Poiché θ_1 rappresenta l'effetto diretto, allora la seconda componente $\theta_2\beta_1$ rappresenta l'effetto indiretto che può essere visto intuitivamente come il prodotto delle due parti che compongono il *pathway* $A \rightarrow M \rightarrow Y$.

Per permettere a questi valori di assumere un'interpretazione causale da un punto di vista controfattuale, avendo utilizzato dei coefficienti di regressione, sono necessarie delle assunzioni forti. L'approccio sarà generalizzato nel paragrafo successivo insieme alla specificazione delle assunzioni necessarie (Formule 4.13 – 4.16).

Il secondo metodo tradizionale che rientra nella visione dei modelli ad equazioni strutturali e permette di ottenere delle stime di effetti diretti e

indiretti è il *difference method* (Susser, 1973). In quest'ultimo, a differenza del metodo precedente, sono costruiti due modelli per l'outcome, uno con e l'altro senza mediatore:

$$E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c \quad (4.7)$$

$$E(Y|A = a, C = c) = \theta_0^\dagger + \theta_1^\dagger a + \theta_4^\dagger c \quad (4.8)$$

Il valore dell'effetto diretto rimane come nel *product method* dato da θ_1 mentre quello dell'effetto indiretto è dato dalla riduzione dell'effetto del trattamento sull'outcome dovuto all'inserimento del mediatore: $\theta_1^\dagger - \theta_1$. L'effetto totale è il coefficiente dell'esposizione nella regressione di A sulla risposta, cioè θ_1^\dagger . Anche con il *difference method* vale la decomposizione degli effetti che permette di ottenere l'effetto complessivo come somma dell'effetto indiretto e dell'effetto diretto.

È stato dimostrato che i due metodi, in caso di risposta continua su scala differenza, portano ai medesimi risultati (MacKinnon, et al., 1995), ma lo stesso non vale nel caso di risposte dicotomiche, per le quali viene applicata una regressione logistica invece dei minimi quadrati (MacKinnon e Dwyer, 1993).

Con questi metodi, quando nel modello viene inserito il termine di interazione tra esposizione e mediatore ci si trova in una situazione particolare di *moderated mediation* (Preacher, et al., 2007).

Formalmente ci si trova in una situazione di *moderated mediation* quando l'effetto indiretto dipende dai livelli di qualche variabile, cioè quando le relazioni di mediazione sono contingentate ai livelli di un moderatore.

In questo caso particolare è il trattamento stesso che funge da moderatore e fa sì che l'effetto di M possa variare con l'esposizione, interagendo con esso e quindi segnando la necessità di aggiungere all'interno dei modelli un termine di interazione. Sono state ottenute (Preacher, et al., 2007) formule per il calcolo degli effetti diretti e indiretti anche in queste situazioni, ma

presentano lo svantaggio di non formare l'effetto totale se sommate, proprietà molto importante per capire la porzione di effetto che effettivamente viene mediata.

Queste tecniche non pongono nemmeno particolare attenzione alle ipotesi sul confondimento, necessarie per dare un'interpretazione causale degli effetti. L'approccio controfattuale sopperisce a questa mancanza permettendo il calcolo degli effetti che sommano a quello totale anche in presenza di interazioni e non linearità, quando invece i metodi tradizionali mostrano molte limitazioni (Robins e Greenland, 1992; Pearl, 2001).

4.3.2 Approccio controfattuale

L'approccio controfattuale dell'inferenza causale può essere applicato per ampliare la definizione di Baron e Kenny anche nelle situazioni più problematiche, ponendo assunzioni precise e forti sul confondimento e permettendo di semplificare dal punto di vista formale il modo in cui vengono espressi e calcolati gli effetti di interesse.

Considerando un modello di mediazione semplice come quello della *Figura 4.1*, si assume un nesso causale tra esposizione e mediatore. Così facendo, come è stato precedentemente descritto per la risposta, anche per il mediatore possono essere identificati due "outcome" potenziali, ottenendo un valore di M , M_a , oppure un altro, M_{a^*} , al variare del livello dell'esposizione.

Siano quindi Y_a e M_a i valori assunti dall'outcome e dal mediatore per un soggetto con esposizione A pari ad a . Sia Y_{am} il valore della risposta quando l'esposizione viene fissata al valore a e il mediatore al valore m .

Si assume inoltre che $Y_a = Y_{aM_a}$, cioè che il valore assunto dalla risposta è lo stesso nella situazione in cui $A = a$ e nella situazione in cui il mediatore viene fissato al valore che assumerebbe quando $A = a$; si fa riferimento a questa assunzione come *composition* (VanderWeele, 2015). Viene fatta anche un'altra assunzione, la *consistency assumption* (VanderWeele, 2015), per cui quando $A = a$, i controfattuali Y_a e M_a equivalgono ai valori osservati Y e M e quando $A = a$ e $M = m$ il controfattuale Y_{am} equivale ad Y .

Dal punto di vista pratico quest'ultima assunzione indica il fatto che per ogni individuo si è in grado di osservare solamente l'outcome corrispondente al livello di esposizione a cui l'individuo è effettivamente stato sottoposto. Considerando per semplicità il caso con esposizione dicotomica, per un soggetto con $A = 0$ è possibile osservare solamente Y_0 e non è possibile conoscere il valore di Y_1 . Analogamente per un soggetto con $A = 1$ è possibile osservare solamente Y_1 e non Y_0 .

Robins e Greenland (1992) e Pearl (2001) hanno mostrato delle formule per calcolare gli effetti di interesse seguendo l'approccio controfattuale con la notazione appena mostrata:

- **Controlled Direct Effect:** esprime di quanto cambia in media l'outcome quando il mediatore M rimane fisso al valore m uniformemente nella popolazione ma l'esposizione passa da a a a^* , cioè

$$CDE(m) = E(Y_{am} - Y_{a^*m} | c) \quad (4.9)$$

Considerando il caso semplice con un mediatore dicotomico, cioè $M = 0$ oppure $M = 1$, è quindi possibile identificare due CDE , che esprimono il valore diretto dell'esposizione sulla risposta mantenendo il mediatore fisso ad uno dei due valori possibili.

- **Natural Direct Effect:** esprime di quanto cambierebbe l'outcome nella situazione in cui A è fissato ad a rispetto a quando l'esposizione è a^* , ma per ogni individuo il mediatore fosse fissato al livello che avrebbe assunto in assenza di esposizione, cioè

$$NDE = E(Y_{aM_{a^*}} - Y_{a^*M_{a^*}} | c) \quad (4.10)$$

È l'effetto dell'esposizione sull'outcome una volta disabilitato il *pathway* tra esposizione e mediatore.

- **Natural Indirect Effect:** esprime il cambiamento medio della risposta quando l'esposizione è fissata al valore a , ma il mediatore cambia dal

valore che assumerebbe quando l'esposizione è a^* a quello che assumerebbe quando è a , cioè

$$NIE = E(Y_{aM_a} - Y_{aM_{a^*}} | c) \quad (4.11)$$

È l'effetto dell'esposizione sulla risposta quando il mediatore cambia e quindi escludendo il *pathway* diretto tra A e Y .

Gli effetti possono essere analogamente definiti senza il condizionamento alle covariate e l'effetto totale può essere espresso come la somma del natural direct effect e del natural indirect effect:

$$\begin{aligned} TE &= E(Y_a - Y_{a^*} | c) = E(Y_{aM_a} - Y_{a^*M_{a^*}} | c) \\ &= E(Y_{aM_a} - Y_{aM_{a^*}} | c) + E(Y_{aM_{a^*}} - Y_{a^*M_{a^*}} | c) \end{aligned} \quad (4.12)$$

Definendo in questo modo gli effetti di interesse, rispetto agli approcci tradizionali, è possibile evitare di fare assunzioni sul modello, sulla forma funzionale e sulla presenza di interazione, come sarà mostrato in seguito (Formule 4.20 – 4.23). Questo è possibile poiché definendo gli effetti come differenze di valori attesi, la forma funzionale e la natura della risposta e del mediatore, sono presi in considerazione da essi, senza andare ad agire sul calcolo finale degli effetti. Tuttavia, per permettere che gli effetti siano stimati attraverso dei modelli di regressione, risultino identificati e abbiano un'interpretazione causale, sono necessarie diverse assunzioni aggiuntive (VanderWeele, 2015) rispetto a quelle già esplicitate (*composition, consistency e no-unmeasured-confounding assumption*):

1. Assenza di confondenti non misurati nella relazione tra l'esposizione e l'outcome

$$Y_{am} \perp A | C \quad (4.13)$$

2. Assenza di confondenti non misurati nella relazione tra il mediatore e l'outcome

$$Y_{am} \perp M|A, C \quad (4.14)$$

3. Assenza di confondenti non misurati nella relazione tra l'esposizione e il mediatore

$$M_a \perp A|C \quad (4.15)$$

4. Assenza di confondenti causati dall'esposizione nella relazione tra il mediatore e l'outcome

$$Y_{am} \perp M_{a^*}|C \quad (4.16)$$

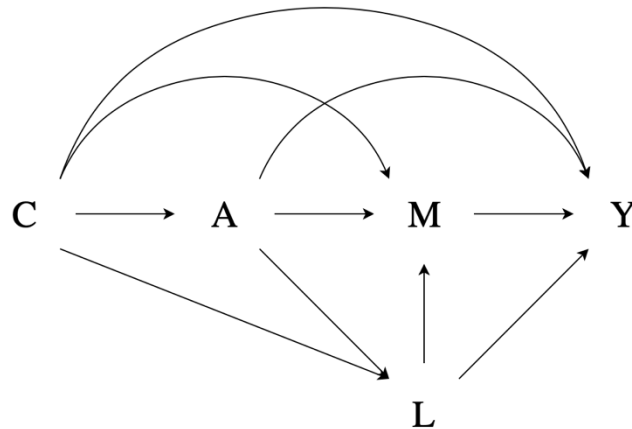


Figura 4.5 – Modello di mediazione in cui è presente un confondente (L) nella relazione tra mediatore e outcome influenzato dall'esposizione

Per la stima del $CDE(m)$ sono necessarie solamente le prime due assunzioni, mentre per la stima di NDE e NIE bisogna aggiungere anche le ultime due. È importante sottolineare che solamente la prima e la terza assunzione verrebbero soddisfatte dalla randomizzazione, che quindi in questo contesto si dimostrerebbe non sufficiente.

Le prime tre assunzioni inoltre richiedono a loro volta l'assunzione di temporalità degli eventi, fondamentale per dare un'interpretazione causale agli effetti.

Sotto queste assunzioni, gli effetti possono essere quindi stimati dai dati nella maniera seguente:

- $$CDE(m) = E(Y_{am} - Y_{a^*m}|c) = E(Y|a, m, c) - E(Y|a^*, m, c) \quad (4.17)$$

- $$NDE = E(Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c) = \sum_m \{E(Y|a, m, c) - E(Y|a^*, m, c)\}P(m|a^*, c) \quad (4.18)$$

- $$NIE = E(Y_{aM_a} - Y_{aM_{a^*}}|c) = \sum_m E(Y|a, m, c)\{P(M|a, c) - P(M|a^*, c)\} \quad (4.19)$$

e alcune volte si fa riferimento alla (4.19) con il termine *mediation formula* (Pearl, 2012).

Per calcolare gli effetti a livelli di popolazione basta applicare le proprietà della torre⁹.

Specificando dei modelli per calcolare $E(Y|a, m, c)$ e $P(M|a, c)$ è possibile calcolare analiticamente con delle semplici regressioni gli effetti a cui si è interessati con una maggiore flessibilità, evitando i problemi in cui solitamente si incorre con i metodi tradizionali in caso di interazioni e non linearità.

Riprendendo l'esempio di Baron e Kenny, mantenendo il modello per il mediatore come prima (Formula 4.6) e modificando solo quello per la risposta (Formula 4.7), è possibile aggiungere il termine di interazione tra il mediatore e l'esposizione di interesse come segue

$$E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c \quad (4.20)$$

⁹ Proprietà della torre. Siano X e Y variabili aleatorie discrete con densità congiunta $\rho_{X,Y}$, allora considerando $E(X|Y = \cdot) = f(\cdot)$,

$$E(X) = E(E(X|Y)) = \sum_{y \in \Omega_Y} E(X|Y = y)P(Y = y)$$

e definire in questo modo i tre effetti mediante i coefficienti delle regressioni come:

- $$CDE(m) = E(Y_{am} - Y_{a^*m}|c) = (\theta_1 + \theta_3 m)(a - a^*) \quad (4.21)$$

- $$NDE = E(Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c) = (\theta_1 + \theta_3\beta_0 + \theta_3\beta_1 a^* + \theta_3\beta_2' c)(a - a^*) \quad (4.22)$$

- $$NIE = E(Y_{aM_a} - Y_{aM_{a^*}}|c) = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*) \quad (4.23)$$

Gli effetti sono sempre condizionati al valore c assunto dalle covariate C e fissandolo al valore medio $E(C)$ è possibile ottenere l'effetto marginale in media per l'intera popolazione.

Una proprietà importante di NDE e NIE è che l'effetto totale può essere determinato dalla somma di loro due, cioè $TE = NDE + NIE$ e grazie all'approccio controfattuale questa proprietà viene mantenuta nonostante la presenza del termine di interazione (Pearl, 2001).

Per il calcolo degli intervalli di confidenza degli effetti gli standard error possono essere calcolati sempre utilizzando il metodo delta oppure le tecniche di *bootstrap*.

Le formula fornite sono molto generiche proprio poiché non presuppongono nessun modello statistico particolare e per tale motivo viene spesso fatto riferimento a loro come a “non parametriche”.

Quanto appena visto può essere ampliato a situazioni in cui l'outcome, il mediatore o entrambi sono variabili binarie e in caso di modelli non lineari con variabili di conteggio, per cui è necessario applicare dei modelli di regressione logistica oppure modelli di Poisson o binomiale negativa (VanderWeele, 2015).

I metodi tradizionali e l'approccio controfattuale coincidono nel caso di modelli lineari o log-lineari quando non c'è interazione e in modelli logistici

quando l'outcome è raro e non c'è interazione (VanderWeele e Vansteelandt, 2009; 2010).

Negli approcci tradizionali, quando viene utilizzata la modellazione con equazioni strutturali, la relazione della *Figura 4.1* viene spesso espressa attraverso modelli lineari additivi, nei quali il termine di interazione tra esposizione e mediatore viene escluso, l'outcome viene solitamente assunto continuo e vengono calcolati gli effetti diretti e indiretti comparando modelli lineari per variabili continue, senza averne dato delle definizioni formali a priori. Quanto appena detto giustifica la nascita e l'utilizzo dell'approccio controfattuale che permette di superare i limiti dei SEM e la loro mancata costruzione teorica, consentendone un'applicazione generale.

La definizione degli effetti con i modelli ad equazioni strutturali può essere chiamata *model driven*, in quanto pone maggior enfasi sulla creazione dei modelli e ragiona su di essi per arrivare ad individuare gli effetti. Al contrario, l'approccio controfattuale pone delle basi teoriche fornendo delle definizioni degli effetti a priori e prevede la stima dei modelli solamente in un secondo momento. Così facendo è possibile superare numerose difficoltà incontrate dai metodi tradizionali, permettendone un'applicazione generale anche in situazioni che necessitano l'utilizzo di modelli differenti da quello lineare oppure l'inclusione di termini di interazione.

Spesso la scelta di includere o meno l'interazione può essere difficile anche quando si utilizza l'approccio controfattuale e nelle analisi che verranno svolte in seguito si seguirà quanto suggerito da VanderWeele (2015), cioè di includerla sempre ed eliminarla solamente se la stima degli effetti diretti e indiretti non cambia molto a seconda che l'interazione venga inserita oppure tolta. In letteratura non viene indicata una misura soglia che permette di valutare il cambiamento delle stime e prendere una decisione in merito all'inserimento dell'interazione nei modelli.

È importante mettere in evidenza come la scelta di ignorare l'interazione quando in realtà è presente può portare a conclusioni sbagliate e che la sua inclusione non va determinata solamente sulla base della significatività del

suo coefficiente, poiché anche nel caso in cui non risulti significativo, la sua presenza può aumentare la potenza con cui viene identificato l'effetto indiretto.

4.3.3 Applicazione

Il software che si utilizza per svolgere le analisi successive è SAS, quindi l'applicazione delle formule fino ad ora viste in maniera puramente teorica fa riferimento alla sua notazione.

In passato l'analisi empirica dei dati applicando modelli di mediazione era svolta mediante l'utilizzo di macro, che permettevano l'inserimento del termine di interazione tra esposizione e mediatore (Valeri e VanderWeele, 2013) e lasciavano una certa libertà nella scelta della natura degli outcome, che potevano richiedere l'applicazione di modelli di regressione lineare, logistici, log-lineari, di Poisson o binomiale negativa, e dei mediatori, che potevano essere continui o binari. Al variare della tipologia di outcome anche gli effetti venivano riportati nella scala corrispondente: scala differenza per outcome continuo, scala di rischi relativi quando si usa il logaritmo come funzione legame, scala di odds ratio quando si usa il logit come legame e come rapporto di tassi se si usa un outcome con distribuzione di Poisson o binomiale negativa.

Dal 2017 tuttavia viene messa a disposizione su SAS¹⁰ anche una procedura specifica, che amplia le macro di cui si parlava prima e permette l'applicazione delle metodologie mostrate nei paragrafi precedenti dal punto di vista teorico: la *proc causalmed*.

Invece di utilizzare i modelli lineari, come nei metodi tradizionali, le nuove procedure che stimano gli effetti fanno ricorso a modelli lineari generalizzati per i modelli relativi alla risposta e per quelli relativi al mediatore (Formule 4.6 e 4.20); nel caso specifico di SAS viene seguito l'approccio basato sulla regressione di VanderWeele (2014).

¹⁰ SAS Institute Inc. 2017. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.

Non vengono purtroppo supportati modelli con risposta nominale oppure ordinale con più di due livelli.

Per quanto riguarda i modelli lineari generalizzati, utilizzando la notazione di SAS, è come se fossero utilizzate delle *proc reg* in caso di outcome o mediatore continuo, delle *proc logistic* in caso di outcome o mediatore binario e delle *proc genmod* nel caso in cui si trattasse di un modello di Poisson, binomiale negativa oppure log – lineare.

Se si considera ϑ il vettore che contiene tutti i parametri nei modelli per il mediatore e per la risposta, sotto le assunzioni di corretta specificazione dei modelli e il rispetto delle assunzioni (Formule 4.13 – 4.16), gli effetti causali (*ef*) sono semplicemente delle funzioni di ϑ condizionatamente al valore delle covariate: $g_{ef}(\vartheta|C = c)$.

Per campioni casuali possono quindi essere stimati gli effetti causali della mediazione tramite il metodo della massima verosimiglianza e, una volta stimata la matrice di varianza e covarianza di $\hat{\vartheta}$, è possibile ottenere applicando il metodo delta gli standard error e gli intervalli di confidenza degli effetti. In alternativa possono essere utilizzate tecniche di bootstrap.

Le formule che vengono utilizzate per le stime degli effetti in caso di outcome continuo sono quelle delle Formule 4.21 – 4.23, mentre per le altre tipologie di outcome si procede in maniera analoga e sono mostrate formule corrispondenti da Valeri e VanderWeele (2013).

In caso di esposizione continua i risultati vengono ottenuti semplicemente valutando l'effetto ad una variazione unitaria del valore medio dell'esposizione, confrontandone i valori $a = \bar{a} + 0.5$ e $a^* = \bar{a} - 0.5$.

Per chiarire ulteriormente il passaggio dalle stime dei coefficienti a quelle degli effetti è riportato di seguito un esempio applicativo¹¹.

Si tratta di uno studio per valutare se l'incoraggiamento da parte dei genitori (*A*) ha un effetto sullo sviluppo cognitivo dei bambini (*Y*). Si vuole capire se

¹¹ SAS Institute Inc. 2017. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.

l'effetto sia in parte dovuto alla motivazione ad imparare dei bambini (M). I confondenti che vengono presi in considerazione sono il numero di membri della famiglia (C_1) e lo stato sociale del bambino (C_2), cioè una misura riassuntiva del reddito familiare, dell'occupazione dei genitori e del loro titolo di studio.

Mediante l'applicazione della *proc causalmed* sono ottenute le stime dei modelli per la risposta e per l'outcome, riportati nelle *Tabelle 4.1* e *4.2*.

Il parametro *scale* rappresenta un parametro di dispersione per una distribuzione data e, nel caso di distribuzione Normale (*default*), è una stima della deviazione standard dell'outcome.

	Coefficiente	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	θ_0	-129.4	39.5990	-207.0	-51.7802	10.6771	0.0011
Incoraggiamento	θ_1	2.2828	1.3022	-0.2695	4.8351	3.0731	0.0796
Motivazione	θ_2	1.8176	0.9763	-0.0960	3.7312	3.4657	0.0627
Interazione: Incoraggiamento e Motivazione	θ_3	0.0533	0.0325	-0.0104	0.1171	2.6915	0.1009
Membri della famiglia	θ_4	-0.1040	0.0628	-0.2271	0.0190	2.7455	0.0975
Stato sociale	θ_5	0.0445	0.0915	-0.1350	0.2239	0.2360	0.6271
Scale		0.1407	0.0315	0.0907	0.2180		

Tabella 4.1 – Stima del modello per la risposta

$$E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c_1 + \theta_5 c_2$$

	Coefficiente	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	β_0	10.6851	6.0716	-1.2151	22.5852	3.0970	0.0784
Incoraggiamento	β_1	0.7296	0.2510	0.2377	1.2215	8.4516	0.0036
Membri della famiglia	β_2	-0.1295	0.0647	-0.2564	-0.00264	4.0031	0.0454
Stato sociale	β_3	0.1271	0.1146	-0.0976	0.3518	1.2283	0.2677
Scale		0.2452	0.0548	0.1582	0.3801		

Tabella 4.2 – Stima del modello per il mediatore

$$E(M|A = a, C = c) = \beta_0 + \beta_1 a + \beta_2 c_1 + \beta_3 c_2$$

Le stime degli effetti ottenute dalla procedura sono riportate nella *Tabella 4.3* e possono essere ottenute tramite l'utilizzo delle Formule 4.21 – 4.23.

Sapendo che i valori medi del mediatore, dell'esposizione, del numero di membri della famiglia e del livello di stato sociale sono rispettivamente $\bar{m} = 39$, $\bar{a} = 39.43$, $\bar{c}_1 = 3.2$ e $\bar{c}_2 = 25.7$, e confrontando i valori di esposizione $a = \bar{a} + 0.5 = 35.4$ e $a^* = \bar{a} - 0.5 = 34.4$, è possibile ottenere le stime degli effetti come:

- $CDE(\bar{m}) = (\theta_1 + \theta_3 \bar{m})(a - a^*) = (2.2828 + 0.0533 \cdot 39)(35.4 - 34.4) = 4.36$;
- $NDE = (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 \bar{c}_1 + \theta_3 \beta_3 \bar{c}_2)(a - a^*) = [2.2828 + 0.0533(10.6851 + 0.7296 \cdot 34.4 - 0.1295 \cdot 3.2 + 0.1271 \cdot 25.7)](35.4 - 34.4) = 4.34$;
- $NIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*) = (1.8176 \cdot 0.7296 + 0.0533 \cdot 0.7296 \cdot 35.4)(35.4 - 34.4) = 2.7$;
- $TE = NDE + NIE = 4.34 + 2.7 = 7.04$;
- La percentuale di effetto attribuibile alla mediazione ottenuta semplicemente come $\frac{NIE}{TE} \cdot 100 = \frac{2.7}{7.04} \cdot 100 = 38.35$;
- La percentuale di effetto eliminata indica la percentuale dell'effetto dell'esposizione sull'outcome che può essere eliminata ponendo il mediatore ad un livello fissato ($M = m$) ed è ottenuta come $PE(m) =$

$100 \cdot (TE - CDE(m))/TE$. In questo caso è calcolata come $PE(\bar{m}) = 100 \cdot (TE - CDE(\bar{m}))/TE = 100 \cdot (7.04 - 4.36)/7.04 = 38.06$.

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value
Effetto Totale	7.0476	0.9409	5.2036	8.8917	7.49	<.0001
Controlled Direct Effect (CDE)	4.3632	0.1965	3.9781	4.7482	22.21	<.0001
Natural Direct Effect (NDE)	4.3437	0.1982	3.9553	4.7321	21.92	<.0001
Natural Indirect Effect (NIE)	2.7039	0.9498	0.8423	4.5655	2.85	0.0044
Percentuale di effetto dovuta alla mediazione	38.3666	8.5344	21.6395	55.0936	4.50	<.0001
Percentuale di effetto dovuta all'interazione	0.2761	0.1839	-0.08437	0.6366	1.50	0.1333
Percentuale di effetto eliminata	38.0905	8.4522	21.5244	54.6565	4.51	<.0001

Tabella 4.3 – Stime degli effetti causali

È stato mostrato (VanderWeele, 2014) che l'effetto totale dell'esposizione sull'outcome, in presenza di interazione tra esposizione e mediatore, può essere scomposto in quattro componenti: l'effetto dell'esposizione in assenza del mediatore; l'effetto interattivo quando il mediatore assume il valore che avrebbe assunto in assenza di esposizione; un'interazione mediata e un effetto di mediazione puro. La decomposizione prende il nome di *four-way decomposition* e può essere espressa analiticamente come

$$TE = CDE + INT_{ref} + INT_{med} + PIE$$

dove

- CDE (Controlled Direct Effect) è la componente non dovuta né alla mediazione né all'interazione;
- INT_{ref} (Reference Interaction) è la componente dovuta solamente all'interazione e non alla mediazione;

- INT_{med} (Mediated Interaction) è la componente dovuta sia all'interazione sia alla mediazione;
- PIE (Pure Indirect Effect) è la componente dovuta solamente alla mediazione e non all'interazione.

Considerando i modelli delle Formule 4.6 e 4.20 è possibile definire (VanderWeele, 2014) le diverse componenti sulla base dei coefficienti dei modelli:

- $CDE(m) = (\theta_1 + \theta_3 m)(a - a^*)$
- $PIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a^*)(a - a^*)$
- $INT_{med} = \theta_3 \beta_1 (a - a^*)(a - a^*)$
- $INT_{ref}(m) = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)\}(a - a^*) - (\theta_1 + \theta_3 m)(a - a^*) = \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c - m)(a - a^*)$, dato dalla differenza tra il PDE (Pure Direct Effect) e il CDE .

Mediante l'applicazione della *four-way decomposition* è quindi possibile ottenere l'ultima componente rimasta della *Tabella 4.3*, cioè la percentuale di effetto dovuta all'interazione: $PAI = 100 \cdot (INT_{ref} + INT_{med})/TE$.

Quindi,

- $INT_{ref}(\bar{m}) = \theta_3(\beta_0 + \beta_1 a^* + \beta_2 \bar{c}_1 + \beta_3 \bar{c}_2 - \bar{m})(a - a^*) = 0.0533 \cdot (10.6851 + 0.7296 \cdot 34.4 - 0.1295 \cdot 3.2 + 0.1271 \cdot 25.7 - 39)(35.4 - 34.4) = -0.019$
- $INT_{med} = \theta_3 \beta_1 (a - a^*)(a - a^*) = 0.0533 \cdot 0.7296 \cdot (35.4 - 34.4)(35.4 - 34.4) = 0.0388$
- $PAI = 100 \cdot (INT_{ref} + INT_{med})/TE = 100 \cdot (-0.019 + 0.0388)/7.04 = 0.28$

Esistono altre decomposizioni possibili (*two-way decomposition*) che vengono calcolate dalla procedura di SAS, ma non essendo di immediato interesse per l'analisi non vengono riportate nella presente, si veda in merito VanderWeele (2014).

Le differenze nei decimali sono da attribuire al diverso arrotondamento dei valori per svolgere i calcoli.

4.4 Mediatori multipli

Si vuole arrivare ad ampliare questi concetti di base in modo da rendere possibile lo svolgimento di un'analisi di mediazione che comprende il tempo. Per fare questo nel modo più semplice possibile è necessario trattare prima il problema della presenza di più mediatori.

Anche nel caso in cui si fosse interessati ad un solo mediatore è importante considerare questa situazione poiché altrimenti si potrebbe correre il rischio di violare l'assunzione (4.16).

Una prima possibilità potrebbe essere quella di considerare un mediatore per volta, ma questo non risulta essere efficace in quanto necessita che i mediatori non si influenzino tra di loro; inoltre questo potrebbe portare ad ottenere degli effetti la cui somma risulta superiore al totale anche se la direzione della mediazione è la stessa.

Ci sono due approcci con cui si può procedere in queste situazioni: il primo si basa sulla regressione mentre il secondo su un sistema di pesi.

4.4.1 Metodo basato sulla regressione

Si considera un vettore di mediatori $\underline{M} = (M^{(1)}, \dots, M^{(K)})$. Le definizioni precedentemente date per $CDE(m)$, NDE e NIE possono essere mantenute allo stesso modo, solo facendo riferimento ad un vettore di mediazione e non a una singola variabile.

È inoltre necessario che le quattro assunzioni riguardanti il confondimento valgano per tutti i K mediatori:

$$1. \quad Y_{am} \perp A|C \tag{4.24}$$

$$2. \quad Y_{am} \perp \underline{M}|A, C \tag{4.25}$$

$$3. \quad \underline{M}_a \perp A|C \tag{4.26}$$

$$4. \quad Y_{\underline{a}\underline{m}} \perp \underline{M}_{a^*} | C \quad (4.27)$$

In questo contesto possono essere gestite con facilità violazioni alla quarta assunzione in quanto basta che la variabile interessata venga inclusa nel vettore di mediatori affinché l'assunzione torni ad essere rispettata.

Come in precedenza (Paragrafo 4.3) possono essere esplicitati due modelli di regressione, uno per il mediatore e uno per la risposta. Considerando il caso più semplice in cui Y e $M^{(i)}$, $i = 1, \dots, K$ sono continui, in assenza di interazione e sotto le assunzioni (4.24) – (4.27) è possibile ottenere:

$$E(Y|a, \underline{m}, c) = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \dots + \theta_2^{(K)} m^{(K)} + \theta_4' c \quad (4.28)$$

$$E(M^{(i)}|a, c) = \beta_0^{(i)} + \beta_1^{(i)} a + \beta_2^{(i)} c, \quad i = 1, \dots, K \quad (4.29)$$

Allora, come mostrato da VanderWeele e Vansteelandt (2013),

$$CDE(m) = \theta_1 (a - a^*) \quad (4.30)$$

$$NDE = \theta_1 (a - a^*) \quad (4.31)$$

$$NIE = (\beta_1^{(1)} \theta_2^{(1)} + \dots + \beta_1^{(K)} \theta_2^{(K)}) (a - a^*) \quad (4.32)$$

Già da qui si può comprendere meglio quanto detto prima in quanto l'effetto indiretto è diverso rispetto a quello che si otterrebbe sommando gli effetti indiretti ottenuti per ogni $M^{(i)}$, $i = 1, \dots, K$. Questo succede poiché seguendo questo approccio si avrebbe una regressione diversa per la risposta a seconda del mediatore che si sta considerando. I due metodi coincidono solamente se i mediatori sono tutti indipendenti condizionatamente all'esposizione e alle covariate (VanderWeele, 2015).

La decisione di considerare tutti i mediatori separatamente causerebbe inoltre la violazione dell'assunzione (4.27) e questo porterebbe alla non identificabilità degli effetti.

Proseguendo senza fare analisi separate per ogni mediatore, può essere gestita qualsiasi tipologia di interazione tra esposizione e mediatore oppure anche tra mediatori.

Poiché le interazioni che possono essere inserite potenzialmente possono avere una diversa natura, per il calcolo degli standard error degli effetti, in queste situazioni si consiglia l'utilizzo del *bootstrap* piuttosto che il metodo delta.

In presenza di interazione tra esposizione e un mediatore il modello per la risposta diventa

$$E(Y|a, \underline{m}, c) = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \dots + \theta_2^{(K)} m^{(K)} + \theta_3^{(i)} a m^{(i)} + \theta_4' c \quad (4.33)$$

e gli effetti

$$CDE(m) = (\theta_1 + \theta_3^{(i)} m^{(i)})(a - a^*) \quad (4.34)$$

$$NDE = \{\theta_1 + \theta_3^{(i)}(\beta_0^{(i)} + \beta_1^{(i)} a^* + \beta_2^{(i)'} c)\}(a - a^*) \quad (4.35)$$

$$NIE = (\beta_1^{(1)} \theta_2^{(1)} + \dots + \beta_1^{(K)} \theta_2^{(K)} + \theta_3^{(i)} \beta_1^{(i)} a)(a - a^*) \quad (4.36)$$

In questo modo si possono gestire qualsiasi numero di interazioni tra esposizione e mediatori e, anche in presenza di mediatori dicotomici, gli effetti possono essere aggiustati in maniera piuttosto semplice (VanderWeele, 2015). Considerando l'esposizione dicotomica e randomizzata può anche essere inserita un'interazione tra i mediatori modificando il modello per la risposta

$$E(Y|a, \underline{m}) = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \dots + \theta_2^{(K)} m^{(K)} + \theta_3^{(i)} a m^{(i)} + \tau^{(ij)} m^{(i)} m^{(j)} \quad (4.37)$$

e aggiungendone uno tra quelli dei mediatori

$$E(M^{(i)}M^{(j)}|a) = \beta_0^{(ij)} + \beta_1^{(ij)}a \quad (4.38)$$

Controlled Direct Effect e Natural Direct Effect rimangono invariati rispetto a prima (Formule 4.34 e 4.35) mentre cambia il Natural Indirect Effect diventando

$$NIE = (\beta_1^{(1)}\theta_2^{(1)} + \dots + \beta_1^{(K)}\theta_2^{(K)} + \tau^{(ij)}\beta_1^{(ij)})(a - a^*) \quad (4.39)$$

Tuttavia è importante sottolineare che con l'aggiunta di covariate questo metodo può avere dei problemi di compatibilità nei modelli per i mediatori singoli e per la loro interazione (VanderWeele e Vansteelandt, 2013), quindi solitamente si ricorre all'approccio proposto nel Paragrafo 4.4.2.

Un'altra importante limitazione di questo approccio è che in presenza di risposta binaria, esso funziona solo con mediatori continui e in assenza di interazioni tra di loro (VanderWeele, 2015); anche questo giustifica l'utilizzo del metodo basato su un sistema di pesi descritto di seguito.

