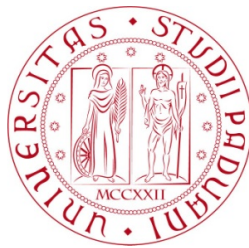


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



RELAZIONE FINALE

**UN CONFRONTO EMPIRICO SU POSSIBILI COMBINAZIONI
TRA TECNICHE DI IMPUTAZIONE E TRATTAMENTO DEL
CONFONDIMENTO TRAMITE PROPENSITY SCORE IN
PRESENZA DI DATI MANCANTI**

Relatrice: Prof.ssa Alessandra Rosalba Brazzale
Dipartimento di Scienze Statistiche

Correlatore: Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche

Laureanda: Anna Zanovello
Matricola N 1191541

Anno Accademico 2021/2022

Indice

Introduzione	7
1 Effetti causali in presenza di confondenti	9
1.1 Introduzione	9
1.2 Il modello causale di Rubin	10
1.3 Il propensity score	12
1.3.1 Propensity score matching	15
1.3.2 La stratificazione basata sul propensity score	18
1.3.3 Propensity score covariate adjustment	20
1.3.4 Inverse probability of treatment weighting	22
1.4 Valutazione del bilanciamento	24
1.5 Confronto tra i quattro metodi basati sul propensity score	27
2 Il problema dei dati mancanti	29
2.1 Introduzione	29
2.2 Nozioni di base	30
2.3 Tecniche di imputazione	32
2.3.1 L'imputazione multipla	35
2.4 Utilizzo del propensity score in presenza di dati mancanti	40

3	Una prima valutazione empirica	43
3.1	Introduzione	43
3.2	Disegno della simulazione	45
3.3	Trattamento dei dati mancanti	52
3.4	Analisi dei dati	53
3.5	Risultati	54
3.6	Discussione	59
3.7	Un esempio di caso di studio reale	64
4	Alcuni scenari realistici	81
4.1	Introduzione	81
4.2	Disegno della simulazione	82
4.3	Risultati	85
4.4	Discussione	86
4.5	Due varianti della simulazione	106
4.5.1	Diminuzione della numerosità campionaria	106
4.5.2	Variazione del modello di amputazione	107
	Conclusioni	107
A	Grafici di dispersione	115
A.1	Simulazione 1 Scenario B	116
A.2	Simulazione 1 Scenario C	120
A.3	Simulazione 1 Scenario D	124
A.4	Simulazione 2 Scenario A	128
A.5	Simulazione 2 Scenario B	132
A.6	Simulazione 2 Scenario C	136
A.7	Simulazione 2 Scenario D	140
A.8	Simulazione 3 Scenario A	144

A.9 Simulazione 3 Scenario B	148
A.10 Simulazione 3 Scenario C	152
A.11 Simulazione 3 Scenario D	156
A.12 Simulazione 4 Scenario A	160
A.13 Simulazione 4 Scenario B	164
A.14 Simulazione 4 Scenario C	168
A.15 Simulazione 4 Scenario D	172
A.16 Simulazione 5 Scenario A	176
A.17 Simulazione 5 Scenario B	180
A.18 Simulazione 5 Scenario C	184
A.19 Simulazione 5 Scenario D	188
A.20 Simulazione 6 Scenario A	192
A.21 Simulazione 6 Scenario B	196
A.22 Simulazione 6 Scenario C	200
A.23 Simulazione 6 Scenario D	204
A.24 Simulazione 7 Scenario A	208
A.25 Simulazione 7 Scenario B	212
A.26 Simulazione 7 Scenario C	216
A.27 Simulazione 7 Scenario D	220
A.28 Simulazione 8 Scenario A	224
A.29 Simulazione 8 Scenario B	228
A.30 Simulazione 8 Scenario C	232
A.31 Simulazione 8 Scenario D	236
A.32 Simulazione 9 Scenario A	240
A.33 Simulazione 9 Scenario B	244
A.34 Simulazione 9 Scenario C	248
A.35 Simulazione 9 Scenario D	252

Introduzione

Nel 1983 Rosenbaum e Rubin definiscono per la prima volta il *propensity score* e propongono un nuovo approccio che permetta di stimare l'effetto causale di una variabile trattamento su una risposta in presenza di confondenti. Da allora l'utilizzo di tale approccio è diventato sempre più diffuso e sono state proposte molte varianti al fine di rendere i metodi basati sul *propensity score* sempre più efficaci nel ridurre la distorsione causata dal confondimento. Esplorando la letteratura statistica si possono individuare molti studi atti a confrontare diverse modalità di utilizzo del *propensity score* per stabilire quali di essi abbiano prestazioni migliori.