4.4.2 *Inverse Probability Weighting* (VanderWeele e Vansteelandt, 2013)

Questo metodo non richiede la specificazione di modelli per i mediatori e può essere applicato con outcome di qualsiasi tipo, tuttavia è stato mostrato che funziona meglio nei casi di esposizione dicotomica o comunque con pochi livelli. Per questo motivo risolve molti problemi che portano i metodi basati sulla regressione presentati nel Paragrafo 4.4.1.

Per il calcolo degli effetti diretti e indiretti marginali (cioè in media rispetto alle covariate), sotto le quattro assunzioni specificate è necessaria fondamentalmente solo la stima di tre medie pesate¹²:

- $E(Y_{a^*M_{a^*}})$, ottenuta usando come pesi $P(A = a^*)/P(A = a^*|C = c_i)$ sui soggetti con esposizione a^* ;
- $E(Y_{aM_a})$, ottenuta usando come pesi $P(A = a)/P(A = a|C = c_i)$ sui soggetti con esposizione a ;
- $E(Y_{aM_{a^*}})$, ottenuta calcolando come prima cosa per ogni soggetto con esposizione a^* i valori predetti per la risposta se il soggetto avesse avuto esposizione a invece di a^* , ricavando $E(Y|A = a, \underline{M} = \underline{m}_i, C = c_i)$ e utilizzando poi questi per il calcolo della media pesandoli con $P(A = a^*)/P(A = a^*|C = c_i)$.

Una volta ottenute queste stime possono essere calcolati gli effetti di interesse e può essere utilizzato il *bootstrap* per il calcolo degli intervalli di confidenza. Il Natural Direct Effect sarà identificato dalla differenza $E(Y_{aM_{a^*}}) - E(Y_{a^*M_{a^*}})$ e il Natural Indirect Effect dalla differenza $E(Y_{aM_a}) - E(Y_{aM_{a^*}})$.

Questo approccio e il precedente hanno in comune il fatto che entrambi sono considerati parametrici e assumono che i modelli utilizzati siano correttamente specificati. L'approccio basato sulla regressione assume che siano correttamente specificati i modelli per la risposta e per ogni mediatore. L'approccio *IPW* invece richiede che siano correttamente specificati i modelli per la risposta e per l'esposizione (utilizzati per calcolare i pesi). Nel primo

¹² L'approccio *IPW* si basa e trova la giustificazione dell'utilizzo di quei pesi specifici nella seguente identità (VanderWeele e Vansteelandt, 2013):

$$\begin{aligned}
 E(Y_{aM_{a^*}}) &= \int E(Y|a, m, c) f(M = m|a^*, c) f(c) dm dc \\
 &= \int I(A = a^*) E(Y|a, M, c) f(M|A, c) f(c) dM dA dc \\
 &= \int \frac{I(A = a^*)}{P(A = a^*|c)} E(Y|a, M, c) f(M, A, c) dM dA dc \\
 &= E \left[\frac{I(A = a^*)}{P(A = a^*|c)} E(Y|a, M, c) \right] \\
 &= E \left[\frac{P(A = a^*)}{P(A = a^*|c)} E(Y|a, M, c) | A = a^* \right]
 \end{aligned}$$

caso non sono necessari modelli per l'esposizione, mentre nel secondo non sono necessari quelli per i mediatori (VanderWeele e Vansteelandt, 2013).

Non è stato ancora affrontato però il caso in cui si sia interessati ad un solo mediatore ma siano presenti altri mediatori precedenti a questo, influenzati dall'esposizione, che a loro volta influenzano sia il mediatore di interesse sia l'outcome. È questo il passaggio che permetterà poi di ampliare il discorso valutando i fenomeni da un punto di vista non solo trasversale ma anche longitudinale.

Nelle situazioni in cui l'assunzione (4.27) viene violata, come è stato precedentemente detto, i metodi di regressione (Paragrafo 4.3) non permettono di ottenere stime non distorte per il $CDE(m)$, mentre NDE e NIE non sono identificati. Per cercare di rimediare a queste situazioni, sono stati creati degli approcci alternativi per il calcolo degli effetti in situazioni nelle quali solitamente non sarebbero identificati.

Sia L la variabile confondente per la relazione tra mediatore e risposta, influenzata dall'esposizione (Figura 4.5). Se si volesse calcolare il $CDE(m)$ di A su Y considerando solamente i *pathway* che non includono M , è come se si cercasse l'effetto di due percorsi: $A - Y$ e $A - L - Y$ (Figura 4.6).

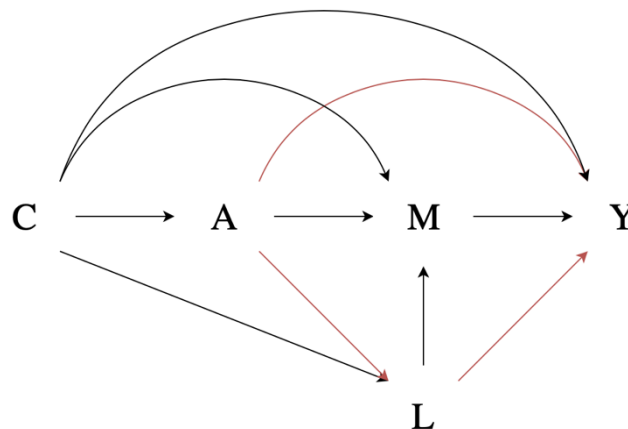


Figura 4.6 – Modello di mediazione in cui è presente un confondente (L) nella relazione tra mediatore e outcome influenzato dall'esposizione. In rosso sono evidenziati i *pathway* di interesse

In questa situazione specifica i metodi esposti nel Paragrafo 4.3 non funzionano poiché L si trova nel percorso dall'esposizione all'outcome e allo stesso tempo rappresenta un confondente per la relazione tra mediatore e outcome. A livello intuitivo, è possibile vedere che gli approcci precedentemente mostrati non portano a risultati affidabili supponendo inizialmente di inserire L insieme alle altre covariate nella regressione per la risposta. Attraverso questo controllo verrebbe bloccando il percorso $A - L - Y$ a cui si è interessati. Si potrebbe quindi pensare di non includere la variabile insieme alle altre covariate, ma questo porterebbe comunque a stime distorte in quanto non verrebbe controllato l'effetto di confondimento che L ha nella relazione tra M e Y (VanderWeele, 2015).

Per poter calcolare correttamente il $CDE(m)$ nelle situazioni in cui i metodi di regressione falliscono, si può far ricorso a due diverse categorie di modelli: i *marginal structural models* (MSM) (Robins et al., 2000; VanderWeele, 2009) e gli *structural mean models* (SMM) (Robins, 1999). Verranno approfonditi solamente i primi, ma è importante evidenziare come entrambe le categorie di modelli sono nate per gestire situazioni in cui l'esposizione è tempo-dipendente al di fuori della mediazione e poi sono state ampliate per gestire anche scenari differenti.

Rispetto a prima sarà necessario modificare l'assunzione (4.14) in quanto è fondamentale che non siano presenti confondenti non misurati nella relazione tra il mediatore e l'outcome anche condizionatamente a L , cioè

$$Y_{am} \perp M | A, C, L \quad (4.40)$$

Se la prima assunzione (4.13) e la seconda modificata (4.40) sono valide allora si può procedere con i metodi basati su MSM e SMM.

I MSM (Robins, 1996; Robins et al., 2000) sono modelli direttamente per le medie degli outcome controfattuali adattati attraverso la tecnica dell'*inverse probability weighting*, cioè utilizzando un sistema di pesi piuttosto che delle

regressioni. L'idea alla base è quella di regredire Y sull'esposizione (A) e sul mediatore (M) controllando per le covariate (C) e per le variabili (L) che violano l'assunzione (4.27), mediante un sistema di pesi.

Sono necessari per questo motivo due gruppi di pesi, uno per l'esposizione e l'altro per il mediatore.

Ipotizzando per semplicità l'esposizione dicotomica, per ogni individuo i verrà calcolato

$$\omega_i^A = P(A = a_i)/P(A = a_i|C = c_i) \quad (4.41)$$

in cui a denominatore si trova la probabilità che l'individuo riceva il trattamento che effettivamente ha ricevuto condizionatamente al valore delle covariate che può essere visto come qualcosa di simile ad un *propensity score* (Rosenbaum e Rubin, 1983).

Se i soggetti non hanno ricevuto il trattamento allora i pesi relativi all'esposizione saranno

$$\omega_i^A = (1 - P(A = a_i))/(1 - P(A = a_i|C = c_i)) \quad (4.42)$$

Il nome del metodo deriva proprio dalla quantità che viene posta a denominatore dei pesi mentre al numeratore si trova semplicemente la proporzione dei soggetti esposti e non esposti nella popolazione rispettivamente.

Allo stesso modo per quanto riguarda il mediatore, per ogni soggetto i verrà calcolato

$$\omega_i^M = P(M = m_i)/P(M = m_i|A = a_i, C = c_i, L = l_i) \quad (4.43)$$

E analogamente a quanto fatto per l'esposizione, in caso di mediatore dicotomico potranno essere utilizzati i complementi a 1 di queste probabilità per i soggetti con valore diverso del mediatore.

I pesi globali sono ottenuti come $\omega_i = \omega_i^A \omega_i^M$ e vengono utilizzati in maniera diretta nella regressione

$$E(Y|a, m) = \gamma_0 + \gamma_1 a + \gamma_2 m + \gamma_3 am \quad (4.44)$$

per ottenere una stima non distorta di

$$CDE(m) = \gamma_1(a - a^*) + \gamma_3 m(a - a^*) \quad (4.45)$$

Il controllo per le covariate viene quindi svolto mediante un sistema di pesi invece che con delle regressioni. Non vengono inserite le covariate nella regressione di Y ma viene semplicemente svolta una regressione pesata della risposta su esposizione e mediatore.

L'inclusione della variabile confondente L solamente all'interno del denominatore dei pesi per il mediatore è quello che permette di aggirare il problema che si presentava con i metodi di regressione precedentemente mostrati, per cui le stime risultavano distorte sia nel caso in cui si attuasse un controllo per L sia che non fosse attuato.

Nel calcolo degli standard error viene utilizzato lo stimatore *sandwich* che permette di tenere conto del sistema di pesi utilizzato.

Le probabilità all'interno dei pesi possono essere calcolate mediante l'utilizzo di regressioni logistiche nel caso in cui le variabili siano binarie, altrimenti si può far ricorso a delle regressioni logistiche ordinali, a delle multinomiali oppure utilizzando le distribuzioni di densità della variabile a seconda della natura dell'esposizione e del mediatore. È stato mostrato che con variabili continue questo metodo non risulta funzionare bene.

L'approccio funziona anche in assenza di L e rappresenta una valida alternativa a quello basato sulla regressione (VanderWeele, 2015).

È già stato evidenziato che in caso di violazione della quarta assunzione, *NDE* e *NIE* non sono identificati, tuttavia è possibile ottenere degli analoghi basati su interventi randomizzati sul mediatore che prendono il nome di *randomized interventional analogues* di *NDE* e *NIE*.

Generalmente ci si è sempre focalizzati su cosa sarebbe successo ad un soggetto esposto se fosse stato fissato il mediatore al valore che avrebbe assunto nel caso in cui il soggetto non fosse stato esposto. Adesso invece si considera un prospettiva differente, cioè cosa sarebbe successo ad un soggetto

esposto se venisse fissato il mediatore ad un valore estratto casualmente dalla popolazione non esposta.

Questa ottica differente permette di identificare degli analoghi di *NDE* e *NIE*.

Sia $G_{a|c}$ un'estrazione casuale dalla distribuzione di M tra coloro con $A = a$ e condizionatamente ai valori assunti dalle covariate $C = c$.

Siano a e a^* i due valori che si vogliono confrontare, allora è possibile definire

- L'effetto sull'outcome di assegnare casualmente un individuo esposto ad un valore del mediatore dalla sua distribuzione tra gli esposti rispetta a quella dei non esposti

$$\begin{aligned} NIE^R &= E\left(Y_{aG_{a|c}}\right) - E\left(Y_{aG_{a^*|c}}\right) \\ &= \sum_{l,m} E(Y|a, l, m, c)P(l|a, c)\{P(m|a, c) \\ &\quad - P(m|a^*, c)\} \end{aligned} \quad (4.46)$$

- L'effetto di essere esposto rispetto a non essere esposto con il valore del mediatore estratto entrambe le volte dalla sua distribuzione dei non esposti

$$\begin{aligned} NDE^R &= E\left(Y_{aG_{a^*|c}}\right) - E\left(Y_{a^*G_{a^*|c}}\right) \\ &= \sum_{l,m} \{E(Y|a, l, m, c)P(l|a, c) \\ &\quad - E(Y|a^*, l, m, c)P(l|a^*, c)\}P(m|a^*, c) \end{aligned} \quad (4.47)$$

- L'effetto sull'outcome di essere esposto rispetto a non esserlo con il valore del mediatore estratto casualmente rispettivamente dalla distribuzione degli esposti e quella dei non esposti

$$TE^R = E\left(Y_{aG_{a|c}}\right) - E\left(Y_{a^*G_{a^*|c}}\right) \quad (4.48)$$

Questo permette anche di mantenere la proprietà di decomposizione per cui

$$TE^R = NDE^R + NIE^R \quad (4.49)$$

Questi effetti sono identificati sotto le assunzioni (4.13), (4.15) e (4.40). La quarta assunzione quindi può essere rilasciata e non è necessaria per l'identificazione di questi effetti poiché il mediatore viene fissato ad un livello che è estratto casualmente dalla distribuzione del mediatore tra coloro con una particolare esposizione e non al livello che avrebbe assunto per quel soggetto se fosse stato esposto diversamente (VanderWeele, 2015).

Se non ci sono variabili confondenti come L e vengono rispettate le assunzioni (4.13 – 4.16) allora gli effetti coincideranno con quelli veri.

Per la stima degli intervalli di confidenza di questi effetti si usano generalmente le tecniche di *bootstrap*.

4.5 *Time – varying exposure and mediators*

4.5.1 Il *g-computational approach* e i *marginal structural models*

Quanto visto fino ad ora può essere ampliato a situazioni in cui esposizione e mediatore variano nel tempo e anche in questo contesto il problema principale riguarda l'identificabilità degli effetti, poiché trattandosi di studi longitudinali spesso la quarta assunzione viene violata.

Gli studi longitudinali tuttavia presentano il vantaggio di poter avere una maggiore certezza per quanto riguarda la temporalità degli eventi, fondamentale quando si vogliono stimare relazioni causali e fare inferenza su di esse.

Si supponga quindi che A , M e L varino nel tempo e di disporre per ciascuno di essi di T osservazioni per ogni soggetto: $(A(1), \dots, A(T))$, $(M(1), \dots, M(T))$ e $(L(1), \dots, L(T))$.

Siano C le covariate baseline costanti nel tempo, definite prima della prima osservazione. Inizialmente si assume l'ordinamento temporale $A(t), M(t), L(t)$ che però può essere modificato in maniera abbastanza semplice.

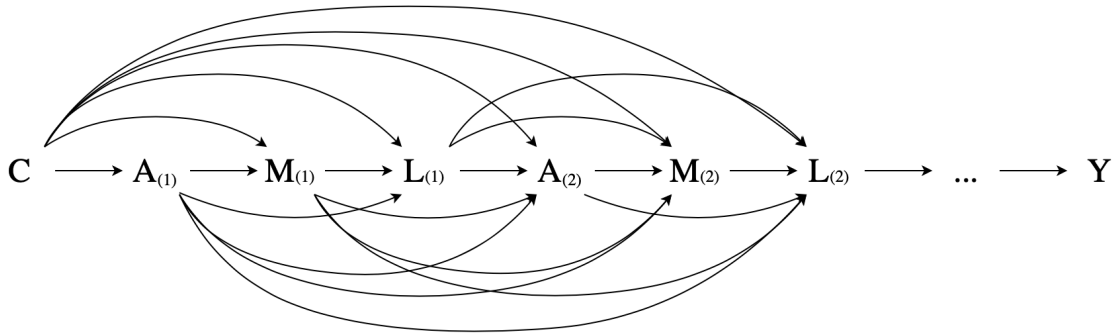


Figura 4.7 – Modello di mediazione con esposizione, mediatore e confondenti che variano nel tempo considerando l'ordinamento temporale $A(t), M(t), L(t)$

Per ogni variabile W si definisce $\overline{W}(t) = (W(1), \dots, W(t))$ la storia della variabile fino al tempo t ; $\overline{W} = \overline{W}(T) = (W(1), \dots, W(T))$ la storia completa della variabile fino a T ; $\underline{W}(t) = (W(t), \dots, W(T))$ la storia della variabile dal tempo t fino al tempo finale T . Per convenzione $W(t)$ indica la situazione in cui $t \leq 0$.

In questo contesto a causa delle relazioni in cui L è coinvolta, NDE e NIE non sono identificati, ma in seguito verrà mostrato come sarà possibile ottenerne degli analoghi basati su interventi randomizzati del mediatore, come nel Paragrafo 4.4.2.

Per gestire questi scenari si fa ricorso ad un approccio basato su un sistema di pesi e a *marginal structural models*. Dovendo valutare più periodi saranno necessari più pesi per l'esposizione e per il mediatore che permetteranno di stimare il CDE . Per calcolarlo verranno confrontati due livelli di esposizione mantenendo costante il mediatore, ma essendo in un contesto longitudinale verranno in realtà confrontate intere traiettorie di esposizione: \overline{a} e \overline{a}^* . Il mediatore invece sarà fissato ad una particolare storia \overline{m} .

Sia $Y_{\overline{am}}$ l'outcome controfattuale quando $\overline{A} = \overline{a}$ e $\overline{M} = \overline{m}$, allora sotto le assunzioni (4.50 e 4.51) è possibile definire il *CDE* a livello di popolazione come $E(Y_{\overline{am}} - Y_{\overline{a^*m}})$.

Le assunzioni devono essere valide in ogni istante temporale e possono quindi essere espresse come:

1. Per ogni istante t l'effetto di $A(t)$ su Y non subisce l'effetto di confondenti dati $\overline{A}(t-1)$, $\overline{M}(t-1)$, $\overline{L}(t-1)$ e C , cioè

$$Y_{\overline{am}} \perp A(t) \mid \overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), C \quad (4.50)$$

2. Per ogni istante t l'effetto di $M(t)$ su Y non subisce l'effetto di confondenti dati $\overline{A}(t-1)$, $\overline{M}(t-1)$, $\overline{L}(t-1)$ e C , cioè

$$Y_{\overline{am}} \perp M(t) \mid \overline{A}(t), \overline{M}(t-1), \overline{L}(t-1), C \quad (4.51)$$

Bastano queste due assunzioni per poter calcolare il *CDE* (Robins, et al., 2000; VanderWeele, 2009), come nel Paragrafo 4.3.2, sarà necessario calcolare i pesi e poi regredire la risposta sul mediatore e sull'esposizione usando i pesi per controllare l'effetto di confondenti che variano nel tempo (Robins, et al., 2000), cioè facendo un MSM.

Per ogni individuo i si definiscono i pesi per ogni istante t come

$$\omega_i^A(t) = \frac{P(A(t) = a_i(t) \mid \overline{a}_i(t-1), \overline{m}_i(t-1))}{P(A(t) = a_i(t) \mid \overline{a}_i(t-1), \overline{m}_i(t-1), \overline{l}_i(t-1), c_i)} \quad (4.52)$$

$$\omega_i^M(t) = \frac{P(M(t) = m_i(t) \mid \overline{a}_i(t), \overline{m}_i(t-1))}{P(M(t) = m_i(t) \mid \overline{a}_i(t), \overline{m}_i(t-1), \overline{l}_i(t-1), c_i)} \quad (4.53)$$

Una volta stimati questi può essere calcolato il peso complessivo per ogni individuo i come

$$\omega_i = \omega_i^A(1) \cdot \omega_i^M(1) \cdots \omega_i^A(t) \cdot \omega_i^M(t) \cdots \omega_i^A(T) \cdot \omega_i^M(T) \quad (4.54)$$

Usati in maniera diretta poi per il modello di regressione finale di Y su \bar{A} e \bar{M} . Al denominatore dei pesi per l'esposizione si trova anche in questo caso la probabilità che un individuo riceva l'esposizione che ha effettivamente ricevuto al tempo t , condizionatamente alla passata storia espositiva, del mediatore, delle covariate che variano nel tempo e di quelle baseline. È l'inclusione delle covariate solamente al denominatore che permettere di correggere per i possibili confondenti che variano nel tempo (Robins, et al., 2000).

Un ulteriore vantaggio di questa procedura è che nella specificazione della relazione tra $E(Y_{\bar{a}\bar{m}})$ e \bar{A} e \bar{M} è concessa molta flessibilità.

Può essere specificato un modello con solo due periodi di esposizione e mantenendo tutte le parti separate come

$$E(Y|a, m) = \gamma_0 + \gamma_1 a(1) + \gamma_2 a(2) + \gamma_3 m(1) + \gamma_4 m(2) \quad (4.55)$$

e ottenere

$$\begin{aligned} E(Y_{\bar{a}\bar{m}}) - E(Y_{\bar{a}^*\bar{m}}) &= \gamma_0 + \gamma_1 a(1) + \gamma_2 a(2) + \gamma_3 m(1) + \gamma_4 m(2) \\ &\quad - (\gamma_0 + \gamma_1 a^*(1) + \gamma_2 a^*(2) + \gamma_3 m(1) + \gamma_4 m(2)) \\ &= \gamma_1 (a(1) - a^*(1)) + \gamma_2 (a(2) - a^*(2)) \end{aligned} \quad (4.56)$$

ma sono anche possibili specificazioni più complesse come

$$E(Y_{\bar{a}\bar{m}}) = \gamma_0 + \gamma_1 cum(\bar{A}) + \gamma_2 cum(\bar{M}) \quad (4.57)$$

oppure con termini quadratici e di interazione.

Per ottenere gli standard error è necessario ricorrere allo stimatore robusto o stimatore *sandwich*.

A seconda della natura di A e M possono essere usate regressioni logistiche, ordinali o multinomiali per calcolare le probabilità da utilizzare nei pesi.

Alcune volte le covariate baseline C possono essere inserite anche nel modello finale per portare a stime più stabili degli effetti ma se si decide di procedere in questo modo è importante che le variabili vengano incluse anche nel condizionamento quando viene calcolata la probabilità al numeratore dei pesi. Non verrà trattato adesso ma è importante sottolineare come queste analisi possano essere ampliate anche al caso in cui la risposta varia nel tempo, tuttavia in questo caso i pesi andrebbero calcolati separatamente per tutti gli istanti temporali e dovrebbero essere applicati poi dei modelli per misure ripetute (VanderWeele, 2009).

Il diverso ordinamento delle variabili può essere preso in considerazione semplicemente modificando il condizionamento nelle probabilità dei pesi, devono essere condizionate solamente al passato (VanderWeele, 2015).

In generale non sono ancora stati sviluppati degli approcci che riescono a soddisfare appieno le situazioni in cui esposizione e mediatore variano nel tempo, ma comunque sono stati fatti notevoli passi avanti rispetto al passato. Il problema adesso riguarda il calcolo di NDE e NIE che in queste situazioni spesso non sono identificati. Per questi motivi non verranno trattati gli effetti come usualmente descritti in letteratura ma si cercherà di ottenerne dei *randomized interventional analogues* che permettono di alleggerire le assunzioni necessarie alla stima.

L'approccio che verrà di seguito descritto si ispira alla *g-formula* di Robins (Robins, 1986) e alla *mediation formula* di Pearl (Pearl, 2001), unendo i due approcci in modo da trovare una soluzione adattabile a questo scenario; quello che ne risulterà prenderà il nome di *mediational g-formula*. L'approccio potrà poi essere implementato semplicemente adattando due *marginal structural models*.

Aggiungendo quindi la terza assunzione

3. Per ogni t , l'effetto di $A(t)$ su $M(t)$ non subisce l'effetto di confondenti dati $\bar{A}(t-1)$, $\bar{M}(t-1)$, $\bar{L}(t-1)$ e C , cioè

$$M_{\bar{a}}(t) \perp A(t) | \bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C \quad (4.58)$$

Sia $M_{\bar{a}}(t)$ il controfattuale di $M(t)$ quando $\bar{A} = \bar{a}$. Si assume la consistenza, cioè che quando $\bar{A} = \bar{a}$ allora $M_{\bar{a}}(t) = M(t)$ e $Y_{\bar{a}}(t) = Y(t)$, quando si osservano $\bar{A} = \bar{a}$ e $\bar{M} = \bar{m}$ allora $Y_{\bar{a}\bar{m}} = Y$.

Sia $\bar{G}_{\bar{a}|c}(t)$ un'estrazione casuale dalla distribuzione di $\bar{M}(t)$ tra coloro con $\bar{A} = \bar{a}$ e condizionatamente ai valori assunti dalle covariate $C = c$. Quindi al tempo t , $\bar{G}_{\bar{a}|c}(t)$ dipende da \bar{a} solo fino al tempo t .

Siano inoltre \bar{a} e \bar{a}^* due differenti storie di esposizione; anche in questo caso vale la decomposizione degli effetti (VanderWeele e Tchetgen Tchetgen, 2016)

$$\begin{aligned} & \underbrace{E\left(Y_{\bar{a}\bar{G}_{\bar{a}|c}(t)} | c\right) - E\left(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|c}(t)} | c\right)}_{\text{Interventional Overall Effect}} \\ &= \underbrace{\left\{E\left(Y_{\bar{a}\bar{G}_{\bar{a}|c}(t)} | c\right) - E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)} | c\right)\right\}}_{\text{Interventional Indirect Effect}} \\ &+ \underbrace{\left\{E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)} | c\right) - E\left(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|c}(t)} | c\right)\right\}}_{\text{Interventional Direct Effect}} \end{aligned} \quad (4.59)$$

Il valore degli effetti ovviamente cambierà al variare delle due traiettorie di esposizione che si stanno comparando.

È importante fare alcune osservazioni sugli analoghi: l'*interventional indirect effect* assumerà un valore diverso da zero solamente se la distribuzione del mediatore viene modificata dall'esposizione e questo cambiamento modifica l'outcome; quando l'assunzione (4.16) non viene violata gli effetti analoghi coincidono con quelli veri; quando i veri effetti non sono identificati solamente in casi molto specifici può capitare che gli analoghi siano diversi da zero ma poi effettivamente non ci sia un effetto indiretto.

Sotto le tre assunzioni (4.50), (4.51) e (4.58) è possibile utilizzare la *mediational g-formula* (VanderWeele e Tchetgen Tchetgen, 2016), definita come

$$\begin{aligned}
& E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)} \mid c\right) \\
&= \int_{\bar{m}} \int_{\bar{l}(T-1)} E(Y \mid \bar{a}, \bar{m}, \bar{l}, c) \prod_{t=1}^{T-1} dP\{\bar{l}(t) \mid \bar{a}(t), \bar{m}(t), \bar{l}(t-1), c\} \\
&\times d\left[\int_{\bar{l}^\dagger(T-1)} \prod_{t=1}^T P\{m(t) \mid \bar{a}^*(t), \bar{m}(t-1), \bar{l}^\dagger(t-1), c\} dP\{\bar{l}^\dagger(t-1) \mid \bar{a}^*(t-1), \bar{m}(t-1), \bar{l}^\dagger(t-2), c\}\right] = Q(\bar{a}, \bar{a}^*)^{13}
\end{aligned} \tag{4.60}$$

e ottenere mediante essa gli analoghi degli effetti

$$E\left(Y_{\bar{a}\bar{G}_{\bar{a}|c}(t)} \mid c\right) - E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)} \mid c\right) = Q(\bar{a}, \bar{a}) - Q(\bar{a}, \bar{a}^*) \tag{4.61}$$

$$E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)} \mid c\right) - E\left(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|c}(t)} \mid c\right) = Q(\bar{a}, \bar{a}^*) - Q(\bar{a}^*, \bar{a}^*) \tag{4.62}$$

In assenza di mediatore la formula si riduce alla *g-formula* di Robins (Robins, 1986) mentre in assenza di variabili confondenti influenzate dall'esposizione, cioè L , la formula si riduce ad un analogo della *mediation formula* di Pearl (Pearl, 2012) che considera anche il tempo.

L'approccio può essere variato per accomodare scenari diversi, magari con un diverso ordinamento temporale delle variabili oppure considerando come mediatore un'estrazione casuale dalla popolazione complessiva esposta senza condizionarsi ai soggetti con un certo set di covariate.

Valori passati di esposizione, covariate e mediatore influiscono sul valore corrente del mediatore quindi nella pratica bisogna utilizzare dei metodi che tengano in considerazione la dipendenza dal tempo per riuscire a stimare la

¹³ Definizione di valore atteso. Sia (Ω, \mathcal{B}, P) uno spazio di probabilità e X una variabile aleatoria a valori reali su tale spazio (ossia una funzione misurabile $X: \Omega \mapsto \mathbb{R}$, dove i numeri si intendono equipaggiati con la loro σ -algebra boreliana). Il valore atteso di X è semplicemente l'integrale di X rispetto alla misura di probabilità P , cioè

$$E(X) := \int_{\Omega} X(\omega) dP(\omega)$$

distribuzione di $\overline{M}_{\overline{a}}(T)$ e poi fare delle estrazioni da questa. Questa dipendenza dal passato viene presa in considerazione dalla *mediational g-formula* ma ci sono anche altri metodi che possono gestirla come ad esempio l'utilizzo di un sistema di pesi e dei *marginal structural models*.

Un primo possibile approccio di stima è quindi quello di utilizzare la formula prima specificata adattando dei modelli parametrici per le diverse componenti e questo prende il nome di *g-computational approach*¹⁴. Solitamente tuttavia non viene scelto come metodo in quanto prevede la specificazione di molti modelli che a volte non risultano essere compatibili tra di loro e con l'ipotesi nulla di assenza di effetto.

Per questo motivo si fa spesso ricorso ad un approccio di stima più parsimonioso che prevede l'adattamento di due modelli marginali strutturali (Robins, et al., 2000; VanderWeele e Tchetgen Tchetgen, 2016), uno per $E(Y_{\overline{am}})$ e uno per $P(\overline{M}_{\overline{a}^*} = \overline{m})$, poi utilizzati per il calcolo degli effetti diretti e indiretti partendo da quello di $E(Y_{\overline{a}\overline{g}_{\overline{a}^*}}) = \int_{\overline{m}} E(Y_{\overline{am}}) dP(\overline{M}_{\overline{a}^*} = \overline{m})$.

Considerando una risposta continua è possibile ipotizzare il seguente MSM lineare

$$E(Y_{\overline{am}}) = \theta_0 + \theta_1 cum(\overline{a}) + \theta_2 cum(\overline{m}) \quad (4.63)$$

dove $cum(\overline{a}) = \sum_{t \leq T} a(t)$ e $cum(\overline{m}) = \sum_{t \leq T} m(t)$.

Il modello assume che esposizione e mediatore hanno un effetto cumulativo nel tempo e, per A e M continui, è come affermare che l'effetto congiunto di \overline{A} e \overline{M} su Y agisce solamente attraverso i rispettivi livelli medi storici e che questi non interagiscono su scala additiva. Il modello può essere reso più complesso e flessibile inserendo ulteriori relazioni e interazioni tra le variabili.

¹⁴ Descritto da Robins e Hernán (2009) in contesti diversi da quelli di mediazione, in situazioni in cui l'esposizione varia nel tempo.

Per il processo di mediazione invece si può ipotizzare il seguente MSM

$$g^{-1}\{E(M_{\bar{a}}(t))\} = \beta_0(t) + \beta_1(t)avg(\bar{a}(t)) \quad (4.64)$$

dove $g^{-1}(\cdot)$ è una funzione legame, i coefficienti potenzialmente possono variare nel tempo e $avg(\bar{a}(t)) = \sum_{j \leq t} a(j)/t$.

Utilizzando i due modelli appena specificati è semplice verificare che

$$\begin{aligned} E(Y_{\bar{a}\bar{G}_{\bar{a}^*}}) &= \int_{\bar{m}} E(Y_{\bar{a}\bar{m}}) dP(\bar{M}_{\bar{a}^*} = \bar{m}) \\ &= \int_{\bar{m}} \{\theta_0 + \theta_1 cum(\bar{a}) + \theta_2 cum(\bar{m})\} dP(\bar{M}_{\bar{a}^*} = \bar{m}) \\ &= \theta_0 + \theta_1 cum(\bar{a}) \\ &\quad + \theta_2 \left(\sum_{t \leq T} g(\beta_0(t) + \beta_1(t)avg(\bar{a}^*(t))) \right) \end{aligned} \quad (4.65)$$

Ed essendo il mediatore continuo può ipotizzarsi una funzione legame identità che permette di ottenere:

$$\bullet \quad \underbrace{E(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)}|c) - E(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|c}(t)}|c)}_{\text{Interventional Direct Effect}} = \theta_1 \{cum(\bar{a}) - cum(\bar{a}^*)\} \quad (4.66)$$

$$\bullet \quad \underbrace{E(Y_{\bar{a}\bar{G}_{\bar{a}|c}(t)}|c) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)}|c)}_{\text{Interventional Indirect Effect}} = \sum_{t \leq T} \theta_2 \beta_1(t) \{avg(\bar{a}(t)) - avg(\bar{a}^*(t))\} \quad (4.67)$$

Le formule si semplificano ulteriormente in caso di esposizione dicotomica e con i coefficienti del modello per il mediatore mantenuti costanti nel tempo (VanderWeele e Tchetgen Tchetgen, 2016).

Per la stima si può utilizzare un sistema di pesi, la cui costruzione varia con le assunzioni che vengono fatte. Se le assunzioni (4.50), (4.51) e (4.58) sono

valide allora la stima dei coefficienti del modello per la risposta può essere ottenuta tramite i minimi quadrati pesati con pesi pari a

$$\begin{aligned} & \prod_{t=1}^{T-1} \hat{P}\{A(t), M(t) | \bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\}^{-1} \\ &= \prod_{t=1}^{T-1} \left(\hat{P}\{M(t) | \bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), C\} \hat{P}\{A(t) | \bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\} \right)^{-1} \end{aligned} \quad (4.68)$$

composto da delle stime di massima verosimiglianza delle probabilità che li caratterizzano, ottenute mediante un modello parametrico.

I coefficienti del modello per il mediatore sono ottenuti mediante *inverse probability weighting* utilizzando i pesi

$$\prod_{s=1}^t \hat{P}\{A(s) | \bar{A}(s-1), \bar{M}(s-1), \bar{L}(s-1), C\}^{-1} \quad (4.69)$$

Lo stesso approccio e ragionamento possono essere seguiti anche nel caso in cui si volessero adattare dei modelli che prevedono un condizionamento a delle covariate baseline.

I modelli possono essere modificati con abbastanza semplicità in modo da includere dei termini con ordine superiore, inserendo dei ritardi oppure considerando gli effetti separatamente per ogni istante temporale. In queste situazioni sarà necessario modificare le formule sopra indicate, ma la procedura con cui si ottengono rimane la medesima (VanderWeele e Tchetgen Tchetgen, 2016). L'approccio può essere ampliato anche ad outcome binari applicando aggiustamenti adeguati¹⁵.

Sembra infine importante osservare che ci sono situazioni in cui le storie di esposizione e mediatore sono poco probabili e quindi i pesi possono diventare molto grandi. Per questo motivo come procedura standard solitamente

¹⁵ Online Supplement a VanderWeele e Tchetgen Tchetgen (2016).

vengono troncate le distribuzioni dei pesi al primo e 99-esimo percentile per ottenere degli stimatori più stabili (Cole e Hernán, 2008).

Per aumentare la stabilità possono essere utilizzati dei pesi stabilizzati (Robins, et al., 2000) utilizzando per il modello per l'outcome al tempo t

$$\frac{\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1)\}\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1)\}}{\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), C\}\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\}} \quad (4.70)$$

e come pesi per il modello per il mediatore al tempo t

$$\frac{\hat{P}\{A(t)|\bar{A}(t-1)\}}{\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\}} \quad (4.71)$$

Questa correzione risulta particolarmente importante quando mediatore ed esposizioni sono variabili continue (Robins, et al., 2000).

4.5.2 Applicazione

Viene proposta l'applicazione dei metodi appena visti in un contesto simile a quello che poi sarà ripreso nei capitoli successivi.

Si fa riferimento all'approccio che utilizza due modelli marginali strutturali per il calcolo degli effetti, come quelli delle Formule 4.63 e 4.64. Esposizione, mediatore e risposta sono considerati variabili continue e per semplicità standardizzate. Le covariate per cui si intende correggere possono essere costanti oppure variabili nel tempo. Affinché l'applicazione di questo metodo abbia valore, sono necessari almeno tre periodi di osservazione: nel primo vengono considerate le covariate baseline, nel secondo o comunque in quelli intermedi sono considerate le variabili che cambiano nel tempo, mentre nell'ultimo periodo viene considerata la risposta finale.

Solitamente i valori dell'outcome precedentemente osservati possono essere considerati come confondenti che variano nel tempo insieme alle altre variabili.

Si considerano cinque periodi di osservazione e l'utilizzo di pesi stabilizzati come quelli riportati nelle Formule 4.70 e 4.71.

L'ordinamento temporale ipotizzato è come quello della *Figura 4.7* ma può essere alterato con abbastanza semplicità semplicemente cambiando i termini nelle regressioni utilizzate per il calcolo dei pesi.

Come prima cosa è necessario calcolare i pesi che poi devono essere applicati ai modelli e si procede in maniera sequenziale calcolando numeratori e denominatori per poi unirli successivamente e ottenere i pesi finali.

Le variabili che vengono considerate sono continue e per questo motivo si utilizzano delle regressioni lineari (*proc reg* su SAS), attraverso queste si ottengono i residui studentizzati poi trasformati nelle componenti dei pesi mediante l'applicazione della distribuzione di densità della Normale standard.

La procedura può essere divisa in passaggi successivi:

1. Sono calcolati i pesi per il numeratore della Formula (4.70).

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1)\}$:

- $M(2) \sim A(1) + A(2) + M(1)$
- $M(3) \sim A(1) + A(2) + A(3) + M(1) + M(2)$
- $M(4) \sim A(1) + A(2) + A(3) + A(4) + M(1) + M(2) + M(3)$

Si costruiranno poi altri tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1)\}$:

- $A(2) \sim A(1) + M(1)$
- $A(3) \sim A(1) + A(2) + M(1) + M(2)$
- $A(4) \sim A(1) + A(2) + A(3) + M(1) + M(2) + M(3)$

2. Sono calcolati i pesi per il denominatore della Formula (4.70).

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), C\}$:

- $M(2) \sim A(1) + A(2) + M(1) + L(1) + C$
- $M(3) \sim A(1) + A(2) + A(3) + M(1) + M(2) + L(1) + L(2) + C$

- $M(4) \sim A(1) + A(2) + A(3) + A(4) + M(1) + M(2) + M(3) + L(1) + L(2) + L(3) + C$

All'interno delle $L(t)$ sono contenuti anche gli outcome nei periodi precedenti all'ultimo.

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\}$:

- $A(2) \sim A(1) + M(1) + L(1) + C$
- $A(3) \sim A(1) + A(2) + M(1) + M(2) + L(1) + L(2) + C$
- $A(4) \sim A(1) + A(2) + A(3) + M(1) + M(2) + M(3) + L(1) + L(2) + L(3) + C$

3. Vengono calcolate le probabilità previste a partire dai residui studentizzati ottenuti negli step (1.) e (2.). Per ogni residuo R viene applicata la distribuzione Normale: $\frac{1}{\sqrt{2\pi}} e^{-R^2/2}$. Sono infine ottenuti i pesi moltiplicando le probabilità previste per il numeratore e dividendole per il prodotto di quelle previste per il denominatore.

4. I pesi ottenuti nello step (3.) sono applicati nella regressione lineare $Y(5) \sim cum(A(t)) + cum(M(t))$.

5. Sono calcolati i pesi per il numeratore della Formula (4.71).

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{A(t)|\bar{A}(t-1)\}$:

- $A(2) \sim A(1)$
- $A(3) \sim A(1) + A(2)$
- $A(4) \sim A(1) + A(2) + A(3)$

6. Sono calcolati i pesi per il denominatore della Formula (4.71).

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), C\}$:

- $A(2) \sim A(1) + M(1) + L(1) + C$
- $A(3) \sim A(1) + A(2) + M(1) + M(2) + L(1) + L(2) + C$
- $A(4) \sim A(1) + A(2) + A(3) + M(1) + M(2) + M(3) + L(1) + L(2) + L(3) + C$

7. Vengono calcolate le probabilità previste a partire dai residui studentizzati ottenuti negli step (5.) e (6.). Per ogni residuo R viene applicata la distribuzione Normale: $\frac{1}{\sqrt{2\pi}} e^{-R^2/2}$. Sono infine ottenuti i pesi moltiplicando le probabilità previste per il numeratore e dividendole per il prodotto di quelle previste per il denominatore.
8. I pesi ottenuti nello step (7.) sono applicati nella regressione $M(5) \sim avg(A(t))$.
9. Una volta stimati i coefficienti dei modelli è possibile ottenere le stime degli effetti diretti e indiretti semplicemente applicando i risultati ottenuti dal punto di vista teorico (Formule 4.66 e 4.67).

Nel caso in cui le variabili fossero dicotomiche il procedimento risulterebbe ancora più semplice in quanto il calcolo delle probabilità previste sarebbe immediato dopo l'applicazione di una regressione logistica. Per il calcolo degli effetti finali bisognerebbe solamente ottenerne dal punto di vista teorico la forma e poi utilizzare le stime dei coefficienti per stimarli.

I coefficienti delle due regressioni (step (4.) e (8.)) possono essere ottenuti mediante l'applicazione di modelli di regressione lineari, modelli di regressione logistica oppure anche di modelli ad effetti fissi oppure casuali (*proc mixed* di SAS). In questo modo è possibile ottenere una grande libertà nella modellazione del fenomeno, rispettandone la sua natura e la natura delle variabili coinvolte, in modo da ottenere le stime dei coefficienti più accurate possibile.

5. ANALISI TRASVERSALI

La relazione a cui si è interessati e che si cerca di analizzare riguarda il legame tra lo stato socio-economico degli individui e il loro benessere, prestando particolare attenzione al potenziale effetto di mediazione determinato dalla loro condizione di salute fisica.

Dopo aver selezionato solamente i soggetti che partecipano alle wave 1, 2, 4, 5 e 6, è stata svolta la medesima analisi parallelamente su tutte le cinque wave.

In questo modo è possibile avere una visione trasversale del fenomeno, congiuntamente ad una longitudinale dello stesso, valutando se nelle stime degli effetti è possibile osservare dei trend oppure delle modifiche potenzialmente dovute alla crisi economica o, più semplicemente, all'invecchiamento del campione in esame.

Come è stato mostrato precedentemente (Capitolo 4) si può ricorrere a specifici modelli di mediazione che permettono di comprendere il modo e la misura in cui le tre variabili possono relazionarsi. È stata posta grande attenzione alla scelta dei possibili confondenti e di tutte le variabili che potevano influire in qualche modo sulla relazione principale.

Procedendo in tal modo è possibile far sì che le assunzioni sottostanti il modello siano rispettate e quindi che si possa conferire un'interpretazione causale alle stime.

In ogni wave si è deciso di procedere in due fasi successive: prima è stato costruito un modello di mediazione senza covariate e successivamente si è ripetuta la stima del modello con l'inserimento delle covariate. È stato così possibile avere una visione iniziale globale del fenomeno e valutare in un secondo momento in che misura l'inclusione dei potenziali confondenti modificasse gli effetti, comprendendo così la loro reale importanza.

Le variabili utilizzate nella seconda fase sono state scelte sulla base delle analisi descrittive svolte nel Capitolo 3, congiuntamente alla valutazione della letteratura svolta nel Capitolo 1.

Le due fasi, distinte dall'assenza e dalla presenza dei confondenti nei modelli, sono caratterizzate dall'applicazione dei metodi su due differenti gruppi di soggetti: il primo che per ogni wave comprende tutti i rispondenti e il secondo nel quale vengono selezionati solamente coloro che hanno partecipato a tutte le wave di rilevazione.

Questo modo di procedere è giustificato dal fatto che l'applicazione dei modelli di mediazione tenendo in considerazione la temporalità del fenomeno (Capitolo 6) prevede la selezione di soggetti che hanno partecipato a tutte le wave di rilevazione. Così facendo, tuttavia, si è consapevoli di commettere un errore da selezione (*selection bias*), poiché considerando solamente i soggetti che partecipano a tutte le rilevazioni vengono eliminati coloro che sono deceduti e coloro che potenzialmente erano in una condizione di salute compromessa che ha loro impedito la partecipazione. Inoltre, vengono eliminati anche i soggetti che SHARE aggiunge in ogni periodo di rilevazione proprio per far fronte all'*attrition* e quindi alla perdita di soggetti al follow-up per qualsiasi tipo di causa.

Si tratta di un errore che, specialmente considerando come mediatore lo stato di salute fisica degli individui, si ritiene possa avere un'influenza abbastanza forte sulla stima dell'effetto indiretto, probabilmente diminuendone l'impatto a causa della selezione dei soggetti più resistenti.

Con lo scopo di cercare di valutare l'errore che si sta commettendo in modo da avere un'idea più chiara anche nel momento in cui vengono svolte analisi longitudinali, si portano avanti in contemporanea per ogni wave le medesime analisi applicate sui due dataset differenti.

Come è stato mostrato nel Capitolo 4 dal punto di vista teorico, sono presenti diversi metodi che permettono di gestire il confondimento anche in contesti di mediazione.

In questa prima parte di analisi si è deciso di utilizzare l'approccio basato sulla regressione proposto da VanderWeele (2015). Questo metodo confida nella corretta specificazione degli effetti delle covariate nei modelli per la risposta e per il mediatore così da correggere le stime degli effetti di interesse. Le analisi sono svolte mediante l'applicazione della *proc causalmed* di SAS e, nonostante la procedura utilizzata permetta l'implementazione di diversi modelli lineari generalizzati, i coefficienti stimati sono il risultato di semplici regressioni lineari, poiché le tre variabili principali dell'analisi sono state mantenute continue dopo la loro creazione mediante l'applicazione di analisi fattoriali.

A tal proposito sembra opportuno ricordare che la ricchezza degli individui o della coppia, nel caso in cui il rispondente avesse un partner, è rappresentata da un fattore che permette di riassumere il valore della residenza principale, il valore di altri possedimenti, il denaro contenuto nei conti bancari e in bond, azioni e fondi comuni; il benessere è un fattore creato tramite l'unione degli indici CASP, una misura del benessere, e EURO – D, una scala che identifica il grado di depressione degli individui; infine la salute fisica è identificata da un terzo fattore che permette di unire le difficoltà che gli individui hanno nello svolgere le attività di vita quotidiana e i problemi di mobilità che presentano. Per maggiori dettagli sulla creazione dei fattori si veda il Capitolo 3.

Nel dataset completo viene applicata l'analisi fattoriale e i fattori che ne risultano presentano per costruzione media nulla e varianza unitaria.

Nel dataset riferito al panel i fattori sono standardizzati dopo la selezione in modo da semplificare la lettura delle stime degli effetti e dei coefficienti dei modelli.

Nelle *Tabelle 5.1 e 5.2* è possibile trovare delle misure che riprendono quanto detto nel Capitolo 3 e riassumono le caratteristiche delle variabili da un punto di vista puramente descrittivo.

A differenza di quanto mostrato nelle analisi descrittive del Capitolo 3, la variabile relativa alla condizione di salute fisica degli individui è stata invertita in modo che una sua crescita indichi un miglioramento nello stato

di salute e questo può rendere più facile l'interpretazione dei termini di interazione e dei coefficienti dei modelli.

La standardizzazione è stata svolta dopo la selezione dei soggetti e per tale motivo sono presenti valori differenti di massimi e minimi tra le variabili confrontando i risultati dei due dataset nella stessa wave.

		Minimo	Massimo	Mediana	Range
	SES	-0.753220	36.795848	-0.26629	37.54907
Wave 1	Salute fisica	-6.4323048	0.5344977	0.534498	6.96680
	Benessere	-4.494962	1.594267	0.197322	6.08923
	SES	-0.788038	15.210859	-0.29440	15.99890
Wave 2	Salute fisica	-6.0478248	0.5142475	0.514248	6.56207
	Benessere	-4.852211	1.541319	0.199934	6.39353
	SES	-0.5741623	18.2498440	-0.30678	18.82401
Wave 4	Salute fisica	-5.710617	0.536435	0.536435	6.24705
	Benessere	-4.811602	1.533589	0.182702	6.34519
	SES	-0.647974	20.683584	-0.29920	21.33156
Wave 5	Salute fisica	-5.5140802	0.5179773	0.517977	6.03206
	Benessere	-4.563403	1.474774	0.219613	6.03818
	SES	-0.696236	20.103308	-0.28417	20.79954
Wave 6	Salute fisica	-5.5937843	0.5228040	0.522804	6.11659
	Benessere	-4.855314	1.498454	0.190809	6.35377

Tabella 5.1 – Misure riassuntive delle variabili relative a SES, salute fisica e benessere in tutte le wave nei dataset completi

		Minimo	Massimo	Mediana	Range
	SES	-0.835103	12.106551	-0.28820	12.94165
Wave 1	Salute fisica	-9.389956	0.525772	0.525772	9.91573
	Benessere	-4.319638	1.576328	0.187828	5.89597
	SES	-0.838069	14.705118	-0.29970	15.54319
Wave 2	Salute fisica	-8.639541	0.513234	0.513234	9.15277
	Benessere	-5.269911	1.565977	0.209977	6.83589
	SES	-0.693864	15.514116	-0.32969	16.20798
Wave 4	Salute fisica	-6.764306	0.565609	0.280307	7.32991
	Benessere	-4.277971	1.558164	0.181155	5.83613
	SES	-0.714920	15.374901	-0.29232	16.08982
Wave 5	Salute fisica	-5.857609	0.571886	0.316654	6.42950
	Benessere	-4.375726	1.473510	0.188049	5.84924
	SES	-0.771518	13.664666	-0.27496	14.43618
Wave 6	Salute fisica	-5.043493	0.585125	0.362433	5.62862
	Benessere	-4.221387	1.512608	0.184215	5.73399

Tabella 5.2 – Misure riassuntive delle variabili relative a SES, salute fisica e benessere in tutte le wave nei dataset riferiti al panel

5.1 Relazioni tra variabili e *selection bias*

Quando si analizza la letteratura che studia le relazioni tra stato socio-economico, stato di salute e benessere, è possibile osservare come non sempre ad un benessere economico sia associato un benessere psicologico (Kaplan, et al., 2008; Blustein, 2008). Tuttavia, nelle analisi descrittive svolte, sono state individuate delle associazioni positive tra queste variabili e ciò porta ad indicare come ci si possa aspettare che in questo campione lo stato socio-economico di un individuo, identificato mediante la sua ricchezza, abbia un effetto positivo sul suo benessere.

Per quanto riguarda il mediatore ci si aspetta una relazione positiva con il benessere, in quanto al miglioramento delle condizioni fisiche un individuo tende ad avere un incremento del proprio benessere.

La relazione che invece risulta essere più problematica da definire è quella tra lo stato socio-economico e la salute fisica, poiché generalmente vengono presentate le società moderne come società del benessere in cui la salute fisica è praticamente garantita a tutti ma, nonostante questo, confrontando paesi differenti e sistemi sanitari differenti, emergono delle realtà nelle quali la buona salute è maggiormente garantita a chi è in migliori condizioni economiche (Whitehead, et al., 2001; Newell, 2000; Barbeau, et al., 2004; Adler, et al., 2016). Questo porta alla consapevolezza che esistono ancora molte società, anche tra quelle maggiormente sviluppate, all'interno delle quali le disparità nell'accesso alle cure e ai trattamenti sono ancora molto presenti.

Il campione di riferimento in questa analisi svolge perciò un ruolo fondamentale, poiché si tratta di soggetti ultracinquantenni che si trovano a dover affrontare un processo di invecchiamento inevitabile che, indipendentemente dalle risorse economiche possedute da essi, li porta ad un peggioramento della condizione di salute fisica.

In questo scenario appare ancora più importante differenziare la salute fisica da quella psicologica, in quanto soggetti con uno stato socio-economico più alto potrebbero essere in grado non solo di ricorrere a cure migliori, ma anche a sistemi di supporto psicologico che possono incrementare il loro benessere.

Oltre all'età del campione di riferimento è importante considerare il fatto che sono analizzate le realtà di paesi europei che, nonostante le inevitabili differenze legate a sistemi sanitari e politici, presentano degli standard di accesso alle cure e di trattamenti molto avanzati e con possibilità di accesso molto alto.

Per questi motivi, seppure ci si aspetta che emergano delle relazioni significative tra le variabili, gli effetti potrebbero essere meno forti rispetto a quelli che si potrebbero ottenere svolgendo questa tipologia di analisi su una

realità più differenziata come può essere quella dei paesi in via di sviluppo o di quelli che, nonostante siano maggiormente sviluppati, presentano sistemi sanitari privati che limitano notevolmente l'accesso della popolazione alle cure.

Per un'interpretazione accurata dei risultati dei modelli riportati di seguito, è fondamentale affrontare il problema dell'errore da selezione in maniera più approfondita. Come è stato detto prima, oltre alle analisi trasversali su tutti gli individui, sono state ripetute le stesse analisi solamente sul panel di soggetti che partecipano a tutte le wave. In tale modo è possibile considerare il fenomeno anche con un'ottica temporale e rendere possibile l'applicazione di modelli di mediazione con esposizione e mediatore che variano nel tempo come quelli presentati alla fine del Capitolo 4 e applicati in seguito nel Capitolo 6.

Confrontando le medie delle variabili relative alla ricchezza, alla condizione di salute e al benessere nei due dataset, è emerso che in tutte le wave nel dataset riferito al panel sono presenti soggetti con uno stato socio-economico che in media risulta essere superiore rispetto a quando si considera la popolazione intera. Inoltre, mentre nelle prime tre wave (1, 2 e 4) i soggetti che partecipano a tutte le rilevazioni presentano una condizione di salute e di benessere migliore rispetto al campione completo, nella quinta presentano livelli di benessere migliore, ma si osserva un peggioramento della condizione di salute e nella sesta oltre ad un continuo peggioramento dello stato di salute, si evidenzia anche che il benessere risulta essere in media inferiore. Si tratta di un risultato legato all'influenza di più effetti che si sommano: da un lato l'effetto dell'invecchiamento dei soggetti nel panel che tuttavia presentavano inizialmente condizioni migliori, dall'altro l'introduzione del *refreshment sample* nel campione completo che riduce l'effetto complessivo del processo di invecchiamento.

Queste considerazioni sono molto importanti anche nell'interpretazione dei risultati finali, poiché selezionando i soggetti viene valutato il loro processo

di invecchiamento andando ad osservare un campione che nella prima wave presenta un'età media di 61 anni fino ad arrivare a 73 nella sesta. Differentemente da questo, considerando il campione intero, ad ogni rilevazione, SHARE per far fronte alla perdita di soggetti al follow-up permette l'ingresso di nuovi soggetti che, purché rispettando i criteri di accesso per la partecipazione, non necessariamente hanno le stesse caratteristiche di coloro che sono stati persi. Questo può essere rilevato con semplicità analizzando anche solamente l'età media degli individui nelle diverse wave e confrontandola con quella del campione selezionato (*Tabella 5.3*).

		Wave 1	Wave 2	Wave 4	Wave 5	Wave 6
Età media (anni)	Campione intero	64	65	66	67	68
	Campione panel	61	64	69	71	73

Tabella 5.3 – Età media dei soggetti nelle diverse wave nel campione intero e in quello riferito al panel

Quando vengono selezionati gli individui, quindi, ad eccezione delle prime due wave, si considerano delle persone che rappresentano una porzione della popolazione più anziana, pertanto inevitabilmente in condizione di salute peggiore e di conseguenza anche di benessere, secondo quanto emerso dalle analisi di correlazione. Da queste analisi, infatti, è possibile osservare una crescita della correlazione tra benessere e salute con il procedere delle wave (Appendice C).

A causa dell'introduzione dei nuovi soggetti ad ogni rilevazione per attenuare l'effetto dell'*attrition*, nel campione complessivo non è possibile valutare allo stesso modo l'evoluzione del campione iniziale e nella stima dei modelli e degli effetti non è possibile comprendere in che misura l'effetto è relativo al campione longitudinale e in che misura a quello appena introdotto. Non è

inoltre possibile osservare un effetto dell'errore da selezione che aumenta con il procedere delle wave, come invece ci si potrebbe aspettare.

5.2 L'effetto di mediazione senza covariate

Nella prima fase viene valutato il modello di mediazione che graficamente può essere rappresentato come in *Figura 5.1*.

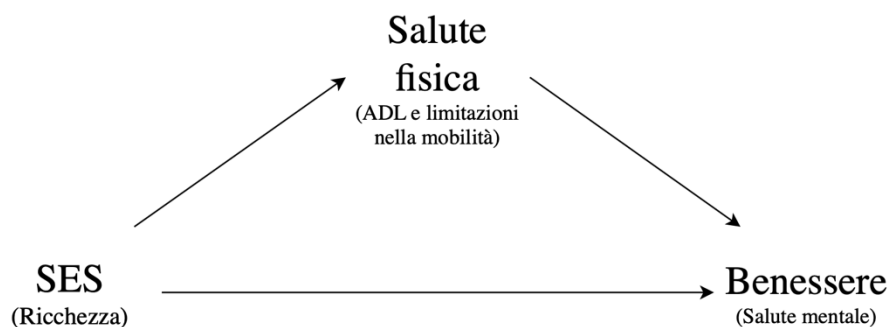


Figura 5.1 – Rappresentazione grafica delle relazioni ipotizzate tra esposizione, mediatore e risposta

Procedendo con il metodo basato sulla regressione, sono specificati due modelli che permettono di prendere in considerazione anche la possibile presenza di interazione tra lo stato socio-economico e le condizioni di salute fisica degli individui:

$$E(Y|A = a, M = m) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am$$

$$E(M|A = a) = \theta_0^\dagger + \theta_1^\dagger a$$

Con A viene indicato lo stato socio-economico dell'individuo, con M la sua condizione di salute e con Y il suo benessere.

È importante ricordare per la lettura dei risultati che a valori alti di A , M e Y corrispondono una maggiore ricchezza, una condizione di salute migliore e un livello alto di benessere e qualità della vita.

Come mostrato nell'esempio del Paragrafo 4.3.3, è possibile interpretare i coefficienti dei modelli come quelli di una regressione lineare semplice.

Le analisi sono state svolte per ogni wave inizialmente sul campione complessivo e successivamente sui 6707 soggetti che hanno partecipato a tutte le interviste.

Non inserendo variabili per il controllo del confondimento le stime saranno diverse da quelle reali e non essendo rispettate le ipotesi necessarie (Formule 4.13 – 4.16), non è possibile conferire agli effetti un'interpretazione causale. Appare tuttavia importante mostrare anche questa situazione per comprendere il fenomeno nella sua complessità e per chiarire l'importanza delle assunzioni che sono state specificate precedentemente solo a livello teorico, ma che permettono di arrivare a delle stime corrette degli effetti.

Si procede quindi mostrando i risultati dei modelli applicati ad ogni wave separatamente e valutando similitudini e differenze iniziali.

5.2.1 Risultati

Per ogni wave le tabelle contenenti le stime dei coefficienti dei modelli con i rispettivi intervalli di confidenza e livelli di significatività sono riportati nell'Appendice D per i campioni completi e nell'Appendice E per i campioni riferiti al panel.

In tutti i modelli, in entrambi i dataset, le stime dei coefficienti sono risultate significative, ad eccezione del termine di interazione tra SES e salute fisica nei modelli per la risposta che lo sono solamente nelle wave 1, 4 e 5, nel caso dei dataset integrali, e nelle wave 4 e 6, in quelli con la selezione dei soggetti. Nonostante questo si è deciso di mantenere ugualmente l'interazione all'interno dei modelli, poiché si ritiene che possa essere fondamentale per aiutare a capire il modello e possa conferire maggiore potenza nell'identificazione degli effetti.

L'effetto dello stato socio-economico sul benessere psicologico degli individui è identificato in tutte le wave da un coefficiente significativo con segno positivo; questo indica che al crescere della ricchezza degli individui il loro

well-being migliora, probabilmente poiché, avendo maggiori disponibilità economiche, possono fare ricorso a sistemi di supporto migliori per gestire problemi di questo tipo, migliorando quindi quegli aspetti della loro vita.

Gli effetti del SES sul benessere ottenuti con i dati del campione completo sono molto simili nelle prime tre wave considerate, mentre nella quinta e nella sesta wave risultano essere leggermente più alti, passando da un valore intorno allo 0.11 ad uno di circa 0.13. Questo indica che per ogni incremento unitario dell'indice relativo alla ricchezza, è possibile individuare una crescita del benessere dell'1-2%¹⁶.

Considerando il campione riferito al panel la situazione appare più complessa, in quanto il coefficiente assume un valore iniziale di 0.14 a cui seguono per due volte una decrescita e una crescita che permettono di arrivare all'ultima wave con un valore del coefficiente di 0.15.

L'analisi del fenomeno risulta complessa poiché, nel campione riferito al panel sono stati selezionati dei soggetti con un maggior livello di benessere ma che, col susseguirsi delle rilevazioni, arrivano alla sesta wave in una condizione mediamente inferiore rispetto al campione completo.

Mediante il confronto degli intervalli di confidenza è possibile valutare se effettivamente i coefficienti possono considerarsi significativamente uno diverso dall'altro.

Il coefficiente corrispondente alla relazione tra SES e benessere nella sesta wave del campione riferito al panel assume valore più alto rispetto a quelli di tutte le altre wave, indicando il forte legame tra le due variabili.

Non è possibile confrontare gli intervalli di confidenza dei coefficienti per valutare differenze tra wave nel campione completo e tra il campione completo e quello riferito al panel poiché fanno riferimento a popolazioni diverse.

¹⁶ Valore ottenuto rapportando il valore relativo all'incremento del benessere dovuto ad un aumento unitario della ricchezza (0.11 – 0.13) con il range del fattore relativo al benessere ($\approx 6 - 6.3$).

Tuttavia, possono essere confrontati i semplici valori dei coefficienti e la differenza maggiore che può essere individuata è tra il coefficiente della sesta wave nel dataset riferito al panel e quello della quarta wave nel campione completo, che riportano rispettivamente come stime il valore più grande e quello più piccolo.

La decrescita del coefficiente nelle prime tre wave del campione completo a cui segue una crescita nelle ultime due potrebbe essere giustificata dal fatto che probabilmente una porzione delle persone nel campione ha risentito della crisi economica e quindi questo porta a una differenza maggiore tra i soggetti nel campione che si riflette in stime diverse. Purtroppo essendo popolazioni diverse non è possibile svolgere un confronto statisticamente valido.

L'ultimo coefficiente stimato nei modelli è quello relativo alla salute fisica degli individui che indica il forte legame con il loro benessere psicologico.

I coefficienti portando tutti segno positivo, indicano che al diminuire delle difficoltà che si hanno nelle regolari attività della vita e degli impedimenti derivanti dal processo di invecchiamento, il benessere psicologico aumenta.

Specialmente trattandosi di persone prevalentemente ultracinquantenni il legame tra i due potrebbe essere più forte rispetto a quello che si verificherebbe in una popolazione più giovane all'interno della quale i disagi a livello psicologico sono molto spesso dovuti ad altri aspetti della vita.

Nel campione panel il coefficiente cresce sempre di più proseguendo con le wave. Come detto prima, questo può essere giustificato dal fatto che i soggetti invecchiano e presentano necessariamente un peggioramento della loro condizione che porta il legame tra le due variabili ad essere sempre più forte e importante. Nel campione completo, al contrario, non è possibile identificare nessun trend particolare.

Nel modello per il mediatore in tutte le situazioni il coefficiente ha segno positivo. Questo indica che al crescere della ricchezza degli individui la loro salute migliora, in quanto diminuiscono i loro problemi nella mobilità e nello svolgimento delle attività di vita quotidiana. Confrontando i risultati nei due campioni l'unica differenza che sembra opportuno evidenziare è il fatto che i

coefficienti riportino valori più alti nel campione panel, ma confrontando gli intervalli di confidenza non è possibile identificare una reale differenza.

Ovviamente analizzare questa relazione senza considerare l'età e il paese di provenienza può apparire fuorviante, tuttavia è importante mostrare come questo confermi le ipotesi iniziali formulate durante lo svolgimento delle analisi descrittive.

Quello che risulta particolarmente importante, trattandosi di un'analisi di mediazione, rimane però la stima degli effetti e la valutazione della possibile presenza di un'effettiva mediazione dell'effetto.

Di seguito nelle *Tabelle 5.4 e 5.5* sono riportate le stime degli effetti stimati e le percentuali degli effetti dovuti alla mediazione nelle diverse wave rispettivamente per il campione completo e per quello riferito al panel.

Come prima cosa è possibile osservare come tutte le stime degli effetti siano risultate significativamente diverse da zero, ad eccezione della percentuale di effetto dovuta all'interazione nella seconda e nella sesta wave del campione completo e nella seconda e quinta wave del campione panel.

Valutando poi la progressione del fenomeno e considerando quindi l'effetto totale che il SES ha sul benessere degli individui nelle diverse wave, è possibile notare per gli effetti stimati dal campione completo una decrescita iniziale della stima fino alla wave 4, a cui segue un aumento nelle ultime due. Nonostante questo, confrontando gli intervalli di confidenza, non c'è una differenza significativa. Lo stesso vale per le stime degli effetti diretti e indiretti che risultano essere simili in tutte le wave.

Nel campione riferito al panel, invece, l'effetto totale subisce una decrescita tra la prima e la seconda wave, per poi rialzarsi e continuare a crescere nelle wave successive. In generale, confrontando gli effetti stimati nei due campioni, i valori ottenuti con il dataset riferito al panel hanno valori sempre superiori o al massimo simili a quelli stimati nel dataset completo e questo riflette le differenze osservate precedentemente nei coefficienti dei modelli.

In generale, comparando le wave, emergono delle differenze maggiori nelle stime degli effetti quando queste sono ottenute sul dataset riferito al panel,

probabilmente poiché si tratta di un gruppo di soggetti che evolve nel tempo, mentre mantenendo il campione completo si ha l'inserimento di nuovi soggetti (*refreshment sample*), che rende i campioni più simili.

Si può osservare che la percentuale dell'effetto totale dovuto al processo di mediazione varia tra il 20% e il 30%, indicando come sia importante l'inclusione della condizione fisica degli individui per la comprensione complessiva delle relazioni tra stato socio-economico e benessere e del fatto che quello che apparentemente sembra un'associazione tra due variabili può nascondere una realtà più complessa.

Inoltre, si pensa che questo valore potrebbe essere ancora più alto se si fossero considerati anche i soggetti deceduti o quelli che hanno smesso di partecipare alla rilevazione, questo poiché sono stati considerati anche soggetti in condizioni peggiori che quindi avrebbero potenzialmente permesso di incrementare l'effetto e l'importanza della salute fisica come mediatore.

Prendendo come riferimento i risultati della sesta wave nel campione riferito al panel, è possibile dare la seguente interpretazione alle stime:

- un valore del $CDE(\bar{m})$, ovvero dell'effetto del passaggio dell'esposizione da a a a^* quando il mediatore M rimane fisso al valore m uniformemente nella popolazione, pari a 0.15 indica l'aumento medio del benessere psicologico che si potrebbe osservare nel caso in cui la salute fisica fosse fissata al suo livello medio e lo stato socio-economico passasse da essere 0.5 sotto alla media a 0.5 sopra alla media. Tale valore corrisponde ad un aumento in media del benessere del 2.6%;
- il NDE, ovvero l'effetto del passaggio dell'esposizione da a a a^* quando per ogni individuo il mediatore M rimane fissato al livello che avrebbe assunto in assenza di esposizione, assume anch'esso valore 0.15 e indica il cambiamento in media del benessere psicologico quando lo stato socio-economico passasse da essere $a = \bar{a} + 0.5$ a $a^* = \bar{a} - 0.5$ e la salute fisica è fissata al livello che avrebbe assunto quando lo stato socio-economico assume il valore a ;

- il NIE, ovvero l'effetto del cambiamento del mediatore dal valore che assumerebbe quando l'esposizione è a^* a quello che assumerebbe quando è a e l'esposizione è fissata al valore a , indica che il valore assunto dal benessere cambierebbe in media di 0.06 quando lo stato socio-economico assume valore $a = \bar{a} + 0.5$ ma la salute fisica passa dal livello che avrebbe assunto quando il SES è fissato ad a , al livello che avrebbe assunto quando il SES è fissato al livello a^* .