I presupposti per il presente elaborato nascono dopo aver osservato che non ci sono altrettanti studi che forniscano delle linee guida per la scelta del metodo migliore quando il dataset da esaminare è affetto dal problema dei dati mancanti. Pertanto questo lavoro si pone l'obiettivo di confrontare le prestazioni dei quattro principali metodi per trattare il confondimento basati sul *propensity score* in combinazione con diverse tecniche di imputazione dei dati mancanti attraverso uno studio di simulazione. Nel condurre la simulazione si è reso necessario considerare anche altri aspetti che potessero influenzare i risultati e si è deciso di svolgere una serie di simulazioni in cui diverse caratteristiche dei dati vengono combinate nei modi possibili al fine di stabilire in ogni circostanza quale sia la combinazione tra i metodi basati

sul *propensity score* e quelli per la gestione dei dati mancanti che produce i risultati migliori.

La tesi è articolata in 4 capitoli. Nel Capitolo 1 si introduce la teoria riguardante il *propensity score* e i metodi di aggiustamento del confondimento basati su di esso. Nel Capitolo 2 si fa una panoramica sulle principali tecniche utilizzate per risolvere il problema dei dati mancanti. Nel Capitolo 3 si descrive il disegno di una prima indagine empirica e si presentano e discutono i risultati ottenuti. Infine nel Capitolo 4 si presenta una seconda indagine empirica ideata per valutare se le conclusioni osservate nel primo studio possano essere estese a situazioni più complesse.

Capitolo 1

Effetti causali in presenza di confondenti

1.1 Introduzione

Uno dei principali obiettivi che si pone l'analisi statistica è quello di stimare l'effetto causale di un dato trattamento su una variabile risposta di interesse. Il contesto migliore in cui è possibile fare ciò è quello degli studi clinici randomizzati, in cui il trattamento viene assegnato casualmente ai partecipanti allo studio. Ciò permette di ottenere due gruppi di soggetti confrontabili tra loro, poiché le caratteristiche individuali si distribuiscono casualmente tra i soggetti che hanno ricevuto il trattamento (trattati) e quelli che non lo hanno ricevuto (non trattati). Purtroppo studi di questo tipo sono spesso non attuabili per motivi etici o pratici, e sono pertanto sostituiti da studi osservazionali, in cui ci si limita a registrare quanto avviene ai componenti di un campione opportunamente selezionato. In questo caso avviene una auto-selezione al trattamento e può accadere che gli individui

con determinate caratteristiche abbiano una maggiore o minore probabilità di appartenere al gruppo dei trattati o a quello dei non trattati e, di conseguenza, si creino differenze sistematiche tra di essi. È perciò fondamentale capire se l'effetto sulla risposta osservato nel campione sia effettivamente dovuto al trattamento o se dipenda invece dalle caratteristiche individuali. Le variabili che hanno un effetto sia sul trattamento sia sulla risposta oggetto di studio vengono dette confondenti. Esistono vari metodi per controllarne l'effetto negli studi osservazionali, ma quelli di cui si occupa questa tesi sono basati sul *propensity score* (Rosenbaum e Rubin, 1983), definito come la probabilità di ricevere il trattamento condizionatamente alle caratteristiche individuali. Il *propensity score* può essere utilizzato essenzialmente in quattro modi diversi, che verranno esplorati nel presente capitolo.