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value	
Wave 1	Effetto Totale	0.1719	0.00878	0.1547	0.1891	19.57	<.0001
	Controlled Direct Effect (CDE)	0.1192	0.00809	0.1034	0.1351	14.74	<.0001
	Natural Direct Effect (NDE)	0.1177	0.00828	0.1015	0.1339	14.21	<.0001
	Natural Indirect Effect (NIE)	0.0542	0.00408	0.04621	0.06218	13.30	<.0001
	Percentuale di effetto dovuta alla mediazione	31.5261	2.3520	26.9162	36.1359	13.40	<.0001
	Percentuale di effetto dovuta all'interazione	0.8700	0.3989	0.08811	1.6520	2.18	0.0292
Wave 2	Effetto Totale	0.1783	0.00701	0.1645	0.1920	25.43	<.0001
	Controlled Direct Effect (CDE)	0.1155	0.00645	0.1029	0.1281	17.90	<.0001
	Natural Direct Effect (NDE)	0.1153	0.00670	0.1022	0.1285	17.22	<.0001
	Natural Indirect Effect (NIE)	0.0629	0.00343	0.05620	0.06966	18.33	<.0001
	Percentuale di effetto dovuta alla mediazione	35.3003	1.9616	31.4555	39.1450	18.00	<.0001
	Percentuale di effetto dovuta all'interazione	0.0917	0.3504	-0.5950	0.7785	0.26	0.7935
Wave 4	Effetto Totale	0.1608	0.00577	0.1495	0.1721	27.85	<.0001
	Controlled Direct Effect (CDE)	0.1109	0.00531	0.1005	0.1213	20.88	<.0001
	Natural Direct Effect (NDE)	0.1124	0.00549	0.1016	0.1231	20.48	<.0001
	Natural Indirect Effect (NIE)	0.0484	0.00257	0.04342	0.05347	18.88	<.0001
	Percentuale di effetto dovuta alla mediazione	30.1262	1.6034	26.9836	33.2689	18.79	<.0001
	Percentuale di effetto dovuta all'interazione	-0.9100	0.2681	-1.4356	-0.3845	-3.39	0.0007

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value
Wave 5	Effetto Totale	0.1812	0.00501	0.1714	0.1910	36.19	<.0001
	Controlled Direct Effect (CDE)	0.1288	0.00455	0.1199	0.1378	28.33	<.0001
	Natural Direct Effect (NDE)	0.1308	0.00468	0.1217	0.1400	27.96	<.0001
	Natural Indirect Effect (NIE)	0.0504	0.00229	0.04588	0.05484	22.03	<.0001
	Percentuale di effetto dovuta alla mediazione	27.7939	1.2124	25.4176	30.1702	22.92	<.0001
	Percentuale di effetto dovuta all'interazione	-1.0974	0.2149	-1.5187	-0.6762	-5.11	<.0001
Wave 6	Effetto Totale	0.1843	0.00524	0.1740	0.1946	35.19	<.0001
	Controlled Direct Effect (CDE)	0.1331	0.00484	0.1236	0.1426	27.48	<.0001
	Natural Direct Effect (NDE)	0.1334	0.00497	0.1237	0.1431	26.84	<.0001
	Natural Indirect Effect (NIE)	0.0509	0.00234	0.04631	0.05548	21.76	<.0001
	Percentuale di effetto dovuta alla mediazione	27.6155	1.2489	25.1676	30.0634	22.11	<.0001
	Percentuale di effetto dovuta all'interazione	-0.1847	0.2048	-0.5862	0.2167	-0.90	0.3671

Tabella 5.4 – Stime degli effetti del modello con solo esposizione e mediatore per la risposta nelle diverse wave nel campione completo

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value
Wave 1	Effetto Totale	0.1940	0.0144	0.1658	0.2222	13.49	<.0001
	Controlled Direct Effect (CDE)	0.1452	0.0134	0.1190	0.1715	10.85	<.0001
	Natural Direct Effect (NDE)	0.1439	0.0137	0.1171	0.1707	10.54	<.0001
	Natural Indirect Effect (NIE)	0.0501	0.00636	0.03766	0.06257	7.88	<.0001
	Percentuale di effetto dovuta alla mediazione	25.8292	3.1990	19.5593	32.0991	8.07	<.0001
	Percentuale di effetto dovuta all'interazione	0.6823	0.5242	-0.3451	1.7096	1.30	0.1930
Wave 2	Effetto Totale	0.1685	0.0125	0.1441	0.1930	13.50	<.0001
	Controlled Direct Effect (CDE)	0.1143	0.0117	0.09140	0.1371	9.79	<.0001
	Natural Direct Effect (NDE)	0.1128	0.0120	0.08926	0.1364	9.39	<.0001

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value	
	Natural Indirect Effect (NIE)	0.0557	0.00573	0.04450	0.06695	9.73	<.0001
	Percentuale di effetto dovuta alla mediazione	33.0637	3.5223	26.1601	39.9673	9.39	<.0001
	Percentuale di effetto dovuta all'interazione	0.8644	0.6495	-0.4086	2.1375	1.33	0.1832
Wave 4	Effetto Totale	0.1875	0.0126	0.1627	0.2123	14.83	<.0001
	Controlled Direct Effect (CDE)	0.1366	0.0117	0.1136	0.1597	11.63	<.0001
	Natural Direct Effect (NDE)	0.1410	0.0122	0.1171	0.1648	11.60	<.0001
	Natural Indirect Effect (NIE)	0.0465	0.00513	0.03646	0.05658	9.06	<.0001
	Percentuale di effetto dovuta alla mediazione	24.8121	2.7542	19.4140	30.2101	9.01	<.0001
	Percentuale di effetto dovuta all'interazione	-2.3075	0.7095	-3.6981	-0.9169	-3.25	0.0011
Wave 5	Effetto Totale	0.1901	0.0127	0.1653	0.2150	15.02	<.0001
	Controlled Direct Effect (CDE)	0.1248	0.0114	0.1024	0.1471	10.95	<.0001
	Natural Direct Effect (NDE)	0.1251	0.0117	0.1022	0.1480	10.72	<.0001
	Natural Indirect Effect (NIE)	0.0650	0.00630	0.05267	0.07738	10.32	<.0001
	Percentuale di effetto dovuta alla mediazione	34.1991	3.1844	27.9578	40.4405	10.74	<.0001
	Percentuale di effetto dovuta all'interazione	-0.1926	0.5138	-1.1996	0.8143	-0.37	0.7077
Wave 6	Effetto Totale	0.2182	0.0130	0.1926	0.2437	16.72	<.0001
	Controlled Direct Effect (CDE)	0.1502	0.0124	0.1260	0.1745	12.14	<.0001
	Natural Direct Effect (NDE)	0.1532	0.0131	0.1276	0.1788	11.73	<.0001
	Natural Indirect Effect (NIE)	0.0650	0.00587	0.05347	0.07648	11.07	<.0001
	Percentuale di effetto dovuta alla mediazione	29.7828	2.8777	24.1426	35.4230	10.35	<.0001
	Percentuale di effetto dovuta all'interazione	-1.3629	0.6227	-2.5834	-0.1424	-2.19	0.0286

Tabella 5.5 – Stime degli effetti del modello con solo esposizione e mediatore per la risposta nelle diverse wave nel campione riferito al panel

Gli effetti riportati nelle *Tabelle 5.4 e 5.5* sono quelli di *default*, ottenuti come è stato mostrato nell'applicazione presentata nel Paragrafo 4.3.3.

Con la finalità di avere una comprensione quanto più possibile completa non solamente di tutto il periodo di osservazione ma, nello specifico, della realtà che risulta più vicina a quella odierna, si è deciso di approfondire lo studio della relazione tra le tre variabili nella sesta wave. Per questo motivo sono stati svolti ulteriori confronti, valutando gli effetti in tre situazioni differenti, caratterizzate dal diverso valore assunto dai livelli di esposizione confrontati:

- livello di esposizione medio *vs* livello minimo;
- livello di esposizione massimo *vs* livello medio;
- livello di esposizione massimo *vs* livello minimo.

Queste tre situazioni sono state valutate anche cambiando il valore a cui viene fissato il mediatore per il calcolo del *CDE*, facendo sì che assuma valore medio, massimo e minimo alternativamente, in modo da avere una visione ancora più complessiva del fenomeno e delle azioni che possono essere attuate sulla salute fisica per migliorare il benessere degli individui.

I risultati ottenuti per i due campioni analizzati sono riportati nelle *Table* 5.6 e 5.7.

Da questa ulteriore analisi appaiono delle relazioni ancora più interessanti, infatti, è possibile notare immediatamente come, mentre quando si confrontano i livelli medi e minimi del SES gli effetti assomigliano a quelli ottenuti precedentemente, nel momento in cui vengono confrontati valori alti di SES con valori medi oppure bassi l'entità degli effetti aumenta notevolmente.

In questo modo si può osservare una cosa che prima non era emersa con la medesima importanza: mentre passando da valori vicini tra di loro e intorno alla media oppure anche tra valori medi e bassi di SES, il benessere subisce un leggero miglioramento che risulta comunque significativo, è dal confronto dei valori alti e medi oppure alti e bassi che emergono le reali differenze.

Gli effetti diretti sul benessere confrontando valori alti di SES con valori medi o bassi, risultano essere circa 28 volte più grandi rispetto a quelli osservati dal confronto di *default* oppure confrontando valori medi di SES con valori bassi; quelli indiretti risultano essere 15 volte più grandi.

Questo mette in evidenza il fatto che possono essere rilevate delle differenze maggiori nel benessere quando vengono confrontati livelli di SES alti con quelli medi oppure con quelli bassi. Questo potrebbe indicare che effettivamente i soggetti più benestanti possono fare ricorso a cure migliori e a un miglior sistema di supporto anche per la qualità della loro salute psicologica. L'aumento dell'effetto indiretto mette ancora più in evidenza il legame in età avanzata dell'importanza di considerare la salute fisica come mediatore nello studio della relazione tra il SES e il benessere.

Tutti gli effetti nel campione completo risultano essere significativi, mentre quelli indiretti nel campione panel nati dagli ultimi due confronti non lo sono. Questo potrebbe essere giustificato dal fatto che nel campione riferito al panel generalmente sono stati selezionati soggetti con ricchezza maggiore e nella sesta wave sono presenti soggetti con età media superiore, per i quali le condizioni fisiche legate al processo di invecchiamento sono inevitabili e quindi possono essere legate in maniera meno forte alla ricchezza e all'accesso alle cure migliori.

	TE	CDE ($m = \text{mean} \text{min} \text{max}$)	NDE	NIE
$a = \bar{a} + 0.5$ $a^* = \bar{a} - 0.5$ (default)	0.1843 (0.1740; 0.1946)	0.1331 (0.1236; 0.1426)	0.1334 (0.1237; 0.1431)	0.0509 (0.04631; 0.05548)
		0.1723 (0.08354; 0.2610)		
$a = \bar{a}$ $a^* = \text{min}$	0.1311 (0.1236; 0.1386)	0.1300 (0.1202; 0.1398)	0.0948 (0.08778; 0.1018)	0.0364 (0.03318; 0.03954)
		0.0944 (0.08770; 0.1012)		
$a = \text{max}$ $a^* = \bar{a}$	3.4276 (2.8460; 4.0093)	0.0923 (0.08533; 0.09922)	2.6732 (2.4825; 2.8638)	0.7545 (0.1437; 1.3653)
		2.6732 (2.4825; 2.8638)		
$a = \text{max}$ $a^* = \text{min}$	3.5588 (2.9771; 4.1405)	2.6121 (2.4154; 2.8088)	2.7777 (2.5726; 2.9827)	0.7811 (0.1487; 1.4135)
		2.7676 (2.5702; 2.9650)		
		2.7044 (2.5008; 2.9080)		

Tabella 5.6 – Stime degli effetti (e relativi intervalli di confidenza al 95%) del modello con solo esposizione e mediatore per la risposta nella sesta wave, nel campione completo, confrontando diversi valori di esposizione

	TE	CDE ($m = \text{mean} \text{min} \text{max}$)	NDE	NIE
$a = \bar{a} + 0.5$ $a^* = \bar{a} - 0.5$ (default)	0.2182 (0.1926; 0.2437)	0.1502 (0.1260; 0.1745)	0.1532 (0.1276; 0.1788)	0.0650 (0.05347; 0.07648)
		0.3906 (0.1693; 0.6118)		
		0.1268 (0.1037; 0.1499)		
$a = \bar{a}$ $a^* = \text{min}$	0.1755 (0.1537; 0.1973)	0.1183 (0.09919; 0.1374)	0.1220 (0.1011; 0.1428)	0.0535 (0.04465; 0.06236)
		0.3075 (0.1333; 0.4817)		
		0.0999 (0.08169; 0.1180)		
$a = \text{max}$ $a^* = \bar{a}$	1.8700 (0.9930; 2.7469)	2.0503 (1.7194; 2.3813)	2.0503 (1.7191; 2.3816)	-0.1804 (-1.1671; 0.8063)
		5.3307 (2.3110; 8.3504)		
		1.7308 (1.4159; 2.0457)		
$a = \text{max}$ $a^* = \text{min}$	2.0454 (1.1734; 2.9175)	2.1686 (1.8186; 2.5187)	2.2362 (1.8544; 2.6181)	-0.1908 (-1.2345; 0.8529)
		5.6382 (2.4443; 8.8321)		
		1.8307 (1.4976; 2.1637)		

Tabella 5.7 – Stime degli effetti (e relativi intervalli di confidenza al 95%) del modello con solo esposizione e mediatore per la risposta nella sesta wave, nel campione riferito al panel, confrontando diversi valori di esposizione

Traendo queste prime conclusioni tuttavia bisogna prestare molta attenzione poiché le assunzioni che stanno alla base di questo metodo, non essendo inserite covariate per il controllo del confondimento, non sono rispettate e di conseguenza alle stime di questi effetti non è possibile dare nessuna interpretazione causale. Come già detto, questa fase risulta comunque utile per una prima comprensione più approfondita del fenomeno, al fine di valutare anche dal punto di vista empirico l'importanza del rispetto delle assunzioni e le differenze che emergono una volta considerati i confondenti.

I valori delle stime dei coefficienti e degli effetti possono sembrare piccole ma bisogna ricordare, come è stato mostrato nella prima parte di analisi descrittive, che le variabili che fungono da mediatore e da risposta sono frutto di un'analisi fattoriale e hanno un range di valori abbastanza limitato (*Tabelle 5.1 e 5.2*). Cambiamenti piccoli possono essere in realtà grandi una volta valutata la scala di valori che si ha di fronte.

5.3 L'effetto di mediazione con covariate

Il fenomeno in esame è complesso e l'eccessiva semplificazione svolta nei primi modelli non può portare a risultati affidabili. Per questo motivo e per permettere che siano rispettate le assunzioni sottostanti all'approccio controfattuale per i modelli di mediazione, è cruciale l'inclusione di altre covariate e di tutte quelle variabili che possono fungere da confondenti nelle relazioni ipotizzate.

Si ricorda che con il termine confondimento si fa riferimento alla distorsione e inaccuratezza delle stime delle misure di associazione che avviene quando sia il trattamento di interesse sia l'outcome risultano associati ad una terza variabile (LaMorte e Sullivan, 2016).

Dalle analisi descrittive iniziali e dallo studio della letteratura trovata sono state selezionate nove variabili ritenute indispensabili all'analisi e che risultano significativamente associate ad almeno due delle tre variabili chiave del modello di mediazione.

La prima variabile fondamentale è il genere dell'individuo (variabile dicotomica che indica se un individuo è maschio o femmina), che insieme all'età (variabile con cinque livelli che divide gli individui nelle seguenti classi dall'età: "< 50", "50 – 59", "60 – 69", "70 – 79", "80 +") sono i principali confondenti in tutte le analisi che vengono svolte. Si tratta di variabili che possono influire su tutte e tre le variabili chiave dell'analisi. Per quanto riguarda lo stato socio-economico degli individui, l'età risulta fondamentale in quanto solitamente ad età più avanzate si ha una ricchezza maggiore, dovuta in gran parte anche al suo accumulo nel tempo. Nonostante la società

odierna risulti essere molto più paritaria rispetto al passato, in molti paesi l'equità fra i sessi non è raggiunta nell'ambito delle retribuzioni e di conseguenza delle ricchezze accumulate, quindi anche il genere può risultare una variabile chiave nella valutazione del fenomeno e nella comprensione delle interrelazioni tra le variabili. Allo stesso modo il sesso può influire sulla salute fisica, soprattutto perché all'avanzare dell'età vengono sviluppate patologie molto differenti a seconda che un individuo sia uomo o donna. Anche il benessere psicologico risulta variare molto sia con il genere che con l'età specialmente per i diversi sistemi di supporto che si hanno al variare dei momenti della vita dei soggetti.

Il livello di istruzione (variabile categoriale con tre livelli, che indicano se un individuo ha un livello di istruzione basso, medio o alto), lo stato civile (variabile categoriale con cinque livelli che indicano se un individuo è divorziato/a o separato/a, sposato/a, celibe/nubile, in un'unione registrata, vedovo/a) e il paese di residenza (variabile categoriale con nove livelli che indica il paese di residenza dell'individuo tra Austria, Belgio, Danimarca, Francia, Germania, Italia, Spagna, Svezia e Svizzera) possono essere considerati ulteriori possibili confondenti della relazione tra esposizione e outcome. Specialmente i paesi ricoprono un ruolo fondamentale, poiché rispecchiano diverse realtà e diversi sistemi di governo e di welfare che concedono agli individui possibilità economiche e sanitarie differenti.

Anche la condizione abitativa (variabile categoriale con due livelli che indica se un individuo è proprietario oppure non proprietario dell'abitazione in cui risiede), l'indice di massa corporea (variabile categoriale con quattro livelli che indica se un individuo risulta essere sottopeso, normopeso, sovrappeso oppure obeso) e le due variabili relative all'attività fisica svolta a seconda che si tratti di attività fisica vigorosa oppure che richieda un moderato livello di energia (variabili categoriali con quattro livelli che indicano la frequenza dell'attività fisica svolta: più di una volta alla settimana, una volta alla settimana, da una a tre volte al mese oppure quasi mai o mai) rientrano nel

gruppo delle variabili fondamentali per la definizione del processo di mediazione di interesse.

Le variabili relative all'attività fisica possono tuttavia risultare difficili da trattare, poiché quando si tratta della relazione fra i problemi di salute e l'attività fisica svolta, si corre il rischio di imbattersi in quella che viene chiamata *reverse causality*. Quando vengono trattate le relazioni tra queste due variabili, infatti, non si ha sempre la certezza della direzione del nesso causale. Trovandosi in una situazione di incertezza, in questa fase delle analisi viene esclusa l'attività fisica tra i possibili confondenti. Successivamente, tuttavia, ampliando l'analisi e valutando il fenomeno dal punto di vista longitudinale, sarà possibile sfruttare la temporalità e le variabili identificate in momenti diversi per escludere la possibilità di causalità inversa e valutare solamente l'influenza che l'attività fisica precedente ha sullo stato di salute presente mediante la corretta specificazione dei pesi. In questo modo possono essere inserite anche le due variabili relative all'attività fisica nei modelli.

Nella *Figura 5.2* è possibile avere una visione grafica delle relazioni di cui si è appena parlato e comprendere come la mancata inclusione di alcune variabili porterebbe immediatamente alla violazione delle assunzioni necessarie a dare un'interpretazione causale agli effetti stimati.

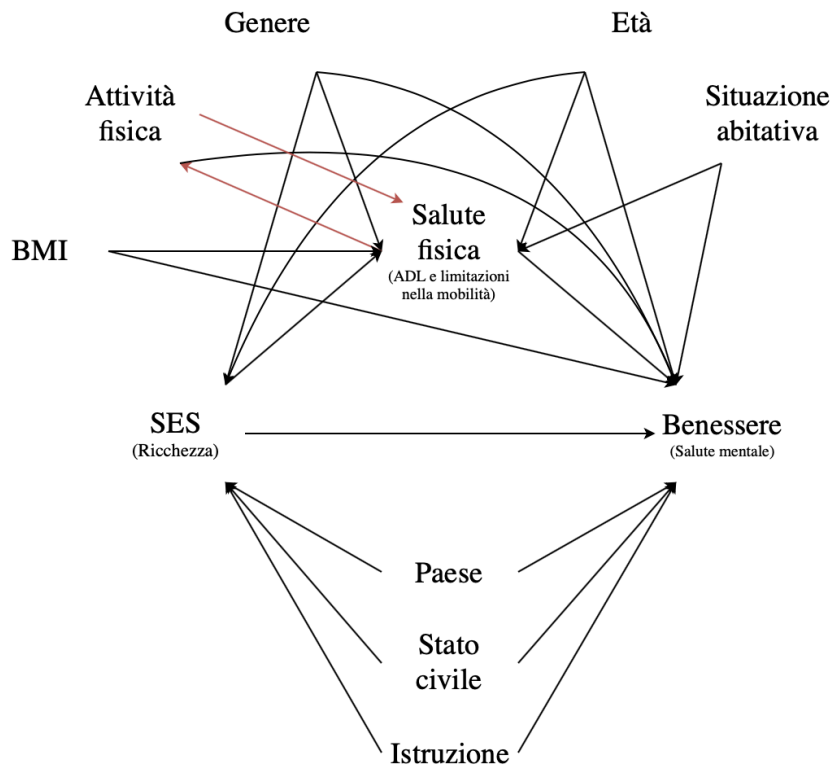


Figura 5.2 – Rappresentazione grafica delle relazioni ipotizzate tra esposizione, mediatore e risposta con i possibili confondenti

Le variabili che entrano in scena nella valutazione del fenomeno sono numerose e tutte le relazioni che le coinvolgono complicano le analisi, ma permettono di ottenere delle stime più accurate degli effetti. Tuttavia, con i metodi basati sulla regressione, è abbastanza immediata e semplice la loro inclusione in quanto tutto il problema si riduce nuovamente alla stima di due modelli semplicemente definiti come

$$E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c$$

$$E(M|A = a, C = c) = \theta_0^\dagger + \theta_1^\dagger a + \theta_4^\dagger c$$

nei quali le variabili A , M e Y sono uguali a quelle definite precedentemente, ma rispetto a prima vengono aggiunte in entrambe le regressioni le covariate, identificate con il generico vettore C .

Le analisi sono state svolte sugli stessi campioni di prima, solamente ampliando i modelli.

Aggiungendo il controllo per il confondimento, si cerca di evitare la violazione delle assunzioni sottostanti l'approccio controfattuale nell'analisi di mediazione (Formule 4.13 – 4.16) e si ottengono stime più accurate rispetto alle precedenti, con le quali sarà possibile fare un confronto per valutare la direzione dell'effetto dovuto all'errore nella stima.

Si procede quindi mostrando i risultati dei modelli applicati in ogni wave separatamente e valutando similitudini e differenze.

5.3.1 Risultati

Ampliando i modelli con l'inserimento dei confondenti emergono diverse cose interessanti che confermano l'importanza delle assunzioni.

I valori specifici dei coefficienti con i relativi intervalli di confidenza e livelli di significatività sono riportati nell'Appendice F per i campioni completi e nell'Appendice G per i campioni riferiti al panel.

La prima cosa che si osserva è un cambiamento rispetto ai modelli del Paragrafo 5.2.1 nelle stime degli effetti; mentre inizialmente, senza correggere per l'effetto dei confondenti, tutti i coefficienti erano significativi, adesso non è più così.

Nei modelli per la salute fisica degli individui, il coefficiente relativo allo stato socio-economico porta sempre il segno positivo, a conferma di quello che è stato detto della relazione positiva della ricchezza sullo stato di salute, e risulta significativo in tutte le wave, sia sul dataset completo che sul dataset del panel. Questo evidenzia l'importanza della relazione tra salute fisica e stato di salute e il ruolo di mediatore svolto da quest'ultima, confermato anche dopo il controllo delle covariate.

Confrontando i risultati ottenuti precedentemente nei modelli che non includevano i confondenti, era emerso come i coefficienti ottenuti dall'applicazione dei modelli nel dataset di panel presentava dei coefficienti in valore assoluto superiori, rispetto a quelli ottenuti dalla stima dei modelli

nel dataset completo. Non è possibile riscontrare la stessa differenza quando viene effettuato il controllo del confondimento dei modelli. Come era già stato sottolineato, infatti, vi era stata una selezione dei soggetti che aveva incluso solamente quelli con determinate caratteristiche ma, una volta inserito il controllo per tali caratteristiche, l'andamento risulta molto più in linea con quello che è stato ottenuto con i dataset completi.

Sia considerando il campione completo, sia il panel, controllando per i possibili confondenti nel modello per la salute fisica, le stime dei coefficienti del SES risultano essere inferiori rispetto a quelli ottenuti precedentemente, confermando l'importanza del controllo per un'identificazione corretta degli effetti.

Nelle analisi svolte sui dataset riferiti al panel, tra le covariate inserite nel modello per il mediatore non risultano significative Danimarca e Svezia rispetto alla Svizzera per la variabile relativa al paese di provenienza. Non risulta significativa neppure la variabile relativa allo stato civile dei soggetti che, insieme al paese di provenienza, erano state precedentemente ipotizzate confondenti per la relazione tra esposizione e risposta. Appare più sorprendente la mancata significatività dell'età, che invece si pensava dovesse essere fondamentale per la descrizione di un fenomeno come lo stato di salute della popolazione presa in considerazione.

Nel dataset completo l'unico coefficiente che non risulta significativo è quello della Svezia nel confronto con la Svizzera quando si considera il paese di provenienza.

Questo indica ancora una volta il problema legato alla selezione dei soggetti e la conseguente differenza nei valori delle stime.

Come ci si aspettava, invece, le variabili relative al genere, alla situazione abitativa, al BMI, allo stato civile e all'istruzione sono risultate significative e permettono di descrivere meglio il modello per il mediatore. Al netto di tutte le altre variabili, l'essere donna corrisponde a condizioni di salute peggiori rispetto a quelle dell'uomo. Al contrario, i soggetti con livelli di istruzione

medi o alti, rispetto a coloro con livelli bassi, presentano condizioni di salute migliori.

L'indice di massa corporea, come previsto, ha un ruolo importante nell'influenzare la condizione di salute, risultando significativo.

Rispetto agli obesi, nelle prime wave, i soggetti sovrappeso e normopeso sono in condizioni migliori mentre i soggetti sottopeso presentano condizioni peggiori. Per questi ultimi il coefficiente non risulta significativo, probabilmente poiché tra i soggetti obesi e quelli sottopeso le condizioni di salute risultano simili.

Questo andamento viene generalmente riscontrato in tutte le wave senza particolari differenze.

Vengono valutati adesso i coefficienti delle variabili nei modelli per la risposta delle varie wave e successivamente saranno analizzate le stime degli effetti diretti, indiretti e totali del modello di mediazione.

Sembra importante sottolineare come non emergano particolari differenze tra le wave nemmeno nei modelli per la risposta, se non per piccoli cambiamenti nei coefficienti che non risultano quasi mai significativi.

Le categorie della variabile relativa all'età non risultano quasi mai significative nei modelli per il benessere, che contengono sia esposizione sia mediatore, in nessuno dei dataset considerati. Evidentemente il benessere, una volta considerate tutte le possibili variabili che lo influenzano, non cambia a seconda della classe di età di appartenenza.

Questo probabilmente può essere dovuto anche al fatto che, indipendentemente dall'età, quello che effettivamente determina situazioni di disagio sia maggiormente legato alle difficoltà fisiche che i soggetti presentano. I coefficienti relativi alla salute fisica in tutte le wave e in entrambi i campioni risultano essere significativi e indicano come al diminuire delle limitazioni funzionali il benessere psicologico aumenti. Nel campione complessivo i coefficienti presentano valori più alti rispetto a quelli ottenuti dall'applicazione dei modelli al panel, che comprende solamente i soggetti che partecipano a tutte le wave.

Anche lo stato socio-economico porta un coefficiente significativo e con segno positivo, testimone del fatto che nei campioni analizzati, ad una crescita della ricchezza corrisponde anche una crescita del benessere.

Il termine di interazione tra le difficoltà fisiche e il SES non risulta essere significativo e ciò può essere un indicatore del fatto che entrambi hanno un ruolo importante nello spiegare il fenomeno, ma non hanno un comportamento sinergico e agiscono con canali separati. Il termine di interazione è stato mantenuto all'interno dei modelli anche senza che fosse significativo, perché la sua esclusione portava ad un cambiamento di circa il 10% nelle stime degli effetti totali che verranno analizzate in seguito.

La situazione abitativa non risulta influire sul benessere.

Il genere del soggetto ha un effetto significativo, con le donne che rispetto agli uomini risultano avere un benessere psicologico inferiore. Come prima, anche il livello di istruzione ha un effetto significativo positivo sul benessere che cresce al crescere del livello educativo raggiunto.

La variabile relativa allo stato civile risulta essere significativa e con un effetto positivo solamente nella categoria di soggetti sposati e che vivono con il marito/moglie, oppure che sono in un'unione registrata, rispetto a coloro che sono vedovi.

L'ultima variabile a cui è importante far riferimento è il paese di provenienza dei soggetti. Come detto precedentemente è noto, ed è stato studiato numerose volte, il fatto che il benessere della popolazione sia molto spesso associato al paese di provenienza. Anche nello studio che si sta presentando sembra emergere qualcosa di simile. Tutti i paesi, ad eccezione della Svezia, dell'Austria e della Germania, risultano avere un coefficiente significativo rispetto alla Svizzera. Questa mancata significatività probabilmente è dovuta al fatto che la Svezia e l'Austria sono i paesi che in quanto a risorse economiche più si avvicinano alla Svizzera, mantenuta come paese di riferimento.

Alla Danimarca corrispondono coefficienti con segno positivo che indicano condizioni di benessere per i soggetti residenti migliori rispetto a quelli della Svizzera.

La Francia, l'Italia, la Spagna e il Belgio invece hanno coefficienti con segno negativo e rispetto alla Svizzera i soggetti residenti si trovano in condizioni di benessere inferiore, in ordine decrescente ci sono Belgio, Francia, Spagna e infine Italia che si trova nella condizione peggiore.

Quanto detto nelle ultime righe può essere interessante anche valutandolo da un punto di vista economico, in quanto i paesi che riportano situazioni di benessere inferiore sono quelli che risultano essere non solo meno ricchi degli altri, ma anche quelli che sono stati maggiormente colpiti dalla crisi economica. I valori dei coefficienti sembrano confermare quanto detto, poiché nelle prime e nelle ultime wave si trovano valori abbastanza simili, ad indicare la situazione iniziale, già problematica e poi quella di lenta ripresa che sta avvenendo dopo la crisi. Nelle wave centrali invece le distanze tra i paesi sembrano accentuarsi, ma si osserva comunque un calo generale del benessere, che può essere dovuto al crollo economico che c'è stato in Europa dal 2008.

Di seguito nelle *Tablelle 5.8 e 5.9* sono riportate le stime degli effetti e le percentuali degli stessi dovuti alla mediazione nelle diverse wave e rispettivamente nel dataset completo e in quello con solamente i soggetti che partecipano a tutte le wave.

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value
Wave 1	Effetto Totale	0.0887	0.00937	0.07029	0.1070	9.46	<.0001
	Controlled Direct Effect (CDE)	0.0715	0.00872	0.05446	0.08863	8.21	<.0001
	Natural Direct Effect (NDE)	0.0710	0.00877	0.05385	0.08824	8.10	<.0001
	Natural Indirect Effect (NIE)	0.0176	0.00366	0.01043	0.02479	4.81	<.0001
	Percentuale di effetto dovuta alla mediazione	19.8632	3.9188	12.1824	27.5439	5.07	<.0001
	Percentuale di effetto dovuta all'interazione	0.5677	0.3475	-0.1133	1.2487	1.63	0.1023
Wave 2	Effetto Totale	0.0817	0.00739	0.06722	0.09618	11.06	<.0001
	Controlled Direct Effect (CDE)	0.0638	0.00683	0.05042	0.07719	9.34	<.0001
	Natural Direct Effect (NDE)	0.0639	0.00690	0.05041	0.07745	9.27	<.0001
	Natural Indirect Effect (NIE)	0.0178	0.00294	0.01201	0.02353	6.04	<.0001
	Percentuale di effetto dovuta alla mediazione	21.7501	3.4234	15.0403	28.4599	6.35	<.0001
	Percentuale di effetto dovuta all'interazione	-0.1509	0.2376	-0.6166	0.3147	-0.64	0.5253
Wave 4	Effetto Totale	0.0705	0.00608	0.05858	0.08241	11.60	<.0001
	Controlled Direct Effect (CDE)	0.0598	0.00564	0.04870	0.07081	10.60	<.0001
	Natural Direct Effect (NDE)	0.0601	0.00568	0.04901	0.07126	10.59	<.0001
	Natural Indirect Effect (NIE)	0.0104	0.00224	0.005975	0.01474	4.63	<.0001
	Percentuale di effetto dovuta alla mediazione	14.6949	2.9658	8.8820	20.5078	4.95	<.0001
	Percentuale di effetto dovuta all'interazione	-0.5425	0.2645	-1.0609	-0.02402	-2.05	0.0403
Wave 5	Effetto Totale	0.0840	0.00527	0.07366	0.09430	15.95	<.0001
	Controlled Direct Effect (CDE)	0.0693	0.00484	0.05979	0.07875	14.32	<.0001
	Natural Direct Effect (NDE)	0.0700	0.00487	0.06042	0.07952	14.36	<.0001
	Natural Indirect Effect (NIE)	0.0140	0.00202	0.01004	0.01797	6.93	<.0001
	Percentuale di effetto dovuta alla mediazione	16.6768	2.2312	12.3037	21.0498	7.47	<.0001
	Percentuale di effetto dovuta all'interazione	-0.8359	0.2625	-1.3504	-0.3214	-3.18	0.0015
Wave 6	Effetto Totale	0.0767	0.00600	0.06492	0.08846	12.77	<.0001
	Controlled Direct Effect (CDE)	0.0631	0.00557	0.05216	0.07398	11.33	<.0001

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value
Natural Direct Effect (NDE)	0.0632	0.00559	0.05225	0.07418	11.30	<.0001
Natural Indirect Effect (NIE)	0.0135	0.00228	0.009014	0.01793	5.92	<.0001
Percentuale di effetto dovuta alla mediazione	17.5684	2.7808	12.1181	23.0186	6.32	<.0001
Percentuale di effetto dovuta all'interazione	-0.1968	0.1471	-0.4851	0.09148	-1.34	0.1809

Tabella 5.8 – Stime degli effetti del modello con esposizione, mediatore e covariate per la risposta nelle diverse wave nel campione completo

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value	
Wave 1	Effetto Totale	0.1054	0.0153	0.07545	0.1353	6.90	<.0001
	Controlled Direct Effect (CDE)	0.0932	0.0143	0.06507	0.1213	6.50	<.0001
	Natural Direct Effect (NDE)	0.0928	0.0144	0.06463	0.1211	6.45	<.0001
	Natural Indirect Effect (NIE)	0.0125	0.00553	0.001683	0.02335	2.26	0.0235
	Percentuale di effetto dovuta alla mediazione	11.8810	4.9467	2.1857	21.5763	2.40	0.0163
	Percentuale di effetto dovuta all'interazione	0.3026	0.3711	-0.4248	1.0300	0.82	0.4148
Wave 2	Effetto Totale	0.0756	0.0130	0.05014	0.1011	5.82	<.0001
	Controlled Direct Effect (CDE)	0.0604	0.0122	0.03651	0.08425	4.96	<.0001
	Natural Direct Effect (NDE)	0.0602	0.0123	0.03613	0.08423	4.90	<.0001
	Natural Indirect Effect (NIE)	0.0154	0.00481	0.006001	0.02485	3.21	0.0013
	Percentuale di effetto dovuta alla mediazione	20.3999	6.1628	8.3211	32.4787	3.31	0.0009
	Percentuale di effetto dovuta all'interazione	0.2622	0.4571	-0.6337	1.1582	0.57	0.5662
Wave 4	Effetto Totale	0.0778	0.0132	0.05201	0.1037	5.91	<.0001
	Controlled Direct Effect (CDE)	0.0665	0.0124	0.04227	0.09069	5.38	<.0001
	Natural Direct Effect (NDE)	0.0675	0.0125	0.04303	0.09187	5.41	<.0001
	Natural Indirect Effect (NIE)	0.0104	0.00428	0.001992	0.01877	2.43	0.0153
	Percentuale di effetto dovuta alla mediazione	13.3351	5.2207	3.1027	23.5676	2.55	0.0106
	Percentuale di effetto dovuta all'interazione	-1.2515	1.0666	-3.3420	0.8390	-1.17	0.2406

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		Z	P – value	
Wave 5	Effetto Totale	0.0736	0.0129	0.04829	0.09890	5.70	<.0001
	Controlled Direct Effect (CDE)	0.0603	0.0118	0.03729	0.08339	5.13	<.0001
	Natural Direct Effect (NDE)	0.0607	0.0118	0.03754	0.08382	5.14	<.0001
	Natural Indirect Effect (NIE)	0.0129	0.00526	0.002599	0.02324	2.45	0.0141
	Percentuale di effetto dovuta alla mediazione	17.5521	6.5442	4.7258	30.3785	2.68	0.0073
	Percentuale di effetto dovuta all'interazione	-0.4592	0.4588	-1.3584	0.4401	-1.00	0.3169
Wave 6	Effetto Totale	0.0839	0.0149	0.05480	0.1130	5.65	<.0001
	Controlled Direct Effect (CDE)	0.0706	0.0139	0.04349	0.09779	5.10	<.0001
	Natural Direct Effect (NDE)	0.0712	0.0140	0.04378	0.09856	5.09	<.0001
	Natural Indirect Effect (NIE)	0.0127	0.00537	0.002222	0.02328	2.37	0.0176
	Percentuale di effetto dovuta alla mediazione	15.1922	6.0402	3.3537	27.0307	2.52	0.0119
	Percentuale di effetto dovuta all'interazione	-0.6287	0.5717	-1.7493	0.4919	-1.10	0.2715

Tabella 5.9 – Stime degli effetti del modello con esposizione, mediatore e covariate per la risposta nelle diverse wave nel campione riferito al panel

Emerge uno scenario diverso da quello che si presentava in precedenza.

Dando un primo sguardo alle tabelle sopra riportate, è possibile osservare che i valori delle stime degli effetti totali, diretti e indiretti, della relazione tra SES e benessere mediata dalla salute fisica, risultano essere significative in tutte le wave, in entrambi i campioni.

Una volta inserite tutte le covariate e i confondenti all'interno dei modelli, cambiano le stime degli effetti, mentre si mantiene la loro significatività. Si osserva un notevole calo nella dimensione delle stime, che risultano essere dimezzate rispetto a quelle presentate nel Paragrafo 5.2.1. Non si può negare la presenza di un effetto sia diretto sia indiretto, ma una volta inclusi tutti i possibili confondenti che agiscono sul fenomeno è possibile affermare che gli effetti sono molto meno forti di quelli che si erano osservati originariamente nei modelli senza correzione per il confondimento.

L'esclusione delle variabili che fungono da confondenti porta le stime degli effetti a risultare *inflated*.

I valori degli effetti sono ottenuti seguendo nei modelli il metodo proposto da VanderWeele per la scelta dell'inclusione o esclusione dell'interazione tra le variabili relative allo stato socio-economico e alla salute fisica degli individui. Poiché l'effetto totale subiva una variazione superiore al 10% quando veniva esclusa l'interazione tra le variabili, si è deciso di mantenerla in tutti i modelli.

L'effetto diretto rappresenta la quasi totalità dell'effetto totale individuato dai modelli in tutte le wave, e anche confrontando i risultati delle diverse wave, tramite la valutazione degli intervalli di confidenza nel campione panel, non è possibile individuare delle differenze significative tra una wave e l'altra.

All'effetto indiretto è possibile attribuire dal 10% al 20% dell'effetto totale confrontando tra di loro le wave, ma nessuno di questi valori risulta essere significativamente diverso l'uno dall'altro.

È quindi possibile confermare l'importanza della salute fisica nello studio della relazione tra lo stato socio-economico e il benessere. Per approfondire ancora di più il legame tra queste variabili, come è stato fatto precedentemente, si è deciso di studiare in che modo cambiano le stime degli effetti confrontando differenti livelli di esposizione e mediatore, all'interno della sesta wave. In questo modo è possibile vedere in che modo l'inclusione delle covariate ha impattato su queste stime.

I risultati ottenuti per i due campioni analizzati sono riportati nelle *Tabelle 5.10* e *5.11*.

Da questo approfondimento emerge uno scenario molto simile a quello evidenziato nelle *Tabelle 5.6* e *5.7*, con l'unica differenza sostanziale legata alla riduzione delle stime dovuta all'introduzione delle variabili confondenti. Questo cambiamento è legato alla riduzione dei valori dei coefficienti che erano stati sottolineati anche all'inizio del paragrafo.

Come precedentemente mostrato emerge che, nonostante siano presenti delle differenze significative nelle condizioni di benessere confrontando soggetti con SES medio con soggetti con SES basso, i valori delle stime degli effetti diretti e totali incrementano notevolmente quando vengono valutate le differenze tra valori alti della ricchezza con valori medi o bassi.

Questo aumento nelle stime degli effetti risulta essere meno pronunciato quando viene svolto il confronto all'interno del campione riferito al panel, nel quale gli effetti indiretti quando si confrontano valori di SES alti con valori bassi oppure valori alti con valori medi, non risultano significativi.

La non significatività degli effetti associata anche ad una stima negativa degli effetti indiretti, si pensa sia dovuta al fatto che il legame tra SES e salute fisica in soggetti anziani¹⁷ venga meno poiché, indipendentemente dalla ricchezza di ciascuno di essi, la salute fisica in età avanzata potrà essere migliorata fino ad un certo punto, senza poter tuttavia invertire o modificare in maniera significativa il processo di invecchiamento che tutti attraversano. Anche nel campione completo gli effetti indiretti non risultano significativi negli ultimi due confronti svolti tra livelli di esposizione, ma in misura molto minore rispetto a quelli del campione panel, con un p-value di 0.053.

Questo permette di sottolineare il fatto che sono presenti delle differenze di benessere legate alla diversa condizione economica degli individui ma, tuttavia, queste differenze, quando vengono confrontati livelli di SES alti con livelli medi oppure bassi, non sono attribuibili ad una differente condizione di salute.

Le differenze nelle stime degli effetti tra i due campioni sono legate al modo in cui è stato svolto il campionamento, che non rispetta la reale composizione della popolazione che è stata originariamente analizzata.

¹⁷ Si ricorda che nella sesta wave l'età media dei soggetti nel campione riferito al panel è di 73 anni.

	TE	CDE ($m = \text{mean} \text{min} \text{max}$)	NDE	NIE
$a = \bar{a} + 0.5$ $a^* = \bar{a} - 0.5$ (default)	0.1286 (0.1177; 0.1395)	0.1029 (0.09278; 0.1131)	0.1032 (0.09290; 0.1134)	0.0254 (0.02127; 0.02959)
		0.0989 (0.08860; 0.1092)		
$a = \bar{a}$ $a^* = \text{min}$	0.0919 (0.08404; 0.09979)	0.0734 (0.06613; 0.08059)	0.0736 (0.06626; 0.08097)	0.0183 (0.01533; 0.02127)
		0.1116 (0.04852; 0.1747)		
$a = \text{max}$ $a^* = \bar{a}$	2.3849 (2.0578; 2.7120)	0.0705 (0.06315; 0.07781)	2.0673 (1.8636; 2.2711)	0.3175 (-0.00485; 0.6399)
		3.1457 (1.3672; 4.9243)		
$a = \text{max}$ $a^* = \text{min}$	2.4768 (2.1469; 2.8067)	1.9861 (1.7796; 2.1926)	2.1480 (1.9333; 2.3626)	0.3288 (-0.00502; 0.6626)
		2.1407 (1.9297; 2.3517)		
		2.0566 (1.8428; 2.2704)		

Tabella 5.10 – Stime degli effetti (e relativi intervalli di confidenza al 95%) del modello con esposizione, mediatore e covariate per la risposta nella sesta wave, nel campione completo, confrontando diversi valori di esposizione

	TE	CDE ($m = \text{mean} \text{min} \text{max}$)	NDE	NIE
$a = \bar{a} + 0.5$ $a^* = \bar{a} - 0.5$ (default)	0.1360 (0.1089; 0.1632)	0.1056 (0.07990; 0.1313)	0.1068 (0.08050; 0.1331)	0.0292 (0.01932; 0.03911)
		0.3115 (0.08942; 0.5335)		
		0.0862 (0.06234; 0.1100)		
$a = \bar{a}$ $a^* = \text{min}$	0.1094 (0.08699; 0.1318)	0.0838 (0.06338; 0.1041)	0.0853 (0.06412; 0.1064)	0.0241 (0.01611; 0.03215)
		0.2471 (0.07093; 0.4232)		
		0.0683 (0.04945; 0.08725)		
$a = \text{max}$ $a^* = \bar{a}$	1.4048 (0.9738; 1.8357)	1.4405 (1.0901; 1.7909)	1.4405 (1.0900; 1.7910)	-0.0358 (-0.5041; 0.4326)
		4.2493 (1.2200; 7.2786)		
		1.1755 (0.8505; 1.5006)		
$a = \text{max}$ $a^* = \text{min}$	1.5142 (1.0776; 1.9508)	1.5243 (1.1535; 1.8950)	1.5520 (1.1669; 1.9371)	-0.0378 (-0.5334; 0.4577)
		4.4963 (1.2909; 7.7018)		
		1.2439 (0.9000; 1.5878)		

Tabella 5.11 – Stime degli effetti (e relativi intervalli di confidenza al 95%) del modello con solo esposizione, mediatore e covariate per la risposta nella sesta wave, nel campione riferito al panel, confrontando diversi valori di esposizione

Da quanto detto fino ad ora in questo capitolo è quindi possibile riassumere il tutto dicendo che sicuramente vi è un effetto dello stato socio-economico e della salute fisica sul benessere degli individui, ma che la mediazione ipotizzata inizialmente da parte della condizione fisica dei soggetti non risulta essere sempre confermata dai risultati di quest'analisi.

Confrontando livelli di SES simili tra di loro, emergono delle differenze significative nelle condizioni di benessere sia legate all'effetto diretto dello stato socio-economico, sia all'effetto indiretto della salute fisica. Quando si confrontano invece livelli di SES molto differenti tra di loro, emergono degli effetti significativi diretti sul benessere, ma viene meno la significatività degli

effetti indiretti. Questo è dovuto al fatto che il secondo confronto viene svolto tra valori massimi e minimi del SES, mentre la maggior parte dei soggetti si trova in condizioni intermedie. Inoltre, l'analisi è svolta a livello europeo, e gli standard di salute e di accesso alle cure mediche risultano essere considerevolmente alti.

La correzione dell'effetto del confondimento è risultata fondamentale, in quanto le stime degli effetti ottenute successivamente alla loro considerazione risultano essere più piccole e si sarebbe potuti giungere a conclusioni parzialmente differenti senza la loro inclusione.

È tuttavia fondamentale sottolineare il fatto che si tratta di un fenomeno molto complesso, nel quale alcune variabili potrebbero non essere state prese in considerazione.

Quello che è stato fino ad ora presentato non è ancora la conclusione delle analisi, poiché, nonostante si sia cercato di analizzare i risultati anche con un'ottica longitudinale, quello che è stato fatto è una lettura trasversale dei risultati per ogni wave, proponendo dei confronti tra di loro per valutare le possibili differenze e dando delle giustificazioni legate al cambiamento socio-politico dello scenario europeo che si è venuto a verificare dal 2004 fino al 2015.

Le analisi verranno ulteriormente ampliate nel capitolo successivo prendendo in considerazione il possibile effetto di confondimento che le variabili osservate in istanti temporali diversi possono avere tra di loro e su sé stesse.

6. ANALISI LONGITUDINALE

L'analisi di mediazione in uno scenario che permette studi longitudinali come quello reso disponibile grazie a SHARE risulterebbe incompleta senza prendere in considerazione la possibile influenza del tempo sulle variabili. In questo modo è possibile valutare sia l'evoluzione dei soggetti e del fenomeno in analisi, ma anche l'influenza che hanno le variabili su loro stesse e sulle altre una volta considerate con *lag* temporali differenti.

Decidere di ampliare l'analisi comprendendo l'effetto temporale e valutando il possibile effetto di confondimento legato alla rilevazione della stessa variabile in istanti temporali diversi, permette di avere una visione più approfondita del fenomeno e delle relazioni tra stato socio-economico, benessere e salute fisica.

Inoltre, come detto alla fine del precedente capitolo, la correzione per tutti i possibili confondenti è fondamentale per interpretare causalmente le stime degli effetti diretti e indiretti. Ampliando l'analisi in un'ottica temporale, è possibile definire con maggiore certezza i nessi causali tra le variabili e, correggendo dall'effetto del confondimento mediante l'utilizzo di un sistema di pesi, è possibile stabilire con maggiore sicurezza gli effetti causali relativi alla relazione tra SES, benessere e salute fisica.

Per questa parte dell'analisi vengono utilizzate le osservazioni dei 6707 soggetti che hanno partecipato a tutte le wave, come fatto per una parte delle analisi trasversali.

La distribuzione delle variabili relative allo stato socio-economico, alla salute fisica e al benessere dei soggetti rimangono quelle riportate nella *Tabella 5.2*. La standardizzazione di queste variabili risulta ancora più utile nelle analisi seguenti poiché permetterà di leggere i risultati ed interpretare i coefficienti dei modelli finali con una maggiore semplicità nonostante l'inserimento dell'effetto del tempo.

6.1 Modello ipotizzato

Le relazioni e le variabili considerate rimangono le stesse definite e descritte nel Capitolo 5. La differenza principale è legata all'inserimento, tra i possibili confondenti, delle variabili relative all'attività fisica svolta dai soggetti. Queste variabili erano state precedentemente escluse poiché non si avevano certezze in merito all'ordinamento temporale quando veniva esaminata la loro relazione con la salute fisica. Inserendo il tempo nelle analisi, tuttavia, può essere eliminato il problema della *reverse causality* e si può considerare solamente l'effetto che l'attività fisica svolta durante le wave precedenti può avere sulle wave successive.

Le due variabili relative all'attività fisica fanno riferimento all'attività fisica vigorosa oppure che richieda un moderato livello di energia. Sono entrambe variabili ordinali con quattro livelli che indicano la frequenza dell'attività fisica svolta: più di una volta alla settimana, una volta alla settimana, da una a tre volte al mese oppure quasi mai o mai.

Le ulteriori decisioni prese prima di incominciare l'analisi riguardano la forma funzionale e la scelta dei modelli che permettono di prendere in considerazione la dipendenza temporale.

Come già mostrato nel Capitolo 4 si può procedere in modi differenti. Invece di seguire il *g-computational approach* che comporterebbe la definizione di molti modelli parametrici per il calcolo delle probabilità che fanno parte della *mediational g-formula* (Formula 4.60), si preferisce fare ricorso all'applicazione più semplice e immediata di due modelli marginali strutturali che permettono la correzione dal confondimento grazie all'utilizzo di un sistema di pesi.

Anche in questo contesto il rispetto delle assunzioni è fondamentale, ma è facilitato dall'introduzione del tempo che molto spesso aiuta a definire meglio le relazioni causali di interesse.

L'importanza dell'inserimento del tempo è legata anche alla natura delle variabili, delle relazioni e del campione a cui si sta facendo riferimento.

Prima di tutto si tratta di un campione analizzato per un periodo di 10 anni, che quindi è portato ad evidenziare cambiamenti a livello individuale dovuti all'invecchiamento che possono avere forti ripercussioni sullo stato di salute e sulle condizioni legate al benessere degli individui in esame, anche in relazione ad eventi traumatici o perdite di familiari che si possono sperimentare. La ricchezza potrebbe subire dei cambiamenti nel corso del decennio trascorso tra il 2004 e il 2015. Come già detto nel Capitolo 5, questi cambiamenti potrebbero essere legati ad aspetti economici come la crisi che nel 2008 ha colpito l'Europa, ma anche individuali, collegati ad avanzamenti di carriera oppure ad eredità percepite e altri eventi che potrebbero aver agito e contribuito ad una condizione finale diversa da quella iniziale.

Ignorare l'evoluzione del campione può portare a risultati incompleti e non completamente corretti.

È già stato osservato come l'esclusione delle covariate avrebbe potuto portare a risultati fuorvianti. Adesso è necessario svolgere un ulteriore passo avanti per cercare di capire fino a che punto l'esclusione del tempo può influenzare i risultati ottenuti a livello trasversale.

È importante conoscere l'ordine temporale per poter definire le strutture e le forme dei modelli. La correzione dell'effetto del confondimento a livello temporale risulta essere efficace solamente se sono ben definite le relazioni tra i confondenti e le variabili relative allo stato socio-economico, al benessere e alla salute fisica, in modo da permettere il calcolo dei pesi nel modo migliore possibile.

Quando il metodo è stato mostrato dal punto di vista teorico nel Capitolo 4, è stato sottolineato come possono essere presi in considerazione diversi ordinamenti temporali e si può correggere l'effetto finale semplicemente modificando il condizionamento delle probabilità quando vengono calcolati i pesi.

Avendo a disposizione cinque wave di osservazione, è stata utilizzata la prima per ottenere le variabili baseline, le tre wave centrali (wave 2, 4 e 5) per la rilevazione di esposizione, mediatore e confondenti tempo-dipendenti, e la

wave finale solamente per la risposta, cioè il benessere psicologico degli individui. Il benessere nelle tre wave centrali è stato considerato come un confondente tempo-dipendente nel calcolo dei pesi.

Nella scelta dei pesi da utilizzare si è deciso di calcolare quelli stabilizzati (Formule 4.70 e 4.71), poiché mediatore ed esposizione sono variabili continue.

Vengono quindi richiamate di seguito le formule da utilizzare nel calcolo dei pesi stabilizzati per il modello per l'outcome, il benessere, e per il modello per il mediatore, la salute fisica, al tempo t , cioè

$$\frac{\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1)\}\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1)\}}{\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), \bar{L}'(t), C\}\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), L'(t), C\}}$$

(5.1)

e

$$\frac{\hat{P}\{A(t)|\bar{A}(t-1)\}}{\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), L'(t), C\}}$$

(5.2)

Il condizionamento è cambiato rispetto a quanto mostrato nel Capitolo 4 poiché è possibile distinguere più gruppi di covariate a seconda dell'influenza temporale che hanno tra di loro.

L'aspetto chiave è la necessità di condizionarsi solamente al passato e per questo motivo le relazioni vengono riassunte nelle *Tablelle 6.1* e *6.2*, in modo da garantire una maggiore chiarezza alle formule appena specificate.

Le variabili baseline, cioè che si mantengono costanti durante tutto il periodo di osservazione sono quelle relative al genere e al livello di istruzione degli individui. Inoltre, per ogni istante t ($t = 2, 4, 5$), sono state calcolate le probabilità che compongono i pesi, condizionandosi sempre alla storia passata di tutte le covariate tempo-dipendenti ($\bar{L}(t-1)$). Prendendo in considerazione l'attività fisica degli individui, ad esempio, quando viene calcolata la probabilità relativa alla salute fisica nella quarta wave, ci si condiziona all'attività fisica svolta nelle wave 1 e 2; se si considera la probabilità relativa

alla salute fisica nella quinta wave, ci si condiziona all'attività fisica svolta nelle wave 1, 2 e 4; e in maniera analoga per tutti gli altri confondenti e per tutte le probabilità a denominatore dei pesi.

Oltre a condizionarsi alla storia passata dei confondenti, riflettendo sulla natura delle variabili e delle relazioni in cui sono coinvolte, è possibile osservare come alcuni di essi possano influenzare l'esposizione e il mediatore anche a livello contemporaneo.

Considerando la salute fisica (*Tabella 6.1*), è possibile ipotizzare che cambiamenti relativi al paese di residenza, allo stato civile, all'età, al BMI e allo stato socio-economico possano essere fonte di stress che può avere ripercussioni sulla salute dei soggetti anche a livello contemporaneo e non solamente relativamente agli anni precedenti.

Allo stesso modo, considerando lo stato socio-economico degli individui (*Tabella 6.2*), è possibile ipotizzare che il cambiamento di residenza possa essere legato a spese che influenzano la ricchezza degli individui; all'età possono essere legati aumenti di stipendio e quindi anche questa può influire sulle disponibilità economiche dei soggetti; e infine la ricchezza degli individui, poiché riguarda la coppia, potrebbe modificarsi a seguito di un divorzio o di una separazione.

Covariate baseline (C)	Dipendenza al tempo t ($L'(t)$)	Dipendenza al tempo $t - 1$ ($\bar{L}(t - 1)$)
Genere	Paese	Situazione abitativa
Livello di istruzione	Stato civile	Attività fisica
	Età	Benessere
	BMI	Paese
	SES	Stato civile
		Età
		BMI
		SES

Tabella 6.1 – Relazioni temporali nei modelli per il calcolo dei pesi specificati per il mediatore al tempo t

Covariate baseline (C)	Dipendenza al tempo t ($L'(t)$)	Dipendenza al tempo $t - 1$ ($\bar{L}(t - 1)$)
Genere	Paese	Salute fisica
Livello di istruzione	Stato civile	Benessere
	Età	BMI
		Attività fisica
		Paese
		Stato civile
		Età

Tabella 6.2 – Relazioni temporali nei modelli per il calcolo dei pesi specificati per l'esposizione al tempo t

Nella scelta dei modelli da applicare dopo aver ottenuto i pesi si è deciso di optare per una specificazione meno semplice della precedente.

È stato ipotizzato un modello marginale strutturale lineare che prendesse in considerazione gli effetti cumulati nel tempo dell'esposizione, cioè lo stato socio-economico, e del mediatore, cioè la salute fisica, nel modello per il benessere degli individui nella sesta wave, ottenendo

$$E(Y_{\bar{am}}) = \theta_0 + \theta_1 cum(\bar{a}) + \theta_2 cum(\bar{m})$$

Per il modello relativo al mediatore è stato utilizzato sempre un MSM lineare ma che considerasse l'effetto medio nel tempo dello stato socio-economico come

$$E(M_{\bar{a}}) = \beta_0 + \beta_1 avg(\bar{a}(t))$$

La specificazione dei modelli appare particolarmente adatta al contesto applicativo a cui si sta facendo riferimento, poiché permette di valutare l'effetto complessivo della salute e della ricchezza considerandoli nella loro interezza, invece di valutare l'influenza di ogni singolo istante temporale separatamente.

Partendo dall'applicazione generale mostrata nel Paragrafo 4.5.2, è possibile affrontare in maniera abbastanza semplice il processo che ha portato alle stime dei modelli.

I cinque periodi di osservazione considerati sono quelli relativi alle wave 1, 2, 4, 5 e 6. Come prima cosa è stato necessario calcolare i pesi che poi sono stati applicati ai modelli; si procede in maniera sequenziale calcolando numeratori e denominatori per poi unirli successivamente e ottenere i pesi finali.

I pesi stabilizzati sono composti da probabilità condizionate dello stato socio-economico e della salute fisica; poiché queste due variabili sono mantenute su scala continua, sono state applicate delle regressioni lineari per calcolare i residui studentizzati dei diversi modelli poi trasformati nelle componenti dei pesi mediante l'applicazione della distribuzione di densità della Normale standard.

La procedura può quindi essere divisa in passaggi successivi:

1. Sono calcolati i pesi per il numeratore della Formula (5.1).

Si costruiranno tre modelli, uno per ciascuno dei tre istanti temporali centrali, per ottenere i residui che poi verranno utilizzati al fine di calcolare le $\hat{P}\{Salute(t)|\overline{SES}(t), \overline{Salute}(t-1)\}$.

Si costruiranno poi altri tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{SES(t)|\overline{SES}(t-1), \overline{Salute}(t-1)\}$.

2. Sono calcolati i pesi per il denominatore della Formula (5.1).

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{Salute(t)|\overline{SES}(t), \overline{Salute}(t-1), \overline{Abitazione}(t-1), \overline{Attività\ fisica}(t-1), \overline{Benessere}(t-1), \overline{Paese}(t), \overline{Stato\ civile}(t), \overline{Età}(t), \overline{BMI}(t), \overline{Genere}, \overline{Istruzione}\}$.

Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{SES(t)|\overline{SES}(t-1), \overline{Salute}(t-1), \overline{Abitazione}(t-1), \overline{Attività\ fisica}(t-1), \overline{Benessere}(t-1), \overline{Paese}(t), \overline{Stato\ civile}(t), \overline{Età}(t), \overline{BMI}(t-1), \overline{Genere}, \overline{Istruzione}\}$.

3. Vengono calcolate le probabilità previste a partire dai residui studentizzati ottenuti negli step (1.) e (2.). Per ogni residuo R viene applicata la distribuzione Normale: $\frac{1}{\sqrt{2\pi}} e^{-R^2/2}$. Sono infine ottenuti i pesi moltiplicando le probabilità previste per il numeratore e dividendole per il prodotto di quelle previste per il denominatore.
4. I pesi ottenuti nello step (3.) sono stati troncati al primo e al 99-esimo percentile, in modo da ottenere delle stime più stabili, e sono stati applicati nella regressione lineare $Benessere(6) \sim cum(SES(t)) + cum(Salute(t))$.
5. Sono calcolati i pesi per il numeratore della Formula (5.2).
Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{Salute(t)|\overline{Salute}(t-1)\}$.
6. Sono calcolati i pesi per il denominatore della Formula (5.2).
Si costruiranno tre modelli per ottenere i residui che poi verranno utilizzati per calcolare le $\hat{P}\{SES(t)|\overline{SES}(t-1), \overline{Salute}(t-1), \overline{Abitazione}(t-1), \overline{Attività\ fisica}(t-1), \overline{Benessere}(t-1), \overline{Paese}(t), \overline{Stato\ civile}(t), \overline{Età}(t), \overline{BMI}(t-1), \overline{Genere}, \overline{Istruzione}\}$.
7. Vengono calcolate le probabilità previste a partire dai residui studentizzati ottenuti negli step (5.) e (6.). Per ogni residuo R viene applicata la distribuzione Normale: $\frac{1}{\sqrt{2\pi}} e^{-R^2/2}$. Sono infine ottenuti i pesi moltiplicando le probabilità previste per il numeratore e dividendole per il prodotto di quelle previste per il denominatore.
8. I pesi ottenuti nello step (7.) sono stati troncati al primo e al 99-esimo percentile, in modo da ottenere delle stime più stabili, e sono stati applicati nella regressione $Salute(6) \sim avg(SES(t))$.
9. Una volta stimati i coefficienti dei modelli è possibile ottenere le stime degli effetti diretti e indiretti come

$$IDE = \theta_1 \{cum(\bar{a}) - cum(\bar{a}^*)\} = T\theta_1 = 3\theta_1$$

e

$$IIE = \sum_{t \leq T} \theta_2 \beta_1(t) \{avg(\bar{a}(t)) - avg(\bar{a}^*(t))\} = T\theta_2\beta_1 = 3\theta_2\beta_1$$

Le formule finali sono ottenute partendo dalle Formule 4.66 e 4.67. Poiché le variabili sono state standardizzate, nel calcolo degli Interventional Direct e Indirect Effects, è stato valutato l'effetto di un cambiamento nello stato socio-economico pari a una deviazione standard attraverso tutti i periodi temporali, così che $a(t) = a^*(t) + 1$ per tutti e tre i periodi di esposizione. Rispetto alla formulazione originale degli effetti, è stato inoltre assunto che i coefficienti nel modello per il mediatore siano costanti nel tempo, cioè $\beta_0(t) = \beta_0$ e $\beta_1(t) = \beta_1$.

I coefficienti dei modelli marginali strutturali per il benessere e per la salute fisica (step (4.) e (8.)) sono stati ottenuti mediante l'applicazione di modelli di regressione lineari, stimati con il metodo dei minimi quadrati pesati.

Gli intervalli di confidenza degli Interventional Direct e Indirect Effects sono stati ottenuti mediante bootstrap con 1000 ripetizioni.

6.1.1 Risultati

Dopo aver stimato i pesi stabilizzati ed aver troncato la loro distribuzione al primo e al 99-esimo percentile per garantire maggior stabilità alle stime, sono stati utilizzati i pesi nelle due regressioni:

- quella del benessere nella sesta wave sull'effetto cumulato del SES e della salute fisica nelle wave 2, 4 e 5;
- quella della salute fisica sull'effetto medio del SES nelle wave 2, 4 e 5.

Di seguito nelle *Table 6.3* e *6.4* sono riportate le stime dei coefficienti dei due modelli con le rispettive significatività e intervalli di confidenza al 95%.

Ad eccezione del termine relativo all'intercetta nel primo modello, tutte le stime dei coefficienti sono risultate significative.

È quindi possibile affermare che esiste un effetto positivo, significativo, dello stato socio-economico degli individui e della loro salute fisica, entrambi cumulati nelle wave centrali (2, 4 e 5), sul loro benessere finale.

Allo stesso modo si osserva un effetto significativo e positivo dello stato socio-economico medio sullo stato di salute degli individui.

	Coefficiente	Stima	Standard Error	t-value	p-value	Intervallo di confidenza (95%)	
Intercetta	θ_0	0.01422	0.01164	1.22	0.2220	-0.00860	0.03704
<i>cum(SES)</i>	θ_1	0.04094	0.00346	11.82	<.0001	0.03415	0.04773
<i>cum(Salute fisica)</i>	θ_2	0.12967	0.00306	42.35	<.0001	0.12367	0.13567

Tabella 6.3 – Stima del modello per il benessere nella sesta wave che valuta l'effetto cumulato dello stato socio-economico e della salute fisica

	Coefficiente	Stima	Standard Error	t-value	p-value	Intervallo di confidenza (95%)	
Intercetta	β_0	0.04825	0.01119	4.31	<.0001	0.02630	0.07019
<i>avg(SES)</i>	β_1	0.15371	0.01311	11.72	<.0001	0.12801	0.17942

Tabella 6.4 – Stima del modello per la salute fisica nella sesta wave che valuta l'effetto medio dello stato socio-economico

Nella *Tabella 6.5* sono riportate delle misure che riassumono le caratteristiche delle variabili da un punto di vista puramente descrittivo che possono aiutare la lettura dei risultati valutando la misura con cui si modificano le variabili.

I parametri dei modelli (*Tabella 6.3* e *6.4*) possono essere interpretati come quelli di un modello di regressione lineare. È quindi possibile affermare che:

- Ad ogni aumento unitario della variabile relativa alla ricchezza cumulata nei tre periodi centrali, segue un aumento pari a 0.04094 della variabile relativa al benessere. Questo corrisponde ad un incremento pari allo 0.71% del benessere causato da un aumento del 2.98% della ricchezza cumulativa.

- Ogni aumento unitario della variabile relativa alla salute fisica cumulata nei tre periodi centrali, segue un incremento pari a 0.12967 della variabile relativa al benessere. Questo corrisponde ad una crescita pari al 2.26% del benessere causata da un aumento del 4.42% della salute fisica complessiva.
- L'incremento unitario della variabile relativa alla ricchezza media nei tre periodi centrali, provoca un aumento pari a 0.15371 della variabile relativa alla salute fisica. Questo corrisponde ad una crescita pari al 2.73% della salute fisica in seguito da un aumento dell'8.93% della ricchezza media.

	Minimo	Massimo	Media	Mediana	Range	Deviazione Standard
<i>cum(SES)</i>	-2.246853	31.355427	0	-0.76261	33.60228	2.48304469
<i>cum(Salute fisica)</i>	-21.006224	1.650729	0	1.015074	22.65695	2.5822789
<i>avg(SES)</i>	-0.748951	10.451809	0	-0.25420	11.20076	0.82768156
Benessere (Wave 6)	-4.221387	1.512608	0	0.184215	5.73399	1
Salute fisica (Wave 6)	-5.043493	0.585125	0	0.362433	5.62862	1

Tabella 6.5 – Misure riassuntive delle variabili coinvolte nei modelli

Successivamente sono stati calcolati gli effetti diretti e indiretti valutando un cambiamento di una deviazione standard nello stato socio-economico degli individui tra i diversi istanti temporali.

Poiché le variabili relative allo stato socio-economico in ogni wave erano state standardizzate, per tre periodi di esposizione, si tratta di considerare $a(t) = a^*(t) + 1$.

L'Interventional Direct Effect rappresenta l'effetto sul benessere dell'ultima wave, di fissare il SES alla traiettoria più alta e la salute fisica al valore che avrebbe ottenuto se fosse estratta casualmente dalla distribuzione della

salute fisica quando il SES segue la traiettoria più bassa, rispetto a fissare il SES alla traiettoria più bassa e la salute fisica al valore che avrebbe ottenuto se fosse estratta casualmente dalla distribuzione della salute fisica quando il SES segue la traiettoria più bassa.

In questo modo l'Interventional Direct Effect può essere ottenuto come

$$\begin{aligned} IDE &= E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)}|c\right) - E\left(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|c}(t)}|c\right) = \theta_1\{cum(\bar{a}) - cum(\bar{a}^*)\} = T\theta_1 = 3\theta_1 \\ &= 3 \cdot 0.04094 = 0.12282 \end{aligned}$$

con intervallo di confidenza al 95% ottenuto mediante bootstrap pari a (0.09136; 0.29569).