1.2 Il modello causale di Rubin

Si consideri una popolazione di N soggetti per i quali si osservano le seguenti variabili: una variabile dicotomica T che assume valore 1 per i soggetti che hanno ricevuto il trattamento e 0 per i soggetti che non hanno ricevuto il trattamento; una variabile risposta Y ; una serie di covariate che indichiamo con X . Secondo il modello causale di Rubin (Holland, 1986), ad ogni individuo sono associate altre due variabili: Y_1 e Y_0 , che rappresentano il valore della risposta rispettivamente quando $T = 1$ e $T = 0$. In un mondo immaginario sarebbe possibile conoscere i valori di entrambe le variabili e si potrebbe quindi facilmente calcolare l'effetto del trattamento per ogni unità tramite la misura $Y_1 - Y_0$. Purtroppo, però, ciò non è possibile, e la variabile che osserviamo è invece: $Y = TY_1 + (1 - T)Y_0$. In sostanza, secondo questo modello si definiscono due mondi paralleli: quello effettivamente osservato

e il cosiddetto controfattuale, il quale rappresenta ciò che si osserverebbe se, contrariamente ai fatti, al soggetto fosse stato assegnato il trattamento opposto a quello ricevuto.

Per la stima dell'effetto del trattamento sono state proposte principalmente due quantità (Imbens, 2004): l'effetto medio del trattamento sulla popolazione totale (*Average Treatment Effect*, ATE) e l'effetto medio del trattamento sui trattati (*Average Treatment effect on Treated*, ATT). Le loro definizioni sono

$$\text{ATE} = \mathbf{E}[Y_1 - Y_0] = \mathbf{E}[Y_1] - \mathbf{E}[Y_0], \quad (1.1)$$

$$\text{ATT} = \mathbf{E}[Y_1 - Y_0 \mid T = 1] = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 1]. \quad (1.2)$$

Entrambe le misure hanno lo scopo di valutare quale sia l'effetto del trattamento, ma, mentre la prima ne valuta l'entità su tutta la popolazione, la seconda si condiziona al caso in cui i soggetti abbiano ricevuto il trattamento. Ciò può essere utile quando l'interesse è quello di valutare l'effetto del trattamento su chi effettivamente ha potuto sperimentarlo. Pertanto la scelta tra di esse deve basarsi sul contesto in cui ci si trova e sugli obiettivi dello studio.

In generale le due quantità non possono essere stimate senza distorsione, poiché, come detto precedentemente, non è possibile osservare entrambe le potenziali risposte per ogni soggetto. Un possibile stimatore dell'ATT è la differenza in media (DIM):

$$\text{DIM} = \mathbf{E}[Y \mid T = 1] - \mathbf{E}[Y \mid T = 0] = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 0]. \quad (1.3)$$

Risulta evidente che la stima non è corretta, in quanto $\mathbf{E}[Y_0 \mid T = 0] \neq \mathbf{E}[Y_0 \mid T = 1]$. La disuguaglianza tra le due quantità è dovuta alle differenze tra il gruppo dei trattati e dei non trattati prima di ricevere il trattamento, ovvero alle caratteristiche individuali di cui si è discusso nell'introduzione, che formalmente sono parte delle covariate X . Sono proprio queste caratteristiche

a rendere diverse le potenziali risposte nei casi in cui si riceva il trattamento o meno, e ciò ancora prima che il trattamento venga assegnato.

La distorsione può essere scomposta in tre diverse componenti (Heckman, Ichimura e Todd, 1997). La prima è dovuta alla mancata sovrapposizione dei supporti delle covariate X tra i gruppi dei trattati e dei non trattati. La seconda e la terza riguardano invece lo sbilanciamento delle distribuzioni dei confondenti osservati e non osservati rispettivamente. Ciò significa che le unità trattate e quelle non trattate non differiscono unicamente a causa del trattamento subito, ma anche rispetto ad altre variabili che possono essere osservate (le variabili X) o mancanti. Affinché queste differenze non abbiano effetto sulle stime, sarebbe auspicabile individuare per ogni trattato un opportuno sostituto che abbia le sue stesse caratteristiche di base ma che abbia ricevuto il trattamento opposto.