L'Interventional Indirect Effect rappresenta l'effetto sul benessere dell'ultima wave, di fissare il SES alla traiettoria più alta e la salute fisica al valore che avrebbe ottenuto se fosse estratta casualmente dalla distribuzione della salute fisica quando il SES segue la traiettoria più alta, rispetto a fissare il SES alla traiettoria più alta e la salute fisica al valore che avrebbe ottenuto se fosse estratta casualmente dalla distribuzione della salute fisica quando il SES segue la traiettoria più bassa.

L'Interventional Indirect Effect può essere ottenuto come

$$\begin{aligned} IIE &= E\left(Y_{\bar{a}\bar{G}_{\bar{a}|c}(t)}|c\right) - E\left(Y_{\bar{a}\bar{G}_{\bar{a}^*|c}(t)}|c\right) \\ &= \sum_{t \leq T} \theta_2\beta_1(t) \left\{avg(\bar{a}(t)) - avg(\bar{a}^*(t))\right\} = T\theta_2\beta_1 = 3\theta_2\beta_1 \\ &= 3 \cdot 0.12967 \cdot 0.15371 = 0.05979 \end{aligned}$$

con intervallo di confidenza al 95% ottenuto mediante bootstrap pari a (0.04086; 0.08938).

Sommando l'effetto diretto e quello indiretto, è possibile ottenere l'effetto totale del SES sul benessere per un cambiamento di una deviazione standard tra le wave, ottenendo

$$TE = IDE + IIE = 0.12282 + 0.05979 = 0.18262$$

con intervallo di confidenza al 95% ottenuto mediante bootstrap pari a (0.14147; 0.37236).

L'effetto totale è l'effetto sul benessere di fissare il SES alla traiettoria più alta e la salute fisica al valore che avrebbe ottenuto se fosse estratta

casualmente dalla distribuzione della salute fisica quando il SES segue la traiettoria più alta, rispetto a fissare il SES alla traiettoria più bassa e la salute fisica al valore che avrebbe ottenuto se fosse estratta casualmente dalla distribuzione della salute fisica quando il SES segue la traiettoria più bassa. Al cambiamento nella storia dell'esposizione pari a una deviazione standard attraverso tutti i periodi temporali, segue un cambiamento nel benessere pari al 3.18%.

Dell'effetto totale dello stato socio-economico sul benessere degli individui, circa il 33%¹⁸ è mediato dall'effetto della salute fisica.

Dalla valutazione degli intervalli di confidenza al 95% ottenuti mediante l'applicazione del bootstrap, è possibile affermare che l'effetto totale, l'Interventional Direct Effect e l'Interventional Indirect Effect risultano essere significativamente diversi da zero.

Tramite l'applicazione dei metodi che permettono di svolgere analisi di mediazione anche in situazioni in cui esposizione e mediatore variano nel tempo, è stato possibile ottenere una conferma del ruolo di mediatore che detiene la salute fisica nella relazione tra stato socio-economico e benessere degli individui.

Nel Capitolo 5 era stato possibile valutare tale relazione sulle singole wave ed erano emersi aspetti interessanti sulla dinamica delle tre variabili.

Adesso è possibile confermare quanto detto precedentemente ed affermare che anche valutando l'influenza cumulata nel tempo dello stato socio-economico e dello stato di salute è possibile identificare un effetto significativo, positivo, di entrambi sul benessere degli individui.

Alla salute fisica può essere attribuito il 33% dell'effetto totale e questo permette di concludere che la sua esclusione avrebbe impedito di valutare nella sua interezza la relazione tra stato socio-economico e benessere.

¹⁸ Poiché l'effetto totale è la somma dell'IDE e dell'IIE, la percentuale di effetto dovuta alla mediazione può essere ottenuta come $100 \cdot \frac{IIE}{TE} = 100 \cdot \frac{0.05979}{0.18262} = 32.74\%$.

CONCLUSIONI

All'inizio di questo studio era stato posto l'obiettivo di analizzare la relazione tra stato socio-economico e salute degli individui nelle società moderne, all'interno delle quali si sta assistendo a una decrescita dei tassi di natalità che, con il prolungarsi dell'aspettativa di vita alla nascita, determinano un aumento nella quota di popolazione anziana.

La salute è un concetto molto ampio e può essere scisso in salute fisica e benessere degli individui, due dimensioni altamente interconnesse. Per questo motivo si è cercato di determinare se nella relazione tra stato socio-economico e benessere, la salute fisica potesse ricoprire il ruolo di mediatore. Per vagliare tale ipotesi, e quindi per valutare in che misura la ricchezza può aiutare a migliorare la condizione di benessere e di salute fisica degli individui, e per determinare l'importanza di garantire un invecchiamento in buona salute dei soggetti, al fine di favorire un alto stato di benessere, sono stati utilizzati i dati provenienti dall'Indagine su Salute, Invecchiamento e Pensioni in Europa (SHARE).

Grazie a questa indagine, è stato possibile analizzare i dati degli individui con più di 50 anni residenti nei nove paesi che hanno partecipato a tutte le rilevazioni sia in ottica trasversale, sia da un punto di vista longitudinale, considerando l'evoluzione del campione in un periodo temporale che va dal 2004 al 2015.

Invece di ricorrere ai metodi classici per svolgere analisi di mediazione come i SEM, si è deciso di ricorrere a modelli di mediazione seguendo l'approccio controfattuale. Tale approccio si differenzia dai metodi "tradizionali" poiché, invece di concentrarsi sulla creazione dei modelli e ragionare su di essi per arrivare ad individuare gli effetti diretti e indiretti, pone delle basi teoriche fornendo delle definizioni degli effetti a priori e prevede la stima dei modelli solamente in un secondo momento. Così facendo è possibile superare numerose difficoltà incontrate dai metodi "tradizionali", permettendone

un'applicazione generale anche in situazioni che necessitano l'utilizzo di modelli differenti da quello lineare oppure l'inclusione di termini di interazione. Inoltre, in letteratura, sono stati proposti metodi per gestire scenari in cui mediatore ed esposizione variano nel tempo, attraverso il ricorso ad un approccio basato su un sistema di pesi e a *marginal structural models* per la stima degli effetti in queste situazioni.

Per svolgere l'analisi di mediazione, anche considerando l'evoluzione temporale del campione, è stato necessario selezionare solamente i soggetti che hanno partecipato a tutte le wave di rilevazione. Per tale motivo, nella prima fase delle analisi, sono state svolte per ogni wave due analisi di mediazione, ognuna applicata una volta sui campioni completi e una sul panel: una senza correggere per il confondimento e una inserendo le covariate nei modelli per correggere dall'effetto dei confondenti.

Agli effetti stimati con i modelli senza covariate non è possibile dare un'interpretazione causale, poiché le assunzioni sottostanti non sono rispettate, tuttavia sono serviti per avere una prima conferma del ruolo di mediatore svolto da parte della salute fisica.

Dopo l'aggiunta delle covariate, è possibile concludere che sicuramente vi è un effetto dello stato socio-economico e della salute fisica sul benessere degli individui, ma che la mediazione ipotizzata inizialmente da parte della condizione fisica dei soggetti non risulta essere sempre confermata dai risultati dell'analisi.

Confrontando livelli di SES simili tra di loro, cioè 3.3% sopra e sotto alla media, emergono delle differenze significative nelle condizioni di benessere sia legate all'effetto diretto dello stato socio-economico, sia all'effetto indiretto della salute fisica. Quando si confrontano invece livelli di SES molto differenti tra di loro, emergono degli effetti significativi diretti sul benessere, ma viene meno la significatività degli effetti indiretti. Questo è dovuto al fatto che il secondo confronto viene svolto tra valori massimi e minimi del SES, mentre la maggior parte dei soggetti si trova in condizioni intermedie.

Si è successivamente deciso di proseguire le analisi prendendo in considerazione il possibile effetto di confondimento che le variabili osservate in istanti temporali diversi possono avere tra di loro e su sé stesse.

Dopo aver ragionato sulle relazioni temporali tra le variabili, tramite l'applicazione dei metodi che permettono di svolgere analisi di mediazione anche in situazioni in cui esposizione e mediatore variano nel tempo, è stato possibile ottenere una conferma del ruolo di mediatore che detiene la salute fisica nella relazione tra stato socio-economico e benessere.

Valutando l'influenza cumulata nel tempo dello stato socio-economico e dello stato di salute è possibile identificare un effetto significativo, positivo, di entrambi sul benessere degli individui.

L'effetto totale stimato indica che al cambiamento nella storia dell'esposizione pari a una deviazione standard attraverso tutti i periodi temporali, segue un cambiamento nel benessere pari al 3.18%.

Alla salute fisica può essere attribuito il 33% dell'effetto totale dello stato socio-economico sul benessere, e questo permette di concludere che la sua esclusione avrebbe impedito di valutare nella sua interezza la relazione tra stato socio-economico e benessere.

Le metodologie presentate sono nuove e non sono ancora stati sviluppati degli approcci che riescono a soddisfare appieno le situazioni in cui esposizione e mediatore variano nel tempo, nonostante comunque siano stati fatti notevoli passi avanti rispetto al passato.

In queste situazioni i *Natural Direct Effect* e *Natural Indirect Effect* spesso non sono identificati e per questo motivo sono stati ottenuti e calcolati dei *randomized interventional analogues* che permettono di alleggerire le assunzioni necessarie alla stima.

Per approfondire l'analisi si potrebbe provare a stratificare per paese di residenza, oppure ad includere un numero maggiore di covariate e di *lag* temporali, in modo da avere una visione ancora più completa della relazione tra le variabili esaminate.

Inoltre, poiché i metodi basati su *inverse probability weighting* sono altamente sensibili a specificazioni scorrette, si potrebbe provare a svolgere la medesima analisi utilizzando un approccio alternativo per costruire i pesi, come quello del *residual balancing* (Zhou e Wodtke, 2018). Tale metodo prevede la modellazione delle medie condizionate dei confondenti causati dall'esposizione, piuttosto che delle distribuzioni condizionate di esposizione e mediatore, e il risultato è un approccio più efficiente e robusto a errori di specificazione rispetto al precedente.

BIBLIOGRAFIA E APPENDICI

BIBLIOGRAFIA

DATASET

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.611

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.611

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 – SHARELIFE. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.611

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.611

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.611

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 6.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.611

LETTERATURA

- Adler, N., Cutler, D., Fielding, J., Galea, S., Glymour, M., Koh, H., & Satcher, D. (2016). Addressing Social Determinants of Health and Health Disparities: A Vital Direction for Health and Health Care. *NAM Perspectives*. Discussion Paper. *National Academy of Medicine*.
- Agresti, A. (2013). *Categorical Data Analysis*. Hoboken: 3rd Edition, John Wiley & Sons Inc.
- Alwin, D., & Hauser, R. (1975). The decomposition of effects in path analysis. *American Sociological Review*(40), 37-47.
- Andersen-Ranberg, K., Petersen, I., Frederiksen, H., Mackenbach, J., & Christensen, K. (2009). Cross-national differences in grip strength among 50+ year-old Europeans: Results from the SHARE study. *European Journal of Aging*, 6(3), 227-236.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Börsch-Supan, A., & Jürges, H. (2005). *Survey of Health, Ageing and Retirement in Europe – Methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., . . . Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology* .
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., . . . Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*.
- Börsch-Supan, A., Brandt, M., K., H., & Schröder, M. (2011). *The individual and the welfare state. Life histories in Europe*. Heidelberg: Springer.

- Börsch-Supan, A., Brandt, M., Litwin, H., & Weber, G. (2013). *Active ageing and solidarity between generations in Europe: First results from SHARE after the economic crisis*. Berlin: De Gruyter.
- Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J., Siegrist, J., & Weber, G. (2005). *Health, ageing and retirement in Europe – First results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Börsch-Supan, A., Kneip, T., Litwin, H., Myck, M., & Weber, G. (2015). *Ageing in Europe - Supporting Policies for an Inclusive Society*. Berlin: De Gruyter.
- Bacigalupe, A., & Escolar-Pujolar, A. (2014). The impact of economic crises on social inequalities in health: what do we know so far? *International Journal for Equity in Health*(13).
- Barbeau, E., Krieger, N., & Soobader, M. (2004). Working class matters: Socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. *American Journal of Public Health*(94), 269-278.
- Baron, R., & and Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*(51), 1173-1182.
- Barr, B., Taylor-Robinson, D., Scott-Samuel, A., McKee, M., & Stuckler, D. (2012). Suicides associated with the 2008–10 economic recession in England: time trend analysis. *British Medical Journal*(345).
- Barrow, L., & Rouse, C. (2005). Does college still pay? *Economist's Voice*(2), 1-8.
- Blustein, D. (2008). The role of work in psychological health and well-being: A conceptual, historical, and public policy perspective. *American Psychologist*(63), 228-240.

- Brey, P. (2012). Well-Being in Philosophy, Psychology, and Economics. In P. Brey, A. Briggie, & E. Spence, *The Good Life in a Technological Age* (p. 15-34). Routledge.
- Bureau, U. C. (2005). *Prevalence and Most Common Causes of Disability Among Adults*. Stati Uniti.
- Carnevale, A., S.J., R., & Cheah, B. (2013). *The college payoff: Education, occupation, and lifetime earnings*. Washington DC: Georgetown University Center on Education and the Workforce.
- Caston, R. J. (1989, September). Dimensions of Occupational Inequality and Duncan's Socioeconomic Index. *Sociological Forum*, 4(3), p. 329-348.
- Catalano, R., & Bellows, B. (2005). Commentary: if economic expansion threatens public health, should epidemiologists recommend recession? *International Journal of Epidemiology*(34), 1212-1213.
- Catalano, R., Goldman-Mellor, S., Saxton, K., Margerison-Zilko, C., Subbaraman, M., LeWinn, K., & Anderson, E. (2011). The health effects of economic decline. *Annual Review of Public Health*(32), 431-450.
- Cohen, H. (2002). Developing media literacy skills to challenge television's portrayal of older women. *Educational Gerontology*, 28, 599-620.
- Cole, S., & Hernán, M. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*(168), 656-664.
- Crisp, R. (2017, Fall). Well-Being. *The Stanford Encyclopedia of Philosophy*.
- De Vogli, R., Marmot, M., & Stuckler, D. (2013). Excess suicides and attempted suicides in Italy attributable to the Great Recession. *Journal of Epidemiology and Community Health*(67), 378-379.
- Diener, E., & Suh, E. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, 40(1-2), 189-216.
- Dodge, R., Daly, A., Huyton, J., & Sanders, L. (2012). The challenge of defining wellbeing. *International Journal of Wellbeing*, 2(3), 222-235.

- Duncan, O. (1961). A Socioeconomic Index for all Occupations. In J. J. Reiss, *Occupations and Social Status*. (p. 109-138). New York: Free Press of Glencoe.
- Edwards, R. (2008). Who is hurt by procyclical mortality? *Social Science & Medicine*(67), 2051-2058.
- Hairi, F., Mackenbach, J., Andersen-Ranberg, K., & Avendano, M. (2009). *Does socio- economic status predict grip strength in older Europeans? Results from SHARE study in non-institutionalised men and women aged 50+*.
- Hirst, S. (2015). Aging in European Societies: Healthy Aging in Europe. *Journal of Women & Aging*, 27(3), 268-269.
- Hume, D. (1748). *An Enquiry concerning Human Understanding*.
- Hyde, M., Wiggins, R., Higgs, P., & Blane, D. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging & mental health*, 7(3), 186-194.
- Hyman, H. (1955). *Survey design and analysis: Principles, cases and procedures*. . Glencoe, IL: Free Press.
- Jagger, C., Matthews, R., Comas-Herrera, A., Wittenberg, R., Pickard, L., Malley, J., . . . CFAS, M. (2009). *Disability in later life: causes, consequences and future prospects*. MAP2030.
- Judd, C., & Kenny, D. (1981). Process analysis: Estimating mediation in evaluation research. *Evaluation Research*(5), 602-619.
- Kaplan, G., Shema, S., & Leite, C. (2008). Socioeconomic determinants of psychological well-being: the role of income, income change, and income sources during the course of 29 years. *Annals of Epidemiology*, 18(7), 531-7.
- Katz, S., Ford, A., Moskowitz, R., Jackson, B., & Jaffe, M. (1963). Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. *Jama*, 185(12), 914-919.
- Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30, 81-93.

- Kendall, M. (1948). *Rank correlation methods*. London: C. Griffin.
- La Torre, G., Sestili, C., Mannocci, A., Sinopoli, A., DePaolis, M., DeFrancesco, S., . . . De Giusti, M. (2018). Association between Work Related Stress and Health Related Quality of Life: The Impact of Socio-Demographic Variables. A Cross Sectional Study in a Region of Central Italy. *International Journal of Environmental Research and Public Health*, 15(1), 159.
- Lahelma, E., Rahkonen, O., & Huuhka, M. (1997). Changes in the patterning of health? The case of Finland 1986–1994. *Social Science & Medicine*(44), 789-799.
- LaMorte, W. W., & Sullivan, L. (2016). Confounding and Effect Measure Modification. *Boston University School of Public Health*.
- Lansley, S. (2012, February 5). Why economic inequality leads to collapse. *The Guardian – International Edition*.
- MacKinnon, D. (2008). *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.
- MacKinnon, D., & Dwyer, J. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*(17), 144-158.
- MacKinnon, D., Warsi, G., & Dwyer, J. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*(30), 41-62.
- Maldonado, F., & Greenland, S. (1987). Simulation Study of Confounder-Selection Strategies. *American Journal of Epidemiology*, 138(11), 923-936.
- Malter, F., & Börsch-Supan, A. (2013). *SHARE Wave 4: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.
- Malter, F., & Börsch-Supan, A. (2015). *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

- Malter, F., & Börsch-Supan, A. (2017). *SHARE Wave 6: Panel innovations and collecting Dried Blood Spots*. Munich: Munich Center for the Economics of Aging (MEA).
- Menzies, P. (2017). Counterfactual Theories of Causation. *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, Edward N. Zalta (ed.).
- Miech, R., & Hauser, R. (2001). Socioeconomic status (SES) and health at midlife; a comparison of educational attainment with occupation-based indicators. *Annals of Epidemiology*(11), 75-85.
- Mueller, C., & Parcel, T. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Dev.*(52), 13-30.
- Nam, C., & Powers, M. (1983). *The socioeconomic approach to status measurement*. Houston, TX: Cap and Gown Press.
- Nam, C., & Terrie, E. (1982). Measurement of socioeconomic status from United States census data. In M. Powers, *Measures of socioeconomic status: current issues*. (p. 29-42). Boulder, CO: Westview Press.
- Nam, C., & Terrie, E. (1986). Comparing the 1980 Nam-Powers and Duncan SEI occupational scores. . *Center for the Study of Population, Florida State University*.
- Newell, S. (2000). Socio-economic status & health inequalities: a review & the implications for health promotion. *Paper presented to Northern Rivers Area Health Service Health Promotion Staff, Lismore, NSW. .*
- Neyman, J. (1923). Sur les applications de la thar des probabilités aux expériences Agaricales: Essay des principe. *Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, translators) in Statistical Science*(5), 463-472.
- Nicholas, S., Huppert, F., McWilliams, B., & Melzer, D. (2003). Health, Wealth and Lifestyles of the Older Population in England: the 2002 English Longitudinal Study of Ageing.
- Nota, L., Rondal, A., & Soresi, S. (2002). *La valutazione delle disabilità*. (Vol. I). Erip. Editrice.

- Oakes, M. J., & Kaufman, J. S. (2017). *Methods in Social Epidemiology*. John Wiley & Sons, Inc.
- OMS. (1946). *Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and e*. New York.
- OMS. (1980). *International classification of impairments, disabilities, and handicaps : a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29.35 of the Twenty-ninth World Health Assembly, May 1976*. Geneva : W.
- OMS. (1995). *Physical status: the use and interpretation of anthropometry*. Report of a WHO Expert Committee. WHO Technical Report Series 854, Geneva: World Health Organization.
- OMS. (2001). *International Classification of Functioning, Disability and Health. FINAL DRAFT, Full Version, 2001*. Classification, Assessment, Surveys and Terminology Team.
- OMS. (2002). *Active ageing – A policy framework. A contribution of the World Health Organization to the Second United Nations World Assembly on Ageing*. Madrid, Spain.
- OMS. (2006). *Constitution of the World Health Organization - Basic Documents, Forty-fifth edition, Supplement, October 2006*.
- OMS. (2009). *Definition of an older or elderly person*.
- OMS. (2013). *Health, health systems and economic crisis in Europe Impact and policy implications. Summary*. European Observatory on Health System, London.
- OMS. (2016). *Disabilità*.
- OMS. (2018, February 5). *Ageing and health*. Tratto da World Health Organization: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>

- Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika*(82), 669-710.
- Pearl, J. (2001). Direct and indirect effects. *Seventeenth Conference on Uncertainty in Artificial Intelligence* (p. 411-420). San Francisco: Morgan Kaufmann.
- Pearl, J. (2011). Principal stratification—A goal or a tool? *International Journal of Biostatistics*, 7(1), Article 20.
- Pearl, J. (2012). The causal mediation formula – a guide to the assessment of pathways and mechanisms. *Prevention Science*(13), 426-436.
- Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research, 2nd edition*. Holt, Rinehard and Winston.
- Preacher, K., Rucker, D., & Hayes, A. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185-227.
- Prince, M., Reischies, F., Beekman, A., Fuhrer, R., Jonker, C., Kivela, S., . . . Copeland, J. (1999). Development of the EURO-D scale – a European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry*, 174(4), 330-338.
- Quetelet, A. (1832). Recherches sur le poids de l'homme aux différents âges. *Nouveaux Memoires de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles*, 1-83.
- Reeves, A., Stuckler, D., McKee, M., Gunnell, D., Chang, S.-S., & Basu, S. (2012). Increase in state suicide rates in the USA during economic recession. *Lancet*(380), 1813-1814.
- Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure period – Application to control of the healthy worker survivor effect. *Mathematical Modelling*(7), 1393-1512.
- Robins, J. (1999). Marginal structural models versus structural nested models as tools for causal inference. In M. Halloran, & D. Berry, *Statistical Models in Epidemiology: The Environment and Clinical Trials*. (Vol. 116, p. 95-134). New York: Springer-Verlag.

- Robins, J., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*(3), 143-155.
- Robins, J., & Hernán, M. (2009). Estimation of the causal effects of time-varying exposures. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs, *Longitudinal Data Analysis*. New York: Chapman and Hall/CRC Press.
- Robins, J., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*(11), 550-560.
- Rose, D., & Pevalin, D. (2003). *A Researcher's Guide to the National Statistics Socio-economic Classification*. London: Sage.
- Rosenbaum, P., & Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*(70), 41-55.
- Rossi, P., Sampson, W., Bose, C., Jasso, G., & Passel, J. (1974, September). Measuring household social standing. *Social Science Research*, 3, 169-190.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*(66), 688-701.
- Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*(6), 34-58.
- SAS Institute, I. (2017). *SAS/STAT® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.
- Schröder, M. (2011). *Retrospective data collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE methodology*. . Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Shankar, J., Ip, E., Khalema, E., Couture, J., Tan, S., Zulla, R., & Lam, G. (2013). Education as a social determinant of health: issues facing indigenous and visible minority students in postsecondary education in Western Canada. *International journal of environmental research and public health*, 10(9), 3908-29.

- Shavers, V. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, 99(9), 1013-23.
- Sobel, M. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhardt, *Sociological methodology* (p. 290-312). San Francisco, CA: Jossey-Bass.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.
- Stephoe, A., Deaton, A., & Stone, A. (2014). Subjective wellbeing, health, and ageing. *Lancet*, 385(9968), 640-648.
- Stewart, A., Prandy, K., & Blackburn, R. (1973). Measuring the class structure. *Nature*, 245(5426), 415-417.
- Stewart, A., Prandy, K., & Blackburn, R. (1980). *Social Stratification and Occupations*. Macmillan.
- Story, L. (2010, August 22). Income Inequality and Financial Crises. . *The New York Times*.
- Susser, M. (1973). *Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology*. New York: Oxford University Press.
- Tamborini, C., Kim, C., & Sakamoto, A. (2015). Education and Lifetime Earnings in the United States. *Demography*, 52(4), 1383-407.
- Toffolutti, V., & Suhrcke, M. (2014). Assessing the short term health impact of the great recession in the European Union: a cross-country panel analysis. *Preventive Medicine*(64), 54-62.
- Treanor, J. (2015, October 13). Half of world's wealth now in hands of 1% of population – report. *The Guardian – International Edition*.
- UNESCO. (2006). *International Standard Classification of Education: ISCED 1997*. Tratto da <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>
- UNESCO. (2012). *International Standard Classification of Education: ISCED 2011*. . Tratto da <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>

- Valeri, L., & VanderWeele, T. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*(18), 137-150.
- VanderWeele, T. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*(20), 18-26.
- VanderWeele, T. (2014). A unification of mediation and interaction: a four-way decomposition. *Epidemiology*(25), 749-761.
- VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford Univ. Pr.
- VanderWeele, T., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, 2(4), 457-468.
- VanderWeele, T., & Tchetgen Tchetgen, E. (2016). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 917-938.
- VanderWeele, T., & Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, 172(12), 1339-1348.
- VanderWeele, T., & Vansteelandt, S. (2013). Mediation analysis with multiple mediators. *Epidemiologic Methods*(2), 95-115.
- VonDemKneesebeck, O., Hyde, M., Higgs, P., Kupfer, A., & Siegrist, J. (2005). Quality of Life and Well-Being. . In A. Börsch-Supan, A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, & G. Weber, *Health, ageing and retirement in Europe – First results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Wadsworth, E., Chaplin, K., & Smith, A. (2010, December 1). The work environment, stress and well-being. *Occupational Medicine*, 60(8), 635-639.

- Whitehead, M., Dahlgren, G., & Gilson, L. (2001). *Developing the policy response to inequities in Health: a global perspective. ill;.-Challenging inequities in health care: from ethics to action*. New York: Oxford University Press.
- Wright, R. (1988). Causation, responsibility, risk, probability, naked statistics, and proof: Pruning the bramble bush by clarifying the concepts. *Iowa Law Review*(73), 1001-1077.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*(20), 557-585.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*(20), 557-585.
- Zhou, X., & Wodtke, G. (2018). *Residual Balancing Weights for Marginal Structural Models: with Application to Analyses of Time-varying Treatments and Causal Mediation*. Harvard University & University of Toronto.

APPENDICE A

Distribuzione delle variabili che compongono la variabile relativa allo stato socio-economico degli individui

Valore della residenza principale	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	5817	26.30	5591	24.98	8644	25.87	10609	24.39	9431	24.20
0 - 50000	1186	5.36	624	2.79	501	1.50	799	1.84	534	1.37
50000 - 100000	2826	12.78	1780	7.95	1715	5.13	2846	6.54	2637	6.77
100000 - 200000	5997	27.11	5127	22.91	7048	21.09	9594	22.06	8958	22.98
200000 - 400000	4802	21.71	6721	30.03	10330	30.91	13289	30.56	11828	30.34
400000 - 600000	1055	4.77	1841	8.23	3313	9.91	4035	9.28	3427	8.79
600000 - 1000000	323	1.46	533	2.38	1375	4.11	1713	3.94	1579	4.05
1000000 - 2000000	106	0.48	137	0.61	349	1.04	457	1.05	491	1.26
2000000 +	7	0.03	27	0.12	143	0.43	149	0.34	94	0.24
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella A.1 – Numero e percentuali di soggetti partecipanti alle wave per valore della residenza principale

Valore di altri possedimenti	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	17444	78.86	17432	77.89	26370	78.91	33997	78.17	30511	78.28
0 - 50000	1300	5.88	958	4.28	983	2.94	1397	3.21	1156	2.97
50000 - 100000	1071	4.84	865	3.86	1119	3.35	1505	3.46	1395	3.58
100000 - 200000	1207	5.46	1347	6.02	1846	5.52	2537	5.83	2235	5.73
200000 - 400000	709	3.21	1203	5.38	1724	5.16	2446	5.62	2144	5.50
400000 - 600000	197	0.89	348	1.55	731	2.19	855	1.97	804	2.06
600000 - 1000000	115	0.52	199	0.89	373	1.12	477	1.10	507	1.30
1000000 - 2000000	53	0.24	21	0.09	201	0.60	223	0.51	173	0.44
2000000 +	23	0.10	8	0.04	71	0.21	54	0.12	54	0.14
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella A.2 – Numero e percentuali di soggetti partecipanti alle wave per valore di altri possedimenti

Conti bancari	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	3487	15.76	1985	8.87	1253	3.75	1809	4.16	2511	6.44
0 - 1000	4740	21.43	4182	18.69	7807	23.36	8939	20.55	6798	17.44
1000 - 2500	2453	11.09	2655	11.86	4150	12.42	6197	14.25	5259	13.49
2500 - 5000	2501	11.31	2660	11.89	3699	11.07	5645	12.98	5194	13.33
5000 - 10000	2619	11.84	2846	12.72	4155	12.43	5657	13.01	4920	12.62
10000 - 20000	2464	11.14	2813	12.57	3627	10.85	4932	11.34	4805	12.33
20000 - 50000	2400	10.85	2892	12.92	4094	12.25	5371	12.35	4945	12.69
50000 - 100000	1041	4.71	1498	6.69	2387	7.14	3054	7.02	2686	6.89
100000 - 150000	294	1.33	479	2.14	1021	3.06	1035	2.38	957	2.46
150000 - 250000	94	0.42	255	1.14	727	2.18	646	1.49	602	1.54
250000 +	26	0.12	116	0.52	498	1.49	206	0.47	302	0.77
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella A.3 – Numero e percentuali di soggetti partecipanti alle wave per valore dei conti bancari

Bond, azioni e fondi comuni	Wave 1		Wave 2		Wave 4		Wave 5		Wave 6	
	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%	Frequenza	%
0	15374	69.51	14533	64.93	23762	71.11	31249	71.85	28144	72.20
0 - 1000	354	1.60	230	1.03	343	1.03	332	0.76	228	0.58
1000 - 2500	475	2.15	477	2.13	562	1.68	667	1.53	434	1.11
2500 - 5000	616	2.78	630	2.81	662	1.98	844	1.94	553	1.42
5000 - 10000	891	4.03	899	4.02	986	2.95	1341	3.08	1014	2.60
10000 - 20000	1303	5.89	1442	6.44	1717	5.14	2109	4.85	1756	4.50
20000 - 50000	1477	6.68	1676	7.49	1976	5.91	2479	5.70	2240	5.75
50000 - 100000	868	3.92	1136	5.08	1427	4.27	1827	4.20	1789	4.59
100000 - 150000	305	1.38	562	2.51	713	2.13	989	2.27	924	2.37
150000 - 250000	247	1.12	437	1.95	573	1.71	754	1.73	784	2.01
250000 +	209	0.94	359	1.60	697	2.09	900	2.07	1113	2.86
Totali	22119	100	22381	100	33418	100	43491	100	38979	100

Tabella A.4 – Numero e percentuali di soggetti partecipanti alle wave per valore di bond, azioni e fondi comuni

APPENDICE B

Analisi di correlazione e analisi fattoriali

Coefficiente di correlazione Tau b di Kendall								
	Limitazioni ADL	Limitazioni mobilità	CASP -12	EURO-D	Valore della residenza principale	Valore di altri possedimenti	Conti bancari	Bond, azioni e fondi comuni
Limitazioni ADL	1.00000	0.41055	-0.19464	-0.21726	-0.06024	-0.08034	-0.05637	-0.07871
Limitazioni mobilità		1.00000	-0.27899	-0.31833	-0.08930	-0.12046	-0.09409	-0.13474
CASP-12		-	1.00000	0.33135	0.09078	0.11184	0.13066	0.16889
EURO-D				1.00000	0.04124	0.06431	0.08779	0.10639
Residenza principale					1.00000	0.13567	0.11270	0.19201
Altri possedimenti						1.00000	0.12746	0.14642
Conti bancari							1.00000	0.29884
Bond, azioni e fondi comuni								1.00000

Tabella B.1 – Coefficienti di correlazione Tau b di Kendall delle variabili che rappresentano il SES, il benessere e la salute fisica degli individui nella prima wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Coefficiente di correlazione Tau b di Kendall

	Limitazioni ADL	Limitazioni mobilità	CASP -12	EURO-D	Valore della residenza principale	Valore di altri possedimenti	Conti bancari	Bond, azioni e fondi comuni
Limitazioni ADL	1.00000	0.43016	-0.20626	-0.21725	-0.06514	-0.09819	-0.06853	-0.09246
Limitazioni mobilità		1.00000	-0.32393	-0.31813	-0.09067	-0.13542	-0.10698	-0.14143
CASP-12		-	1.00000	0.35852	0.09100	0.11370	0.18027	0.20201
EURO-D				1.00000	0.04102	0.06204	0.09524	0.10221
Residenza principale					1.00000	0.15619	0.13817	0.20124
Altri possedimenti						1.00000	0.14733	0.15536
Conti bancari							1.00000	0.28294
Bond, azioni e fondi comuni								1.00000

Tabella B.2 – Coefficienti di correlazione Tau b di Kendall delle variabili che rappresentano il SES, il benessere e la salute fisica degli individui nella seconda wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Coefficiente di correlazione Tau b di Kendall								
	Limitazioni ADL	Limitazioni mobilità	CASP -12	EURO-D	Valore della residenza principale	Valore di altri possedimenti	Conti bancari	Bond, azioni e fondi comuni
Limitazioni ADL	1.00000	0.42922	-0.21895	-0.21446	-0.06828	-0.10536	-0.08215	-0.08126
Limitazioni mobilità		1.00000	-0.32493	-0.31459	-0.09535	-0.13977	-0.11852	-0.12084
CASP-12		-	1.00000	0.37183	0.08929	0.11897	0.16682	0.15165
EURO-D				1.00000	0.04502	0.07715	0.08856	0.08296
Residenza principale					1.00000	0.16921	0.15810	0.21509
Altri possedimenti						1.00000	0.19675	0.20555
Conti bancari							1.00000	0.29997
Bond, azioni e fondi comuni								1.00000

Tabella B.3 – Coefficienti di correlazione Tau b di Kendall delle variabili che rappresentano il SES, il benessere e la salute fisica degli individui nella quarta wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Coefficiente di correlazione Tau b di Kendall