Sulla base di questo ragionamento si fondano i metodi basati sul *propensity score* e, più in generale, sui *balancing score*, i quali permettono di superare i problemi descritti finora e di ottenere stime non distorte dell'effetto del trattamento, a patto di rispettare delle opportune assunzioni presentate nel prossimo paragrafo.

1.3 Il propensity score

Si definisce *balancing score*, $b(X)$, una funzione delle covariate osservate tale che la distribuzione condizionale di X dato $b(X)$ è la stessa per i trattati ($T = 1$) e per i non trattati ($T = 0$), ovvero $X \perp\!\!\!\perp T \mid b(X)$ (Rosenbaum e Rubin, 1983). Un esempio di *balancing score* sono le stesse variabili X . Una funzione più interessante è il *propensity score*

$$e(X) = \mathbf{P}(T = 1 \mid X),$$

che corrisponde alla probabilità condizionata di avere il trattamento date le covariate X . È stato dimostrato che il *propensity score* è la versione più grezza possibile di *balancing score* (Rosenbaum e Rubin, 1983).

Per fare uso dei metodi basati sul *propensity score* è necessario fare alcune assunzioni. Innanzitutto la selezione del trattamento deve essere avvenuta prima della realizzazione di Y . Successivamente bisogna assumere che l'assegnazione al trattamento sia *strongly ignorable*: $(Y_1, Y_0) \perp\!\!\!\perp T \mid X$, $0 < \mathbf{P}(T = 1 \mid X) < 1$. La prima parte dell'assunzione è anche chiamata *unconfoundedness*, ed equivale ad assumere che non ci siano confondenti non osservati. La seconda invece asserisce che la probabilità per ogni soggetto di ricevere il trattamento è positiva. Infine deve essere rispettata l'assunzione di *stable unit treatment value*, secondo cui le due potenziali risposte per un individuo non sono influenzate dal valore del trattamento di nessun altro individuo (Williamson et al., 2012).

Sotto l'assunzione di assegnazione al trattamento *strongly ignorable*, la DIM (1.3) calcolata per ogni valore del *balancing score* è una stima non distorta dell'effetto del trattamento per tale valore di $b(X)$ (Rosenbaum e Rubin, 1983). Questo ci rassicura sul fatto che utilizzare i metodi basati sui *balancing score*, e quindi anche sul *propensity score*, possa portare a stime non distorte dell'effetto del trattamento.

Il primo passo per utilizzare il *propensity score* è ottenerne una stima appropriata, poiché il suo valore esatto non può essere noto in uno studio osservazionale. In generale, sostituire il *propensity score* con una sua stima non introduce ulteriore distorsione negli stimatori dell'effetto del trattamento (Drake, 1993). Il metodo più intuitivo e comune è quello di adattare un modello di regressione logistica, facendo una regressione del trattamento T sulle covariate X . Si può ottenere facilmente una stima del *propensity sco-*

re per ogni individuo calcolando i valori predetti della probabilità di avere il trattamento. Negli ultimi anni sono state proposte altre tecniche di *machine learning* che permettessero di ottenere stime più accurate del *propensity score*, quali alberi di regressione (CART), reti neurali e foreste casuali (McCaffrey, Ridgeway e Morral, 2004; Setoguchi et al., 2008; Lee, Lessler e Stuart, 2010; Westreich, Lessler e Funk, 2010; Wyss et al., 2014). Questi metodi sono spesso di difficile interpretazione e risultano più complessi da utilizzare, soprattutto per i profani dell'analisi statistica, tuttavia la loro flessibilità può portare ad un miglior bilanciamento dei confondenti e ad una riduzione della distorsione nella stima dell'effetto del trattamento. Proprio questi sono i principali obiettivi dell'inferenza causale. Nel caso si utilizzi un metodo parametrico, è naturale chiedersi quali siano le variabili da inserire nella regressione come esplicative. In merito a questo si possono trovare in letteratura alcune linee guida (Austin, 2011b; Williamson et al., 2012): tutte le variabili che sono al contempo predittori della risposta e del trattamento (i confondenti) devono essere incluse per poter ottenere stime corrette dell'effetto del trattamento; inoltre, se il modello comprende anche le variabili che influenzano unicamente la risposta Y , tipicamente si ottiene una riduzione della varianza dell'effetto del trattamento senza che ci sia un aumento della distorsione; si ha un risultato opposto considerando, invece, i predittori del solo trattamento, che dovrebbero essere preferibilmente omessi (Brookhart et al., 2006; Rubin, 2004). Non sempre è possibile valutare come le covariate influenzino la risposta e il trattamento, perciò si può incorrere nella scorretta specificazione del modello. In ogni caso, è fondamentale che le esplicative non siano state influenzate o modificate dal trattamento, ovvero siano precedenti alla sua assegnazione (Austin, 2011b).