	Limitazioni ADL	Limitazioni mobilità	CASP -12	EURO-D	Valore della residenza principale	Valore di altri possedimenti	Conti bancari	Bond, azioni e fondi comuni
Limitazioni ADL	1.00000	0.44303	-0.22627	-0.22870	-0.05568	-0.09881	-0.08125	-0.08031
Limitazioni mobilità		1.00000	-0.32059	-0.32661	-0.08845	-0.14241	-0.11603	-0.11634
CASP-12		-	1.00000	0.37525	0.10340	0.12460	0.19352	0.15725
EURO-D				1.00000	0.05930	0.08210	0.10607	0.08633
Residenza principale					1.00000	0.17000	0.16107	0.20835
Altri possedimenti						1.00000	0.18187	0.18393
Conti bancari							1.00000	0.25645
Bond, azioni e fondi comuni								1.00000

Tabella B.4 – Coefficienti di correlazione Tau b di Kendall delle variabili che rappresentano il SES, il benessere e la salute fisica degli individui nella quinta wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

Autovalori della matrice di correlazione: Totale = 8 Media = 1				
	Autovalore	Differenza	Proporzione	Cumulativa
1	2.42141753	0.88662550	0.3027	0.3027
2	1.53479203	0.62688088	0.1918	0.4945
3	0.90791114	0.05864199	0.1135	0.6080
4	0.84926915	0.06955597	0.1062	0.7142
5	0.77971318	0.10922290	0.0975	0.8116
6	0.67049028	0.23749476	0.0838	0.8954
7	0.43299552	0.02958434	0.0541	0.9496
8	0.40341118		0.0504	1.0000

Tabella B.5 – Matrice riassuntiva dell'estrazione dei fattori nella sesta wave

	Factor 1	Factor 2
Valore della residenza principale	-0.04895	0.69729
Valore di altri possedimenti	0.02624	0.60842
Conti bancari	-0.02394	0.62414
Bond, azioni e fondi comuni	0.06038	0.65475
EURO-D	-0.73779	0.00311
CASP	-0.72225	0.14177
Limitazioni ADL	0.70697	0.10625
Limitazioni mobilità	0.82251	0.02904

*Tabella B.6 – Matrice dei fattori selezionati ruotati con trasformazione Promax nella
sesta wave (default)*

	Factor1	Factor2
Factor1	1.00000	-0.17927
Factor2	-0.17927	1.00000

*Tabella B.7 – Matrice di correlazione dei fattori selezionati ruotati con trasformazione
Promax nella sesta wave (default)*

APPENDICE C

Correlazioni tra variabili chiave nei due dataset

Coefficiente di correlazione Tau b di Kendall

		Benessere e salute fisica	Benessere e SES	SES e salute fisica
Wave 1	Campione intero	0.32919	0.15166	0.14821
	Campione ridotto	0.30205	0.16901	0.14046
Wave 2	Campione intero	0.34853	0.16328	0.15957
	Campione ridotto	0.31849	0.15873	0.14820
Wave 4	Campione intero	0.35352	0.15935	0.16053
	Campione ridotto	0.33991	0.16315	0.15784
Wave 5	Campione intero	0.35547	0.17325	0.15764
	Campione ridotto	0.36125	0.17677	0.17787
Wave 6	Campione intero	0.33504	0.17371	0.15432
	Campione ridotto	0.35234	0.17595	0.18060

Tabella C.1 – Coefficiente di correlazione Tau b di Kendall delle variabili relative a SES, benessere e salute fisica in tutte le wave (tutti i coefficienti sono risultati significativi e presentano un p-value <.0001)

APPENDICE D

L'effetto di mediazione senza covariate – Modelli per il mediatore e per l'outcome sul campione completo

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0385	0.0079	-0.0540	-0.0230	23.7172	<.0001
SES	0.1174	0.0083	0.1011	0.1337	198.7162	<.0001
Salute fisica	0.4885	0.0093	0.4702	0.5068	2748.8870	<.0001
Salute fisica* SES	0.0278	0.0117	0.00487	0.0507	5.6471	0.0175
Scale	0.8890	0.0055	0.8783	0.8998		

Tabella D.1 – Stime del modello per la risposta nella wave 1

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0627	0.0077	0.0477	0.0777	67.2095	<.0001
SES	0.1077	0.0076	0.0927	0.1227	199.2030	<.0001
Scale	0.8764	0.0054	0.8659	0.8871		

Tabella D.2 – Stime del modello per il mediatore nella wave 1

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0306	0.0062	-0.0428	-0.0184	24.2142	<.0001
SES	0.1153	0.0067	0.1023	0.1284	298.5486	<.0001
Salute fisica	0.5228	0.0077	0.5078	0.5379	4620.8356	<.0001
Salute fisica* SES	0.00272	0.0104	-0.0176	0.0231	0.0690	0.7928
Scale	0.8704	0.0043	0.8620	0.8788		

Tabella D.3 – Stime del modello per la risposta nella wave 2

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0542	0.0061	0.0422	0.0661	79.0136	<.0001
SES	0.1200	0.0060	0.1082	0.1319	395.7806	<.0001
Scale	0.8777	0.0043	0.8693	0.8861		

Tabella D.4 – Stime del modello per il mediatore nella wave 2

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0238	0.0050	-0.0336	-0.0139	22.3043	<.0001
SES	0.1124	0.0055	0.1017	0.1232	419.2453	<.0001
Salute fisica	0.5054	0.0059	0.4938	0.5170	7262.7987	<.0001
Salute fisica* SES	-0.0296	0.0084	-0.0461	-0.0131	12.4163	0.0004
Scale	0.8758	0.0035	0.8690	0.8827		

Tabella D.5 – Stime del modello per la risposta nella wave 4

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0502	0.0050	0.0403	0.0600	99.9589	<.0001
SES	0.0988	0.0049	0.0891	0.1085	398.7880	<.0001
Scale	0.8897	0.0035	0.8828	0.8967		

Tabella D.6 – Stime del modello per il mediatore nella wave 4

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0241	0.0043	-0.0327	-0.0156	30.8456	<.0001
SES	0.1309	0.0047	0.1217	0.1401	782.4706	<.0001
Salute fisica	0.5159	0.0051	0.5059	0.5259	10243.4445	<.0001
Salute fisica* SES	-0.0392	0.0070	-0.0529	-0.0255	31.4371	<.0001
Scale	0.8672	0.0030	0.8613	0.8731		

Tabella D.7 – Stime del modello per la risposta nella wave 5

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0519	0.0044	0.0433	0.0604	141.2647	<.0001
SES	0.1015	0.0044	0.0930	0.1101	539.2561	<.0001
Scale	0.8872	0.0031	0.8811	0.8932		

Tabella D.8 – Stime del modello per il mediatore nella wave 5

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0440	0.0047	-0.0533	-0.0347	86.1610	<.0001
SES	0.1336	0.0051	0.1237	0.1436	694.7812	<.0001
Salute fisica	0.5196	0.0060	0.5079	0.5314	7484.9784	<.0001
Salute fisica* SES	-0.00690	0.0077	-0.0219	0.00814	0.8096	0.3682
Scale	0.8846	0.0033	0.8782	0.8910		

Tabella D.9 – Stime del modello per la risposta nella wave 6

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0812	0.0043	0.0728	0.0897	358.1911	<.0001
SES	0.0986	0.0042	0.0903	0.1069	538.4528	<.0001
Scale	0.8207	0.0030	0.8148	0.8267		

Tabella D.10 – Stime del modello per il mediatore nella wave 6

APPENDICE E

L'effetto di mediazione senza covariate – Modelli per il mediatore e per l'outcome sul campione riferito al panel

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0188	0.0135	-0.0454	0.00772	1.9319	0.1645
SES	0.1446	0.0135	0.1182	0.1711	114.8193	<.0001
Salute fisica	0.4255	0.0148	0.3964	0.4546	821.4670	<.0001
Salute fisica* SES	0.0231	0.0163	-0.00883	0.0551	2.0119	0.1561
Scale	0.8942	0.0095	0.8758	0.9129		

Tabella E.1 – Stime del modello per la risposta nella wave 1

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0220	0.0142	-0.00582	0.0498	2.4017	0.1212
SES	0.1145	0.0136	0.0878	0.1411	70.7596	<.0001
Scale	0.9466	0.0100	0.9271	0.9664		

Tabella E.2 – Stime del modello per il mediatore nella wave 1

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0104	0.0114	-0.0327	0.0120	0.8251	0.3637
SES	0.1139	0.0117	0.0909	0.1369	94.1769	<.0001
Salute fisica	0.4324	0.0129	0.4071	0.4578	1115.8651	<.0001
Salute fisica* SES	0.0232	0.0164	-0.00887	0.0553	2.0124	0.1560
Scale	0.8960	0.0079	0.8807	0.9117		

Tabella E.3 – Stime del modello per la risposta nella wave 2

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0134	0.0120	-0.0101	0.0370	1.2504	0.2635
SES	0.1254	0.0118	0.1023	0.1486	112.4505	<.0001
Scale	0.9600	0.0085	0.9435	0.9768		

Tabella E.4 – Stime del modello per il mediatore nella wave 2

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.00469	0.0112	-0.0266	0.0172	0.1760	0.6749
SES	0.1388	0.0119	0.1154	0.1621	135.5549	<.0001
Salute fisica	0.4436	0.0125	0.4192	0.4680	1267.3998	<.0001
Salute fisica* SES	-0.0754	0.0169	-0.1085	-0.0423	19.9856	<.0001
Scale	0.8820	0.0078	0.8669	0.8974		

Tabella E.5 – Stime del modello per la risposta nella wave 4

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0276	0.0117	0.00457	0.0505	5.5221	0.0188
SES	0.1147	0.0117	0.0919	0.1376	96.7148	<.0001
Scale	0.9409	0.0083	0.9248	0.9573		

Tabella E.6 – Stime del modello per il mediatore nella wave 4

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0246	0.0110	-0.0460	-0.00311	5.0342	0.0249
SES	0.1250	0.0116	0.1023	0.1478	116.2839	<.0001
Salute fisica	0.5283	0.0128	0.5032	0.5534	1702.5103	<.0001
Salute fisica* SES	-0.00592	0.0158	-0.0369	0.0251	0.1401	0.7082
Scale	0.8617	0.0076	0.8469	0.8767		

Tabella E.7 – Stime del modello per la risposta nella wave 5

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0462	0.0111	0.0245	0.0680	17.3485	<.0001
SES	0.1238	0.0113	0.1017	0.1459	120.1359	<.0001
Scale	0.8907	0.0078	0.8754	0.9062		

Tabella E.8 – Stime del modello per il mediatore nella wave 5

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	-0.0421	0.0114	-0.0645	-0.0197	13.5740	0.0002
SES	0.1542	0.0133	0.1281	0.1803	134.3058	<.0001
Salute fisica	0.5362	0.0149	0.5070	0.5654	1294.3266	<.0001
Salute fisica* SES	-0.0469	0.0208	-0.0875	-0.00617	5.0956	0.0240
Scale	0.8677	0.0077	0.8527	0.8830		

Tabella E.9 – Stime del modello per la risposta nella wave 6

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta	0.0835	0.0103	0.0634	0.1037	66.1714	<.0001
SES	0.1269	0.0101	0.1071	0.1467	157.6655	<.0001
Scale	0.8159	0.0073	0.8018	0.8303		

Tabella E.10 – Stime del modello per il mediatore nella wave 6

APPENDICE F

L'effetto di mediazione con covariate – Modelli per il mediatore e per l'outcome sul campione completo

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		-0.0592	0.0662	-0.1889	0.0706	0.7985	0.3715
Genere	Femmina	-0.1431	0.0156	-0.1738	-0.1124	83.6369	<.0001
	Maschio	0
Livello di istruzione	Alto	0.1984	0.0235	0.1525	0.2444	71.6033	<.0001
	Medio	0.0943	0.0197	0.0558	0.1329	23.0090	<.0001
	Basso	0
Età	50 - 59	0.0852	0.0438	-0.00072	0.1711	3.7770	0.0520
	60 - 69	0.1065	0.0444	0.0194	0.1936	5.7448	0.0165
	70 - 79	0.1098	0.0462	0.0193	0.2002	5.6543	0.0174
	80 +	0.0330	0.0528	-0.0704	0.1364	0.3922	0.5311
Stato civile	< 50	0
	Celibe/nubile	0.0287	0.0385	-0.0468	0.1041	0.5552	0.4562
	Separato/a o divorziato/a	-0.0278	0.0352	-0.0967	0.0411	0.6269	0.4285
Paese di residenza	Sposato/a	0.1226	0.0250	0.0736	0.1716	24.0466	<.0001
	Unione registrata	0.2331	0.0610	0.1136	0.3526	14.6170	0.0001
	Vedovo/a	0
	Austria	0.0204	0.0427	-0.0633	0.1041	0.2275	0.6334
Situazione abitativa	Belgio	-0.2072	0.0392	-0.2841	-0.1303	27.8873	<.0001
	Danimarca	0.1209	0.0430	0.0367	0.2051	7.9203	0.0049
	Francia	-0.4543	0.0439	-0.5404	-0.3682	106.8720	<.0001
	Germania	-0.0694	0.0408	-0.1494	0.0106	2.8908	0.0891
	Italia	-0.5858	0.0429	-0.6698	-0.5018	186.8127	<.0001
Situazione abitativa	Spagna	-0.4568	0.0434	-0.5418	-0.3718	111.0155	<.0001
	Svezia	-0.0237	0.0402	-0.1024	0.0550	0.3472	0.5557
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0528	0.0185	-0.0891	-0.0165	8.1176	0.0044

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Proprietario	0
Indice di massa corporea	Normopeso	0.0407	0.0220	-0.00243	0.0838	3.4210	0.0644
	Sottopeso	-0.2154	0.0672	-0.3471	-0.0836	10.2671	0.0014
	Sovrappeso	0.0408	0.0216	-0.00148	0.0831	3.5768	0.0586
	Obeso	0
SES		0.0699	0.0089	0.0524	0.0873	61.3492	<.0001
Salute fisica		0.4349	0.0095	0.4162	0.4535	2081.6770	<.0001
Salute fisica* SES		0.0256	0.0112	0.00373	0.0475	5.2599	0.0218
Scale		0.8463	0.0052	0.8362	0.8566		

Tabella F.1 – Stime del modello per la risposta nella wave 1

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Intercetta	0.1612	0.0633	0.0372	0.2852	6.4909	0.0108
Genere	Femmina	-0.1776	0.0149	-0.2068	-0.1484	142.2245	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2111	0.0224	0.1673	0.2550	89.0854	<.0001
	Medio	0.1661	0.0188	0.1293	0.2029	78.3535	<.0001
	Basso	0
Età	50 - 59	-0.1178	0.0419	-0.2001	-0.0356	7.8922	0.0050
	60 - 69	-0.2096	0.0425	-0.2929	-0.1263	24.3367	<.0001
	70 - 79	-0.4455	0.0440	-0.5318	-0.3593	102.4940	<.0001
	80 +	-0.9921	0.0497	-1.0896	-0.8946	397.7574	<.0001
	< 50	0
Stato civile	Celibe/nubile	0.0337	0.0368	-0.0385	0.1059	0.8353	0.3608
	Separato/a o divorziato/a	-0.0153	0.0336	-0.0812	0.0506	0.2070	0.6491
	Sposato/a	0.0848	0.0239	0.0380	0.1317	12.5837	0.0004
	Unione registrata	-0.00424	0.0584	-0.1186	0.1101	0.0053	0.9421
Paese di residenza	Vedovo/a	0
	Austria	-0.1416	0.0408	-0.2216	-0.0616	12.0319	0.0005
	Belgio	-0.1475	0.0375	-0.2210	-0.0740	15.4764	<.0001
	Danimarca	-0.0708	0.0411	-0.1513	0.00979	2.9642	0.0851

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Francia	-0.0805	0.0420	-0.1629	0.00191	3.6655	0.0555
	Germania	-0.2084	0.0390	-0.2848	-0.1319	28.5352	<.0001
	Italia	-0.1213	0.0410	-0.2016	-0.0410	8.7628	0.0031
	Spagna	-0.2526	0.0414	-0.3338	-0.1715	37.2229	<.0001
	Svezia	-0.0446	0.0384	-0.1199	0.0306	1.3518	0.2450
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0611	0.0177	-0.0957	-0.0264	11.9122	0.0006
	Proprietario	0
Indice di massa corporea	Normopeso	0.3141	0.0209	0.2732	0.3550	226.2525	<.0001
	Sottopeso	-0.0406	0.0643	-0.1667	0.0855	0.3989	0.5276
	Sovrappeso	0.2634	0.0205	0.2232	0.3036	164.5830	<.0001
	Obeso	0
SES		0.0393	0.0081	0.0234	0.0551	23.5173	<.0001
Scale		0.8100	0.0050	0.8003	0.8199		

Tabella F.2 – Stime del modello per il mediatore nella wave 1

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Intercetta	0.1884	0.0542	0.0821	0.2946	12.0821	0.0005
Genere	Femmina	-0.1752	0.0121	-0.1988	-0.1515	210.1948	<.0001
	Maschio	0
Livello di istruzione	Alto	0.1592	0.0183	0.1233	0.1950	75.7042	<.0001
	Medio	0.1114	0.0153	0.0815	0.1413	53.2579	<.0001
	Basso	0
Età	50 - 59	-0.0553	0.0408	-0.1352	0.0246	1.8417	0.1748
	60 - 69	0.0264	0.0409	-0.0538	0.1066	0.4164	0.5187
	70 - 79	0.00282	0.0420	-0.0794	0.0850	0.0045	0.9463
	80 +	-0.0661	0.0451	-0.1545	0.0222	2.1520	0.1424
Stato civile	< 50	0
	Celibe/nubile	-0.0118	0.0304	-0.0713	0.0477	0.1517	0.6969
	Separato/a o divorziato/a	-0.0809	0.0272	-0.1342	-0.0276	8.8627	0.0029
	Sposato/a	0.0623	0.0193	0.0245	0.1000	10.4490	0.0012
	Unione registrata	0.0984	0.0472	0.00575	0.1910	4.3335	0.0374

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Vedovo/a	0
	Austria	-0.0454	0.0335	-0.1111	0.0202	1.8394	0.1750
	Belgio	-0.2898	0.0274	-0.3436	-0.2361	111.6177	<.0001
	Danimarca	0.0756	0.0279	0.0209	0.1303	7.3257	0.0068
	Francia	-0.4116	0.0283	-0.4669	-0.3562	212.2046	<.0001
Paese di residenza	Germania	-0.0541	0.0286	-0.1101	0.00185	3.5922	0.0581
	Italia	-0.6647	0.0284	-0.7203	-0.6090	548.3256	<.0001
	Spagna	-0.4487	0.0300	-0.5075	-0.3899	223.9861	<.0001
	Svezia	-0.0539	0.0282	-0.1091	0.00125	3.6693	0.0554
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0769	0.0144	-0.1052	-0.0486	28.3704	<.0001
	Proprietario	0
	Normopeso	0.0336	0.0168	0.000665	0.0664	3.9985	0.0455
Indice di massa corporea	Sottopeso	-0.1063	0.0496	-0.2035	-0.00897	4.5825	0.0323
	Sovrappeso	0.0340	0.0163	0.00207	0.0660	4.3550	0.0369
	Obeso	0
SES		0.0642	0.0071	0.0504	0.0781	82.4503	<.0001
Salute fisica		0.4710	0.0079	0.4554	0.4865	3524.3505	<.0001
Salute fisica* SES		-0.00649	0.0102	-0.0264	0.0134	0.4079	0.5231
Scale		0.8211	0.0040	0.8133	0.8291		

Tabella F.3 – Stime del modello per la risposta nella wave 2

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.1949	0.0523	0.0924	0.2975	13.8831	0.0002
Genere	Femmina	-0.1616	0.0116	-0.1844	-0.1388	193.2990	<.0001
	Maschio	0
	Alto	0.1882	0.0176	0.1536	0.2228	113.7930	<.0001
Livello di istruzione	Medio	0.1443	0.0147	0.1154	0.1731	96.1492	<.0001
	Basso	0
	50 - 59	-0.1152	0.0394	-0.1925	-0.0380	8.5568	0.0034
Età	60 - 69	-0.2057	0.0396	-0.2832	-0.1282	27.0419	<.0001
	70 - 79	-0.3985	0.0405	-0.4778	-0.3191	96.9467	<.0001
	80 +	-0.8953	0.0431	-0.9798	-0.8107	430.6475	<.0001

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	< 50	0
	Celibe/nubile	0.0776	0.0293	0.0201	0.1351	7.0015	0.0081
	Separato/a o divorziato/a	0.0644	0.0263	0.0129	0.1158	5.9996	0.0143
Stato civile	Sposato/a	0.1185	0.0186	0.0821	0.1550	40.6044	<.0001
	Unione registrata	0.0958	0.0457	0.00628	0.1853	4.3993	0.0360
	Vedovo/a	0
	Austria	-0.2247	0.0324	-0.2881	-0.1613	48.2320	<.0001
	Belgio	-0.1828	0.0265	-0.2347	-0.1309	47.6148	<.0001
	Danimarca	-0.0817	0.0270	-0.1346	-0.0287	9.1474	0.0025
	Francia	-0.1355	0.0273	-0.1890	-0.0820	24.6519	<.0001
Paese di residenza	Germania	-0.1850	0.0276	-0.2391	-0.1310	44.9898	<.0001
	Italia	-0.2512	0.0274	-0.3049	-0.1976	84.1675	<.0001
	Spagna	-0.2264	0.0289	-0.2832	-0.1697	61.2458	<.0001
	Svezia	-0.0448	0.0272	-0.0981	0.00861	2.7023	0.1002
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0926	0.0139	-0.1199	-0.0653	44.3345	<.0001
	Proprietario	0
	Normopeso	0.3080	0.0161	0.2765	0.3395	366.6840	<.0001
	Sottopeso	-0.1031	0.0480	-0.1972	-0.00906	4.6167	0.0317
Indice di massa corporea	Sovrappeso	0.2308	0.0157	0.2000	0.2615	216.4549	<.0001
	Obeso	0
SES		0.0380	0.0062	0.0258	0.0502	37.2994	<.0001
Scale		0.7940	0.0039	0.7864	0.8017		

Tabella F.4 – Stime del modello per il mediatore nella wave 2

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.0788	0.0429	-0.00517	0.1629	3.3832	0.0659
Genere	Femmina	-0.1627	0.0099	-0.1822	-0.1432	267.8802	<.0001
	Maschio	0
	Alto	0.2070	0.0153	0.1770	0.2369	183.1782	<.0001
Livello di istruzione	Medio	0.1539	0.0130	0.1285	0.1794	140.6427	<.0001
	Basso	0

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Età	50 - 59	-0.0613	0.0326	-0.1252	0.00258	3.5370	0.0600
	60 - 69	0.0408	0.0327	-0.0232	0.1048	1.5614	0.2115
	70 - 79	0.00924	0.0334	-0.0563	0.0748	0.0763	0.7824
	80 +	-0.0109	0.0358	-0.0811	0.0594	0.0919	0.7617
	< 50	0
Stato civile	Celibe/nubile	0.0219	0.0240	-0.0251	0.0689	0.8319	0.3617
	Separato/a o divorziato/a	-0.0248	0.0213	-0.0665	0.0169	1.3537	0.2446
	Sposato/a	0.1027	0.0159	0.0714	0.1339	41.4859	<.0001
	Unione registrata	0.0910	0.0404	0.0118	0.1701	5.0745	0.0243
Paese di residenza	Vedovo/a	0
	Austria	0.0856	0.0193	0.0477	0.1235	19.5803	<.0001
	Belgio	-0.3275	0.0189	-0.3646	-0.2904	299.4174	<.0001
	Danimarca	0.0963	0.0232	0.0508	0.1418	17.2346	<.0001
	Francia	-0.3053	0.0193	-0.3431	-0.2675	250.2633	<.0001
	Germania	-0.1062	0.0263	-0.1579	-0.0546	16.2592	<.0001
	Italia	-0.5908	0.0214	-0.6328	-0.5489	763.6264	<.0001
	Spagna	-0.3642	0.0220	-0.4073	-0.3210	273.7535	<.0001
	Svezia	-0.0434	0.0245	-0.0914	0.00473	3.1225	0.0772
	Svizzera	0
Situazione abitativa	Non proprietario	-0.1248	0.0117	-0.1478	-0.1018	113.0310	<.0001
	Proprietario	0
Indice di massa corporea	Normopeso	0.0274	0.0136	0.000686	0.0542	4.0411	0.0444
	Sottopeso	-0.1518	0.0404	-0.2310	-0.0726	14.1010	0.0002
	Sovrappeso	0.0417	0.0133	0.0156	0.0679	9.7927	0.0018
	Obeso	0
SES		0.0617	0.0058	0.0503	0.0732	112.0529	<.0001
Salute fisica		0.4536	0.0062	0.4415	0.4658	5371.0103	<.0001
Salute fisica* SES		-0.0323	0.0081	-0.0482	-0.0163	15.6956	<.0001
Scale		0.8315	0.0033	0.8250	0.8380		

Tabella F.5 – Stime del modello per la risposta nella wave 4

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.1553	0.0413	0.0744	0.2362	14.1702	0.0002
Genere	Femmina	-0.1572	0.0095	-0.1759	-0.1385	271.4097	<.0001
	Maschio	0
	Alto	0.2067	0.0147	0.1779	0.2355	197.9967	<.0001
Livello di istruzione	Medio	0.1506	0.0125	0.1261	0.1750	145.7423	<.0001
	Basso	0
	50 - 59	-0.1019	0.0314	-0.1634	-0.0403	10.5133	0.0012
Età	60 - 69	-0.1554	0.0315	-0.2171	-0.0937	24.3819	<.0001
	70 - 79	-0.3469	0.0322	-0.4100	-0.2839	116.2061	<.0001
	80 +	-0.8693	0.0342	-0.9363	-0.8023	646.6490	<.0001
Stato civile	< 50	0
	Celibe/nubile	0.0629	0.0231	0.0177	0.1082	7.4229	0.0064
	Separato/a o divorziato/a	0.0436	0.0205	0.00341	0.0838	4.5214	0.0335
Paese di residenza	Sposato/a	0.1014	0.0153	0.0713	0.1315	43.6298	<.0001
	Unione registrata	0.0222	0.0389	-0.0541	0.0985	0.3240	0.5692
	Vedovo/a	0
Situazione abitativa	Austria	-0.1476	0.0186	-0.1842	-0.1111	62.7783	<.0001
	Belgio	-0.2534	0.0182	-0.2890	-0.2177	194.4004	<.0001
	Danimarca	-0.0419	0.0224	-0.0857	0.00196	3.5058	0.0612
	Francia	-0.1405	0.0186	-0.1769	-0.1041	57.1528	<.0001
	Germania	-0.2673	0.0254	-0.3169	-0.2176	111.0955	<.0001
	Italia	-0.2278	0.0206	-0.2681	-0.1875	122.6173	<.0001
	Spagna	-0.3124	0.0211	-0.3539	-0.2710	218.3745	<.0001
	Svezia	-0.0528	0.0236	-0.0991	-0.00642	4.9801	0.0256
	Svizzera	0
	Non proprietario	-0.1025	0.0113	-0.1247	-0.0804	82.4141	<.0001
Indice di massa corporea	Proprietario	0
	Normopeso	0.3354	0.0130	0.3099	0.3609	664.3517	<.0001
	Sottopeso	0.2032	0.0390	0.1268	0.2795	27.2074	<.0001
SES	Sovrappeso	0.2699	0.0128	0.2448	0.2949	447.0735	<.0001
	Obeso	0
SES		0.0237	0.0051	0.0137	0.0337	21.6168	<.0001

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Scale	0.8016	0.0032	0.7953	0.8079		

Tabella F.6 – Stime del modello per il mediatore nella wave 4

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Intercetta		-0.0428	0.0409	-0.1231	0.0374	1.0938	0.2956
Genere	Femmina	-0.1517	0.0085	-0.1683	-0.1351	321.2433	<.0001
	Maschio	0
	Alto	0.2172	0.0131	0.1915	0.2429	274.0885	<.0001
Livello di istruzione	Medio	0.1603	0.0115	0.1379	0.1828	195.4268	<.0001
	Basso	0
	50 - 59	0.0193	0.0319	-0.0431	0.0818	0.3683	0.5440
Età	60 - 69	0.1112	0.0318	0.0488	0.1735	12.2175	0.0005
	70 - 79	0.0711	0.0323	0.00774	0.1344	4.8369	0.0279
	80 +	0.0289	0.0340	-0.0378	0.0956	0.7210	0.3958
	< 50	0
Stato civile	Celibe/nubile	-0.00921	0.0209	-0.0501	0.0317	0.1946	0.6591
	Separato/a o divorziato/a	-0.0782	0.0185	-0.1144	-0.0419	17.8787	<.0001
	Sposato/a	0.0787	0.0138	0.0517	0.1057	32.6622	<.0001
Paese di residenza	Unione registrata	0.0315	0.0358	-0.0386	0.1016	0.7769	0.3781
	Vedovo/a	0
	Austria	0.1619	0.0209	0.1209	0.2030	59.7840	<.0001
	Belgio	-0.2462	0.0196	-0.2846	-0.2077	157.4999	<.0001
	Danimarca	0.1621	0.0207	0.1214	0.2027	61.0378	<.0001
	Francia	-0.2425	0.0209	-0.2835	-0.2016	134.7348	<.0001
	Germania	-0.1044	0.0200	-0.1435	-0.0652	27.3216	<.0001
Situazione abitativa	Italia	-0.6607	0.0211	-0.7019	-0.6194	983.4340	<.0001
	Spagna	-0.3327	0.0205	-0.3728	-0.2925	263.5244	<.0001
	Svezia	-0.0202	0.0202	-0.0599	0.0194	1.0013	0.3170
	Svizzera	0
	Non proprietario	-0.1285	0.0102	-0.1484	-0.1086	159.6644	<.0001
Situazione abitativa	Proprietario	0
	Normopeso	0.0687	0.0116	0.0460	0.0914	35.2461	<.0001

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Indice di massa corporea	Sottopeso	-0.0863	0.0361	-0.1570	-0.0155	5.7131	0.0168
	Sovrappeso	0.0679	0.0114	0.0456	0.0901	35.6350	<.0001
	Obeso	0
SES		0.0722	0.0050	0.0625	0.0820	210.0947	<.0001
Salute fisica		0.4641	0.0053	0.4538	0.4745	7741.3815	<.0001
Salute fisica* SES		-0.0443	0.0067	-0.0574	-0.0311	43.4538	<.0001
Scale		0.8158	0.0028	0.8103	0.8214		

Tabella F.7 – Stime del modello per la risposta nella wave 5

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.1661	0.0401	0.0876	0.2447	17.1697	<.0001
Genere	Femmina	-0.1503	0.0083	-0.1665	-0.1341	330.9627	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2211	0.0128	0.1960	0.2462	297.7872	<.0001
	Medio	0.1492	0.0112	0.1272	0.1712	176.9777	<.0001
	Basso	0
	50 - 59	-0.1085	0.0312	-0.1697	-0.0473	12.0793	0.0005
Età	60 - 69	-0.1802	0.0312	-0.2412	-0.1191	33.4454	<.0001
	70 - 79	-0.3641	0.0316	-0.4261	-0.3021	132.4832	<.0001
	80 +	-0.8544	0.0331	-0.9193	-0.7896	666.7065	<.0001
	< 50	0
Stato civile	Celibe/nubile	0.0487	0.0205	0.00859	0.0888	5.6646	0.0173
	Separato/a o divorziato/a	0.0374	0.0181	0.00192	0.0729	4.2675	0.0388
	Sposato/a	0.1084	0.0135	0.0820	0.1349	64.6021	<.0001
Paese di residenza	Unione registrata	0.1087	0.0350	0.0400	0.1774	9.6264	0.0019
	Vedovo/a	0
	Austria	-0.1437	0.0205	-0.1839	-0.1035	49.1207	<.0001
	Belgio	-0.2488	0.0192	-0.2864	-0.2112	168.5335	<.0001
	Danimarca	-0.0858	0.0203	-0.1256	-0.0460	17.8418	<.0001
Paese di residenza	Francia	-0.1524	0.0205	-0.1925	-0.1123	55.5413	<.0001
	Germania	-0.2034	0.0195	-0.2417	-0.1651	108.3460	<.0001

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Italia	-0.2438	0.0206	-0.2842	-0.2035	140.0619	<.0001
	Spagna	-0.1849	0.0201	-0.2242	-0.1456	85.0162	<.0001
	Svezia	-0.0134	0.0198	-0.0522	0.0255	0.4538	0.5005
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0846	0.0099	-0.1041	-0.0651	72.3534	<.0001
	Proprietario	0
Indice di massa corporea	Normopeso	0.3141	0.0112	0.2921	0.3361	782.2000	<.0001
	Sottopeso	0.1169	0.0354	0.0476	0.1862	10.9329	0.0009
	Sovrappeso	0.2446	0.0111	0.2229	0.2663	487.5575	<.0001
	Obeso	0
SES		0.0317	0.0045	0.0228	0.0406	48.6368	<.0001
Scale		0.7996	0.0028	0.7941	0.8051		

Tabella F.8 – Stime del modello per il mediatore nella wave 5

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Intercetta	-0.0376	0.0487	-0.1331	0.0578	0.5971	0.4397
Genere	Femmina	-0.1712	0.0093	-0.1894	-0.1531	342.0317	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2042	0.0144	0.1759	0.2325	200.4306	<.0001
	Medio	0.1600	0.0126	0.1353	0.1847	161.1353	<.0001
	Basso	0
Età	50 - 59	-0.0202	0.0397	-0.0980	0.0577	0.2580	0.6115
	60 - 69	0.0784	0.0395	0.00101	0.1558	3.9427	0.0471
	70 - 79	0.0439	0.0399	-0.0344	0.1222	1.2092	0.2715
	80 +	-0.00790	0.0415	-0.0892	0.0734	0.0363	0.8490
Stato civile	< 50	0
	Celibe/nubile	-0.0416	0.0226	-0.0860	0.00273	3.3830	0.0659
	Separato/a o divorziato/a	-0.0466	0.0197	-0.0852	-0.00802	5.6052	0.0179
Paese di residenza	Sposato/a	0.0917	0.0148	0.0627	0.1206	38.4894	<.0001
	Unione registrata	0.0567	0.0374	-0.0166	0.1301	2.2962	0.1297
	Vedovo/a	0
	Austria	0.1501	0.0241	0.1030	0.1973	38.9388	<.0001

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Belgio	-0.1956	0.0214	-0.2375	-0.1537	83.7528	<.0001
	Danimarca	0.1774	0.0232	0.1320	0.2228	58.6226	<.0001
	Francia	-0.2514	0.0237	-0.2978	-0.2050	112.5617	<.0001
	Germania	-0.0736	0.0229	-0.1185	-0.0287	10.3307	0.0013
	Italia	-0.5657	0.0230	-0.6107	-0.5206	605.5801	<.0001
	Spagna	-0.2900	0.0236	-0.3363	-0.2436	150.4881	<.0001
	Svezia	-0.00168	0.0228	-0.0463	0.0429	0.0054	0.9412
	Svizzera	0
Situazione abitativa	Non proprietario	-0.1067	0.0114	-0.1291	-0.0842	86.8554	<.0001
	Proprietario	0
Indice di massa corporea	Normopeso	0.0602	0.0126	0.0355	0.0850	22.7436	<.0001
	Sottopeso	-0.0651	0.0411	-0.1456	0.0155	2.5068	0.1134
	Sovrappeso	0.0511	0.0124	0.0269	0.0754	17.0816	<.0001
	Obeso	0
SES		0.0641	0.0058	0.0527	0.0754	122.5909	<.0001
Salute fisica		0.4816	0.0062	0.4694	0.4938	5988.7201	<.0001
Salute fisica* SES		-0.0107	0.0074	-0.0252	0.00387	2.0695	0.1503
Scale		0.8393	0.0031	0.8332	0.8455		

Tabella F.9 – Stime del modello per la risposta nella wave 6

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.1546	0.0430	0.0702	0.2389	12.8975	0.0003
Genere	Femmina	-0.1497	0.0082	-0.1656	-0.1337	337.0936	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2043	0.0127	0.1793	0.2292	258.2420	<.0001
	Medio	0.1159	0.0111	0.0941	0.1377	108.4951	<.0001
	Basso	0
Età	50 - 59	-0.1192	0.0351	-0.1881	-0.0504	11.5166	0.0007
	60 - 69	-0.1775	0.0349	-0.2459	-0.1090	25.8238	<.0001
	70 - 79	-0.3226	0.0353	-0.3917	-0.2534	83.5418	<.0001
	80 +	-0.7438	0.0365	-0.8153	-0.6723	415.5511	<.0001
	< 50	0
Stato civile	Celibe/nubile	0.0542	0.0200	0.0150	0.0935	7.3373	0.0068

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Separato/a o divorziato/a	0.0467	0.0174	0.0126	0.0809	7.2052	0.0073
	Sposato/a	0.1101	0.0131	0.0845	0.1357	71.1240	<.0001
	Unione registrata	0.1114	0.0331	0.0465	0.1763	11.3266	0.0008
	Vedovo/a	0
Paese di residenza	Austria	-0.1325	0.0213	-0.1742	-0.0909	38.8335	<.0001
	Belgio	-0.2425	0.0189	-0.2794	-0.2055	165.2574	<.0001
	Danimarca	-0.0670	0.0205	-0.1072	-0.0269	10.7006	0.0011
	Francia	-0.1515	0.0209	-0.1925	-0.1105	52.3262	<.0001
	Germania	-0.1565	0.0202	-0.1961	-0.1168	59.7058	<.0001
	Italia	-0.1114	0.0203	-0.1512	-0.0715	30.0338	<.0001
	Spagna	-0.0977	0.0209	-0.1386	-0.0567	21.8341	<.0001
	Svezia	0.00401	0.0201	-0.0355	0.0435	0.0396	0.8423
	Svizzera	0
	Situazione abitativa	Non proprietario	-0.1056	0.0101	-0.1254	-0.0859	109.5556
Proprietario		0
Indice di massa corporea	Normopeso	0.3310	0.0110	0.3094	0.3526	899.6096	<.0001
	Sottopeso	0.0821	0.0364	0.0109	0.1534	5.1008	0.0239
	Sovrappeso	0.2705	0.0108	0.2493	0.2918	622.1543	<.0001
	Obeso	0
SES		0.0283	0.0048	0.0190	0.0376	35.4325	<.0001
Scale		0.7424	0.0028	0.7370	0.7478		

Tabella F.10 – Stime del modello per il mediatore nella wave 6

APPENDICE G

L'effetto di mediazione con covariate – Modelli per il mediatore e per l'outcome sul campione riferito al panel

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		-0.1578	0.1143	-0.3818	0.0662	1.9068	0.1673
Genere	Femmina	-0.2041	0.0268	-0.2566	-0.1516	58.0669	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2175	0.0400	0.1392	0.2958	29.6277	<.0001
	Medio	0.1445	0.0342	0.0775	0.2114	17.8834	<.0001
	Basso	0
Età	50 - 59	-0.00758	0.0807	-0.1657	0.1506	0.0088	0.9252
	60 - 69	0.0544	0.0815	-0.1053	0.2141	0.4459	0.5043
	70 - 79	0.1225	0.0854	-0.0450	0.2899	2.0547	0.1517
	80 +	0.0372	0.1177	-0.1936	0.2680	0.0999	0.7519
	< 50	0
Stato civile	Celibe/nubile	0.1290	0.0687	-0.00572	0.2637	3.5223	0.0605
	Separato/a o divorziato/a	0.0964	0.0609	-0.0229	0.2158	2.5066	0.1134
	Sposato/a	0.1907	0.0443	0.1038	0.2776	18.4878	<.0001
Paese di residenza	Unione registrata	0.3661	0.1069	0.1565	0.5757	11.7218	0.0006
	Vedovo/a	0
	Austria	0.1697	0.0699	0.0326	0.3068	5.8880	0.0152
	Belgio	-0.1591	0.0594	-0.2756	-0.0426	7.1665	0.0074
	Danimarca	0.2033	0.0646	0.0767	0.3300	9.9070	0.0016
Situazione abitativa	Francia	-0.3703	0.0735	-0.5143	-0.2263	25.3937	<.0001
	Germania	0.00805	0.0682	-0.1257	0.1417	0.0139	0.9061
	Italia	-0.5679	0.0658	-0.6968	-0.4390	74.5671	<.0001
	Spagna	-0.3786	0.0689	-0.5137	-0.2436	30.2081	<.0001
	Svezia	0.0596	0.0630	-0.0639	0.1831	0.8945	0.3443
Situazione abitativa	Svizzera	0
	Non proprietario	-0.0234	0.0337	-0.0894	0.0427	0.4802	0.4884
	Proprietario	0

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Indice di massa corporea	Normopeso	0.1063	0.0368	0.0341	0.1784	8.3275	0.0039
	Sottopeso	0.0979	0.1323	-0.1614	0.3571	0.5475	0.4593
	Sovrappeso	0.0549	0.0359	-0.0154	0.1253	2.3411	0.1260
	Obeso	0
SES		0.0927	0.0144	0.0644	0.1210	41.2861	<.0001
Salute fisica		0.3586	0.0147	0.3298	0.3873	597.8352	<.0001
Salute fisica* SES		0.0188	0.0154	-0.0114	0.0489	1.4899	0.2222
Scale		0.8386	0.0089	0.8214	0.8562		

Tabella G.1 – Stime del modello per la risposta nella wave 1

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Intercetta		0.0486	0.1217	-0.1900	0.2872	0.1595	0.6896
Genere	Femmina	-0.2818	0.0282	-0.3372	-0.2265	99.6258	<.0001
	Maschio	0
	Alto	0.2795	0.0424	0.1965	0.3626	43.5015	<.0001
Livello di istruzione	Medio	0.2360	0.0362	0.1650	0.3070	42.4091	<.0001
	Basso	0
	50 - 59	-0.0600	0.0860	-0.2285	0.1085	0.4869	0.4853
Età	60 - 69	-0.1612	0.0868	-0.3314	0.00890	3.4501	0.0632
	70 - 79	-0.2935	0.0910	-0.4718	-0.1152	10.4119	0.0013
	80 +	-0.6820	0.1251	-0.9272	-0.4369	29.7308	<.0001
	< 50	0
	Celibe/nubile	-0.0280	0.0732	-0.1715	0.1155	0.1466	0.7018
Stato civile	Separato/a o divorziato/a	-0.0333	0.0649	-0.1606	0.0939	0.2638	0.6076
	Sposato/a	0.0386	0.0473	-0.0540	0.1312	0.6664	0.4143
	Unione registrata	-0.0554	0.1140	-0.2788	0.1680	0.2363	0.6269
	Vedovo/a	0
Paese di residenza	Austria	-0.0824	0.0745	-0.2284	0.0636	1.2232	0.2687
	Belgio	-0.2052	0.0632	-0.3290	-0.0813	10.5422	0.0012
	Danimarca	-0.0558	0.0688	-0.1907	0.0791	0.6565	0.4178
	Francia	-0.1792	0.0782	-0.3326	-0.0259	5.2474	0.0220
	Germania	-0.2649	0.0726	-0.4072	-0.1226	13.3166	0.0003

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Situazione abitativa	Italia	-0.1460	0.0700	-0.2832	-0.00874	4.3461	0.0371
	Spagna	-0.3473	0.0732	-0.4909	-0.2038	22.4856	<.0001
	Svezia	-0.0117	0.0671	-0.1432	0.1198	0.0304	0.8617
	Svizzera	0
	Non proprietario	-0.1249	0.0358	-0.1951	-0.0547	12.1477	0.0005
	Proprietario	0
Indice di massa corporea	Normopeso	0.3868	0.0388	0.3108	0.4629	99.3695	<.0001
	Sottopeso	0.0240	0.1410	-0.2523	0.3003	0.0289	0.8649
	Sovrappeso	0.2622	0.0380	0.1877	0.3368	47.5239	<.0001
	Obeso	0
SES		0.0340	0.0149	0.00476	0.0632	5.1968	0.0226
Scale		0.8938	0.0095	0.8755	0.9126		

Tabella G.2 – Stime del modello per il mediatore nella wave 1

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Intercetta	0.4883	0.1245	0.2443	0.7324	15.3772	<.0001
Genere	Femmina	-0.2364	0.0225	-0.2804	-0.1924	110.8960	<.0001
	Maschio	0
Livello di istruzione	Alto	0.1819	0.0334	0.1165	0.2473	29.7312	<.0001
	Medio	0.1326	0.0283	0.0772	0.1880	21.9963	<.0001
	Basso	0
Età	50 - 59	-0.3517	0.1058	-0.5591	-0.1443	11.0428	0.0009
	60 - 69	-0.2654	0.1057	-0.4727	-0.0582	6.3012	0.0121
	70 - 79	-0.2458	0.1073	-0.4561	-0.0356	5.2540	0.0219
	80 +	-0.2679	0.1149	-0.4932	-0.0427	5.4370	0.0197
	< 50	0
Stato civile	Celibe/nubile	0.0200	0.0561	-0.0899	0.1299	0.1270	0.7215
	Separato/a o divorziato/a	-0.0108	0.0494	-0.1076	0.0861	0.0473	0.8278
	Sposato/a	0.0916	0.0351	0.0229	0.1603	6.8226	0.0090
	Unione registrata	0.1683	0.0904	-0.00884	0.3455	3.4676	0.0626
	Vedovo/a	0
Paese di residenza	Austria	0.0215	0.0622	-0.1005	0.1434	0.1189	0.7303

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Belgio	-0.2521	0.0502	-0.3506	-0.1536	25.1818	<.0001
	Danimarca	0.0979	0.0556	-0.0111	0.2069	3.1000	0.0783
	Francia	-0.3781	0.0556	-0.4871	-0.2691	46.2477	<.0001
	Germania	-0.0908	0.0579	-0.2042	0.0226	2.4640	0.1165
	Italia	-0.6990	0.0545	-0.8058	-0.5922	164.5436	<.0001
	Spagna	-0.4689	0.0574	-0.5814	-0.3563	66.6927	<.0001
	Svezia	-0.0533	0.0536	-0.1583	0.0517	0.9892	0.3199
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0443	0.0283	-0.0997	0.0111	2.4526	0.1173
	Proprietario	0
Indice di massa corporea	Normopeso	0.0107	0.0304	-0.0489	0.0703	0.1247	0.7240
	Sottopeso	-0.1074	0.1101	-0.3232	0.1084	0.9520	0.3292
	Sovrappeso	0.00672	0.0292	-0.0504	0.0639	0.0531	0.8177
	Obeso	0
SES		0.0602	0.0123	0.0362	0.0842	24.1340	<.0001
Salute fisica		0.3761	0.0129	0.3509	0.4013	854.1546	<.0001
Salute fisica* SES		0.00980	0.0155	-0.0207	0.0402	0.3976	0.5283
Scale		0.8419	0.0075	0.8274	0.8566		

Tabella G.3 – Stime del modello per la risposta nella wave 2

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.1647	0.1326	-0.0952	0.4246	1.5421	0.2143
Genere	Femmina	-0.2745	0.0237	-0.3209	-0.2281	134.4994	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2186	0.0354	0.1491	0.2880	38.0217	<.0001
	Medio	0.1438	0.0301	0.0849	0.2027	22.8952	<.0001
	Basso	0
Età	50 - 59	-0.1426	0.1128	-0.3636	0.0785	1.5982	0.2062
	60 - 69	-0.2825	0.1126	-0.5033	-0.0618	6.2924	0.0121
	70 - 79	-0.4356	0.1142	-0.6594	-0.2118	14.5582	0.0001
	80 +	-0.7498	0.1221	-0.9891	-0.5105	37.7132	<.0001
	< 50	0
Stato civile	Celibe/nubile	0.00629	0.0597	-0.1107	0.1233	0.0111	0.9161

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Separato/a o divorziato/a	0.0521	0.0526	-0.0511	0.1553	0.9793	0.3224
	Sposato/a	0.0722	0.0373	-0.00099	0.1454	3.7384	0.0532
	Unione registrata	0.0643	0.0963	-0.1245	0.2530	0.4452	0.5046
	Vedovo/a	0
Paese di residenza	Austria	-0.1997	0.0663	-0.3296	-0.0699	9.0875	0.0026
	Belgio	-0.2110	0.0534	-0.3158	-0.1063	15.5906	<.0001
	Danimarca	-0.0513	0.0592	-0.1674	0.0648	0.7494	0.3867
	Francia	-0.1276	0.0592	-0.2436	-0.0115	4.6419	0.0312
	Germania	-0.2180	0.0616	-0.3387	-0.0973	12.5316	0.0004
	Italia	-0.2304	0.0580	-0.3441	-0.1168	15.7918	<.0001
	Spagna	-0.2178	0.0611	-0.3375	-0.0980	12.7072	0.0004
	Svezia	-0.0130	0.0571	-0.1248	0.0988	0.0518	0.8200
	Svizzera	0
	Situazione abitativa	Non proprietario	-0.0995	0.0300	-0.1584	-0.0406	10.9738
Proprietario		0
Indice di massa corporea	Normopeso	0.4534	0.0319	0.3909	0.5159	202.0831	<.0001
	Sottopeso	0.2316	0.1173	0.00174	0.4615	3.8999	0.0483
	Sovrappeso	0.3242	0.0308	0.2638	0.3845	110.7539	<.0001
	Obeso	0
SES		0.0405	0.0125	0.0160	0.0649	10.5244	0.0012
Scale		0.8972	0.0080	0.8817	0.9129		

Tabella G.4 – Stime del modello per il mediatore nella wave 2

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Intercetta	0.4262	0.1977	0.0387	0.8136	4.6461	0.0311
Genere	Femmina	-0.2208	0.0220	-0.2640	-0.1776	100.2841	<.0001
	Maschio	0
Livello di istruzione	Alto	0.2073	0.0329	0.1429	0.2718	39.7688	<.0001
	Medio	0.1363	0.0278	0.0819	0.1907	24.0894	<.0001
	Basso	0
Età	50 - 59	-0.2366	0.1879	-0.6049	0.1317	1.5851	0.2080
	60 - 69	-0.1164	0.1867	-0.4823	0.2495	0.3888	0.5329

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	70 - 79	-0.1644	0.1873	-0.5315	0.2027	0.7704	0.3801
	80 +	-0.1509	0.1893	-0.5220	0.2201	0.6355	0.4253
	< 50	0
	Celibe/nubile	-0.0187	0.0539	-0.1243	0.0869	0.1203	0.7287
	Separato/a o divorziato/a	-0.0463	0.0471	-0.1386	0.0461	0.9641	0.3262
Stato civile	Sposato/a	0.0585	0.0329	-0.00602	0.1231	3.1585	0.0755
	Unione registrata	0.0693	0.0869	-0.1010	0.2396	0.6359	0.4252
	Vedovo/a	0
	Austria	-0.0131	0.0630	-0.1366	0.1104	0.0432	0.8353
	Belgio	-0.2928	0.0497	-0.3902	-0.1953	34.6750	<.0001
	Danimarca	0.0646	0.0557	-0.0447	0.1738	1.3427	0.2466
	Francia	-0.2999	0.0558	-0.4093	-0.1905	28.8744	<.0001
Paese di residenza	Germania	-0.1843	0.0581	-0.2982	-0.0703	10.0490	0.0015
	Italia	-0.7392	0.0545	-0.8460	-0.6323	183.9200	<.0001
	Spagna	-0.4905	0.0575	-0.6033	-0.3777	72.6943	<.0001
	Svezia	-0.1043	0.0532	-0.2085	-0.00004	3.8441	0.0499
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0903	0.0272	-0.1437	-0.0370	11.0355	0.0009
	Proprietario	0
	Normopeso	-0.0130	0.0299	-0.0716	0.0455	0.1903	0.6627
Indice di massa corporea	Sottopeso	-0.1749	0.1034	-0.3776	0.0278	2.8590	0.0909
	Sovrappeso	0.00692	0.0288	-0.0495	0.0633	0.0578	0.8100
	Obeso	0
	SES	0.0687	0.0125	0.0441	0.0932	29.9899	<.0001
	Salute fisica	0.3835	0.0127	0.3586	0.4083	914.7321	<.0001
	Salute fisica* SES	-0.0657	0.0161	-0.0974	-0.0341	16.5842	<.0001
	Scale	0.8295	0.0073	0.8152	0.8440		

Tabella G.5 – Stime del modello per la risposta nella wave 4

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.2018	0.2056	-0.2012	0.6047	0.9629	0.3264
Genere	Femmina	-0.2588	0.0227	-0.3033	-0.2143	129.8449	<.0001
	Maschio	0
	Alto	0.1559	0.0341	0.0890	0.2228	20.8627	<.0001
Livello di istruzione	Medio	0.1180	0.0288	0.0615	0.1745	16.7684	<.0001
	Basso	0
	50 - 59	-0.0408	0.1955	-0.4241	0.3424	0.0437	0.8345
	60 - 69	-0.1410	0.1942	-0.5217	0.2397	0.5272	0.4678
Età	70 - 79	-0.3162	0.1948	-0.6980	0.0657	2.6335	0.1046
	80 +	-0.7625	0.1967	-1.1481	-0.3769	15.0226	0.0001
	< 50	0
	Celibe/nubile	0.0669	0.0560	-0.0429	0.1767	1.4241	0.2327
	Separato/a o divorziato/a	0.1105	0.0490	0.0144	0.2065	5.0840	0.0241
Stato civile	Sposato/a	0.1039	0.0342	0.0369	0.1710	9.2368	0.0024
	Unione registrata	-0.0278	0.0904	-0.2049	0.1494	0.0944	0.7587
	Vedovo/a	0
	Austria	-0.1349	0.0655	-0.2633	-0.00644	4.2366	0.0396
	Belgio	-0.2720	0.0516	-0.3730	-0.1709	27.8252	<.0001
	Danimarca	-0.0812	0.0580	-0.1948	0.0324	1.9630	0.1612
	Francia	-0.2450	0.0580	-0.3586	-0.1314	17.8615	<.0001
Paese di residenza	Germania	-0.3064	0.0604	-0.4247	-0.1881	25.7634	<.0001
	Italia	-0.3086	0.0566	-0.4195	-0.1977	29.7616	<.0001
	Spagna	-0.3844	0.0597	-0.5013	-0.2675	41.5154	<.0001
	Svezia	-0.1092	0.0553	-0.2176	-0.00089	3.9051	0.0481
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0986	0.0283	-0.1540	-0.0432	12.1795	0.0005
	Proprietario	0
	Normopeso	0.4320	0.0306	0.3720	0.4920	198.9407	<.0001
Indice di massa corporea	Sottopeso	0.3650	0.1075	0.1543	0.5757	11.5243	0.0007
	Sovrappeso	0.3439	0.0296	0.2858	0.4019	134.7847	<.0001
	Obeso	0
SES		0.0296	0.0121	0.00586	0.0534	5.9671	0.0146

	Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Scale	0.8630	0.0076	0.8482	0.8781		

Tabella G.6 – Stime del modello per il mediatore nella wave 4

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Intercetta		0.0652	0.2748	-0.4734	0.6039	0.0563	0.8124
Genere	Femmina	-0.1843	0.0212	-0.2258	-0.1428	75.7520	<.0001
	Maschio	0
	Alto	0.1906	0.0315	0.1289	0.2523	36.6609	<.0001
Livello di istruzione	Medio	0.1524	0.0268	0.0999	0.2048	32.4073	<.0001
	Basso	0
	50 - 59	-0.0101	0.2698	-0.5390	0.5187	0.0014	0.9700
Età	60 - 69	0.1252	0.2674	-0.3989	0.6493	0.2191	0.6397
	70 - 79	0.0870	0.2677	-0.4376	0.6116	0.1058	0.7450
	80 +	0.0818	0.2687	-0.4448	0.6084	0.0927	0.7608
	< 50	0
	Celibe/nubile	-0.0764	0.0507	-0.1758	0.0230	2.2687	0.1320
Stato civile	Separato/a o divorziato/a	-0.0467	0.0445	-0.1339	0.0404	1.1043	0.2933
	Sposato/a	0.0463	0.0301	-0.0128	0.1054	2.3599	0.1245
	Unione registrata	0.0900	0.0868	-0.0800	0.2601	1.0773	0.2993
	Vedovo/a	0
	Austria	0.1029	0.0618	-0.0181	0.2239	2.7776	0.0956
Paese di residenza	Belgio	-0.1582	0.0489	-0.2541	-0.0623	10.4519	0.0012
	Danimarca	0.1653	0.0546	0.0583	0.2724	9.1657	0.0025
	Francia	-0.2163	0.0543	-0.3227	-0.1100	15.9025	<.0001
	Germania	-0.1177	0.0569	-0.2293	-0.00611	4.2737	0.0387
	Italia	-0.7707	0.0539	-0.8764	-0.6650	204.1993	<.0001
	Spagna	-0.4803	0.0569	-0.5918	-0.3688	71.3013	<.0001
	Svezia	-0.0803	0.0518	-0.1819	0.0213	2.3996	0.1214
	Svizzera	0
	Non proprietario	-0.0835	0.0263	-0.1350	-0.0319	10.0617	0.0015
Situazione abitativa	Proprietario	0
	Normopeso	0.0106	0.0283	-0.0449	0.0662	0.1410	0.7073

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Indice di massa corporea	Sottopeso	-0.0803	0.1080	-0.2919	0.1313	0.5536	0.4568
	Sovrappeso	0.0382	0.0273	-0.0154	0.0917	1.9532	0.1622
	Obeso	0
SES		0.0617	0.0120	0.0383	0.0852	26.6501	<.0001
Salute fisica		0.4618	0.0130	0.4363	0.4873	1263.2300	<.0001
Salute fisica* SES		-0.0235	0.0149	-0.0527	0.00564	2.4998	0.1139
Scale		0.7990	0.0071	0.7853	0.8130		

Tabella G.7 – Stime del modello per la risposta nella wave 5

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		0.3201	0.2773	-0.2234	0.8637	1.3326	0.2483
Genere	Femmina	-0.2160	0.0212	-0.2576	-0.1745	103.8258	<.0001
	Maschio	0
Livello di istruzione	Alto	0.1686	0.0317	0.1065	0.2307	28.2883	<.0001
	Medio	0.1228	0.0270	0.0699	0.1756	20.7366	<.0001
	Basso	0
Età	50 - 59	-0.1613	0.2723	-0.6951	0.3724	0.3510	0.5536
	60 - 69	-0.2188	0.2699	-0.7478	0.3101	0.6575	0.4174
	70 - 79	-0.3791	0.2701	-0.9085	0.1503	1.9700	0.1605
	80 +	-0.7349	0.2711	-1.2662	-0.2037	7.3514	0.0067
Stato civile	< 50	0
	Celibe/nubile	0.1134	0.0512	0.0131	0.2136	4.9136	0.0266
	Separato/a o divorziato/a	0.0867	0.0449	-0.00126	0.1746	3.7319	0.0534
Paese di residenza	Sposato/a	0.1326	0.0304	0.0731	0.1922	19.0500	<.0001
	Unione registrata	0.1313	0.0875	-0.0403	0.3029	2.2492	0.1337
	Vedovo/a	0
Paese di residenza	Austria	-0.1792	0.0622	-0.3011	-0.0574	8.3138	0.0039
	Belgio	-0.2673	0.0490	-0.3634	-0.1712	29.7162	<.0001
	Danimarca	-0.0843	0.0550	-0.1920	0.0234	2.3541	0.1250
	Francia	-0.2614	0.0545	-0.3681	-0.1547	23.0426	<.0001
	Germania	-0.2500	0.0573	-0.3622	-0.1378	19.0715	<.0001

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
Situazione abitativa	Italia	-0.3569	0.0541	-0.4629	-0.2509	43.5395	<.0001
	Spagna	-0.3534	0.0570	-0.4652	-0.2416	38.3735	<.0001
	Svezia	-0.0521	0.0522	-0.1544	0.0502	0.9967	0.3181
	Svizzera	0
	Non proprietario	-0.1066	0.0265	-0.1586	-0.0546	16.1490	<.0001
	Proprietario	0
Indice di massa corporea	Normopeso	0.3934	0.0282	0.3382	0.4486	195.1612	<.0001
	Sottopeso	0.2887	0.1089	0.0752	0.5022	7.0270	0.0080
	Sovrappeso	0.2718	0.0274	0.2182	0.3254	98.6974	<.0001
	Obeso	0
SES		0.0287	0.0117	0.00587	0.0516	6.0716	0.0137
Scale		0.8065	0.0071	0.7927	0.8206		

Tabella G.8 – Stime del modello per il mediatore nella wave 5

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Intercetta	-0.3589	0.4749	-1.2897	0.5718	0.5712	0.4498
Genere	Femmina	-0.1924	0.0219	-0.2354	-0.1494	76.8481	<.0001
	Maschio	0
	Alto	0.1641	0.0325	0.1003	0.2278	25.4652	<.0001
Livello di istruzione	Medio	0.1138	0.0276	0.0597	0.1679	17.0073	<.0001
	Basso	0
	50 - 59	0.4132	0.4759	-0.5195	1.3459	0.7540	0.3852
Età	60 - 69	0.4744	0.4711	-0.4490	1.3978	1.0139	0.3140
	70 - 79	0.4282	0.4712	-0.4952	1.3517	0.8260	0.3634
	80 +	0.4082	0.4716	-0.5160	1.3325	0.7494	0.3867
	< 50	0
Stato civile	Celibe/nubile	-0.0483	0.0517	-0.1497	0.0531	0.8715	0.3505
	Separato/a o divorziato/a	0.0156	0.0450	-0.0727	0.1039	0.1198	0.7292
	Sposato/a	0.0893	0.0304	0.0298	0.1488	8.6473	0.0033
	Unione registrata	0.1032	0.0901	-0.0734	0.2799	1.3124	0.2520
	Vedovo/a	0
Paese di residenza	Austria	0.1065	0.0650	-0.0210	0.2340	2.6790	0.1017

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
	Belgio	-0.0773	0.0520	-0.1791	0.0246	2.2124	0.1369
	Danimarca	0.1760	0.0580	0.0623	0.2897	9.2031	0.0024
	Francia	-0.2570	0.0582	-0.3711	-0.1429	19.4842	<.0001
	Germania	-0.0802	0.0608	-0.1993	0.0389	1.7404	0.1871
	Italia	-0.6343	0.0578	-0.7475	-0.5211	120.5404	<.0001
	Spagna	-0.4087	0.0613	-0.5288	-0.2885	44.4054	<.0001
	Svezia	-0.0337	0.0553	-0.1421	0.0747	0.3715	0.5422
	Svizzera	0
Situazione abitativa	Non proprietario	-0.0453	0.0274	-0.0990	0.00847	2.7261	0.0987
	Proprietario	0
Indice di massa corporea	Normopeso	0.0188	0.0292	-0.0383	0.0760	0.4176	0.5181
	Sottopeso	-0.0489	0.0972	-0.2394	0.1415	0.2537	0.6145
	Sovrappeso	0.0195	0.0283	-0.0359	0.0749	0.4750	0.4907
	Obeso	0
SES		0.0745	0.0148	0.0455	0.1035	25.3174	<.0001
Salute fisica		0.4889	0.0151	0.4592	0.5186	1043.299 5	<.0001
Salute fisica* SES		-0.0388	0.0201	-0.0781	0.000530	3.7386	0.0532
Scale		0.8135	0.0073	0.7993	0.8279		

Tabella G.9 – Stime del modello per la risposta nella wave 6

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P – value
Intercetta		-0.2227	0.4310	-1.0675	0.6222	0.2668	0.6055
Genere	Femmina	-0.1775	0.0198	-0.2163	-0.1387	80.3850	<.0001
	Maschio	0
Livello di istruzione	Alto	0.1548	0.0294	0.0971	0.2124	27.6741	<.0001
	Medio	0.1022	0.0250	0.0532	0.1512	16.7203	<.0001
	Basso	0
Età	50 - 59	0.4578	0.4320	-0.3889	1.3044	1.1229	0.2893
	60 - 69	0.3969	0.4277	-0.4414	1.2351	0.8610	0.3534
	70 - 79	0.2524	0.4277	-0.5859	1.0907	0.3483	0.5551
	80 +	-0.1111	0.4281	-0.9502	0.7279	0.0674	0.7952
	< 50	0
Stato civile	Celibe/nubile	0.0696	0.0469	-0.0224	0.1615	2.1968	0.1383

		Stima	Standard Error	Intervallo di confidenza (Wald 95%)		χ^2 di Wald	P-value
	Separato/a o divorziato/a	0.0282	0.0409	-0.0519	0.1083	0.4756	0.4904
	Sposato/a	0.0707	0.0275	0.0167	0.1246	6.5902	0.0103
	Unione registrata	0.1372	0.0818	-0.0231	0.2976	2.8150	0.0934
	Vedovo/a	0
	Austria	-0.1932	0.0590	-0.3089	-0.0776	10.7207	0.0011
	Belgio	-0.2576	0.0470	-0.3498	-0.1655	30.0320	<.0001
	Danimarca	-0.0804	0.0526	-0.1835	0.0228	2.3314	0.1268
	Francia	-0.2332	0.0528	-0.3366	-0.1298	19.5292	<.0001
Paese di residenza	Germania	-0.2038	0.0551	-0.3118	-0.0958	13.6765	0.0002
	Italia	-0.2399	0.0524	-0.3425	-0.1373	20.9926	<.0001
	Spagna	-0.2391	0.0556	-0.3481	-0.1302	18.5001	<.0001
	Svezia	-0.00957	0.0502	-0.1079	0.0888	0.0364	0.8488
	Svizzera	0
Situazione abitativa	Non proprietario	-0.1342	0.0247	-0.1827	-0.0858	29.4726	<.0001
	Proprietario	0
	Normopeso	0.3544	0.0261	0.3033	0.4056	184.5067	<.0001
	Sottopeso	0.1824	0.0881	0.00978	0.3551	4.2891	0.0384
Indice di massa corporea	Sovrappeso	0.2926	0.0254	0.2428	0.3424	132.7666	<.0001
	Obeso	0
SES		0.0272	0.0114	0.00487	0.0495	5.7016	0.0169
Scale		0.7385	0.0066	0.7257	0.7516		

Tabella G.10 – Stime del modello per il mediatore nella wave 6

RINGRAZIAMENTI

“Appreciation is a wonderful thing. It makes what is excellent in others belong to us as well.”

– Voltaire

Vorrei ringraziare in primis la mia relatrice, la Prof.ssa G. Boccuzzo, per avermi proposto un argomento così interessante, per la disponibilità e la pazienza dimostratemi sia nello svolgimento delle analisi, sia durante tutto il periodo di stesura, e per le innumerevoli opportunità che mi sta concedendo anche in vista del mio futuro dopo la laurea.

So di non essere una persona facile con cui avere a che fare e per questo motivo vorrei ringraziare una delle persone che mi sopporta da più tempo.

Franci ormai da ben 22 anni sei una delle mie più care amiche, la mia compagna di avventure e vacanze, grazie perché per ogni momento e ogni risata, per ogni consiglio che mi hai dato. Dopo aver iniziato la nostra carriera scolastica insieme mi sembra giusto essere arrivate anche alla fine di questo percorso universitario contemporaneamente. Ora possiamo andare in panico pensando al nostro futuro, ma almeno non lo faremo da sole.

Nonostante abbia deciso di studiare due anni in un'altra città, alcune persone ci sono sempre state, anche solo per parlare una manciata di minuti. Grazie Meli e Carlo per avermi ascoltata nei momenti di difficoltà, nonostante le mille paranoie e i mille castelli in aria che mi sono costruita. Grazie per la pazienza e per il supporto che mi date ogni giorno.

Un ringraziamento speciale va alla mia coinquilina, nonché migliore amica. Grazie per aver deciso di iniziare questa avventura a Padova con me, grazie

per avermi sopportata sempre, anche nei momenti peggiori, per avermi aiutata e spronata in tutti i momenti di difficoltà, per avermi sempre dato la tua opinione sincera, per le risate, per le serate infinite, per tutti i momenti di follia, per le mattine in silenzio ad ascoltare la musica, per le giornate di shopping e per le passeggiate che non finivano mai.

Mary non so come avrei potuto sopravvivere questi due anni senza di te, sai che sei e sarai sempre la mia *soul mate* e non importa in che parte del mondo andrai, non ti libererai di me e ci saremo sempre l'una per l'altra.

You keep me wild, I'll keep you safe.

A tutti gli amici, i compagni di studio e le persone che mi sono state accanto questi anni, che siate parte della mia vita da quando eravamo piccoli oppure da qualche mese non importa, mi avete aiutata a crescere e diventare quella che sono, quindi grazie.

Il ringraziamento più grande devo farlo ai miei genitori, Chiara e Roberto, e a mio fratello Davide. Il raggiungimento di questo traguardo è anche grazie a voi. Da 24 anni mi supportate non solo economicamente ma anche moralmente, mi consigliate e mi ascoltate. Avete sempre creduto in me e supportato le mie decisioni. Grazie per avermi sempre fatto credere di poter essere qualunque cosa nella vita e per l'amore che ogni giorno mi trasmettete.

Grazie di cuore a tutti,

Sara