Una volta stimato, il *propensity score* può essere usato per bilanciare i

confondenti nei due gruppi di trattati e non trattati e ottenere così stime non distorte dell'effetto del trattamento. I metodi di utilizzo sono essenzialmente quattro: il *propensity score matching*, la stratificazione basata sul *propensity score*, il *propensity score covariate adjustment* e l'*inverse probability of treatment weighting*.

1.3.1 Propensity score matching

Il *matching* è la tecnica più comune basata sul *propensity score*. Esso consiste nel creare un campione di dati appaiati costituito da coppie di soggetti trattati e non trattati con valori simili di *propensity score*. In questo modo si ottiene un nuovo insieme di dati che simula le caratteristiche di quelli ottenuti con gli studi clinici randomizzati, in cui i confondenti osservati sono bilanciati tra i due gruppi di trattamento. Successivamente è possibile stimare l'effetto del trattamento sul nuovo campione tramite la differenza delle medie delle risposte di trattati e non trattati (1.3). Per come è costruito il campione di dati appaiati, questo metodo è adatto alla stima dell'ATT (Abdia et al., 2017).

Svariati sono i vantaggi che si ottengono dall'utilizzo del *propensity score matching*. In generale, nei campioni di dati appaiati per qualsiasi possibile confondente si ottiene una varianza dell'effetto medio del trattamento stimato inferiore rispetto a quella che si otterrebbe in campioni casuali, poiché le distribuzioni delle covariate X sono più simili tra i gruppi di trattati e non trattati nei primi che nei secondi (Rosenbaum e Rubin, 1983). Inoltre, l'aggiustamento basato sulla regressione risulta solitamente più robusto rispetto a distaccamenti dalla forma assunta dal modello sottostante i dati se applicato su un campione di dati appaiati invece che su un campione casuale (Rubin, 1973; Rubin, 1979). Eseguire il *matching* rispetto a tutti i

possibili confondenti può essere, però, molto complesso, soprattutto in campioni di dimensione ridotta. Pertanto il *propensity score matching* rappresenta una buona soluzione, poiché in grado di bilanciare tutte le covariate contemporaneamente.

Ci sono diversi modi di applicare il *matching*. Prima di tutto, si può decidere di fare un appaiamento uno a uno, in cui ad ogni soggetto trattato viene associato un unico non trattato, oppure altre varianti come il *matching* uno a molti o il *full matching*, in cui si possono formare gruppi di appaiamento con un trattato e più non trattati o viceversa. Poiché questi ultimi sono poco diffusi, d'ora in poi faremo riferimento unicamente alla modalità uno a uno. Si rimanda ad altri articoli per un approfondimento sugli altri metodi: Rosenbaum (1991), Gu e Rosenbaum (1993), Ming e Rosenbaum (2000) e Hansen (2004).

In secondo luogo, si può eseguire un *matching* senza o con ripetizione. Nel primo caso, una volta fatto l'appaiamento tra due unità, queste vengono escluse in modo che non possano più essere estratte. Al contrario, il secondo prevede che i soggetti possano far parte di più appaiamenti. In questo caso, nello stimare la varianza dell'effetto del trattamento è necessario tenere conto del fatto che alcune unità possono essere presenti in più di una coppia di individui appaiati. Austin (2014) scoraggia l'uso del *matching* con ripetizione perché non sembra avere una migliore performance rispetto all'alternativa senza ripetizione.

Il *matching* si differenzia poi tra *greedy* o *optimal*. L'algoritmo alla base del *greedy matching* è il seguente: un soggetto nel gruppo dei trattati viene casualmente selezionato, poi si individua tra i non trattati quello con valore di *propensity score* più vicino e i due soggetti vengono appaiati. La procedura viene ripetuta fino a che non si terminano i trattati o fino a che non è più

possibile trovare coppie. Questo metodo può essere chiamato anche *nearest neighbour*, ed è definito *greedy* (in italiano ingordo) poiché in ogni passo del ciclo il soggetto non trattato viene selezionato per l'appaiamento senza valutare se possa essere più affine ad un altro trattato presente nel campione. Per questo motivo è stato proposto l'utilizzo dell'*optimal matching*, che consiste nel minimizzare una funzione obiettivo definita dalla differenza totale dei *propensity score* tra le coppie di trattati e non trattati generate. Questa tecnica però non sembra portare dei particolari miglioramenti nel bilanciare i confondenti (Gu e Rosenbaum, 1993). Per quanto riguarda invece il metodo *nearest neighbor*, una sua variante prevede l'utilizzo di una tolleranza massima prestabilita (*caliper*) per la distanza tra i valori di *propensity score*. In sostanza, quando viene selezionato il soggetto non trattato con *propensity score* più vicino al soggetto trattato estratto, si introduce il vincolo che la differenza assoluta tra i valori di *propensity score* sia al di sotto di un certa quantità. In questo modo, può accadere che alcune unità non trovino un opportuno appaiamento e siano escluse dal campione. Tuttavia, questo vincolo permette di non avere coppie costituite da individui troppo distanti tra loro in termini di *propensity score*. Vari studi sono stati fatti con lo scopo di stabilire quale sia il valore più adatto per la tolleranza. In modo concorde viene proposto di appaiare i soggetti basandosi sul *logit* del *propensity score*, cioè sulla quantità $\log\{\widehat{e}(X)/[1 - \widehat{e}(X)]\}$ con $\widehat{e}(X)$ pari al *propensity score* stimato, poiché è più facile che questa quantità segua una distribuzione normale, e viene suggerito di usare una tolleranza proporzionale alla deviazione standard del *logit* del *propensity score*. In particolare, se tale deviazione standard è uguale per i soggetti trattati e quelli non trattati, una tolleranza pari allo 0.2 della deviazione standard complessiva del *logit* del *propensity score* porta ad una riduzione della distorsione dovuta ai confondenti approssimativamen-

te del 99% (Rosenbaum e Rubin, 1985). Inoltre, secondo Austin (2011d), questo valore, o uno simile, minimizza l'errore quadratico medio dell'effetto del trattamento stimato. Negli anni altre modifiche delle procedure sopra spiegate sono state proposte e valutate. Si segnala il seguente articolo per un confronto approfondito tra le metodologie più comuni: Austin (2014).

Sebbene il *matching* permetta di ottenere un campione bilanciato rispetto ai confondenti osservati, è possibile che i dati appaiati presentino uno sbilanciamento residuo. Questo problema può essere risolto eseguendo, come ulteriore passo, una regressione della risposta Y con le variabili T e X come esplicative, così da ottenere una stima dell'effetto del trattamento completamente depurata dall'effetto dei confondenti.

1.3.2 La stratificazione basata sul propensity score

La stratificazione basata sul *propensity score* consiste nel suddividere i soggetti in sottogruppi mutualmente esclusivi, detti strati, tali che le unità appartenenti allo stesso strato abbiano valori simili di *propensity score*. Se gli strati ottenuti sono perfettamente omogenei rispetto al *propensity score*, o più in generale rispetto ad un qualsiasi *balancing score*, allora è garantito che le covariate X abbiano la stessa distribuzione per i soggetti trattati e quelli non trattati in ognuno degli strati (Rosenbaum e Rubin, 1983). Nella pratica ciò non è possibile, e rimarrà quindi uno sbilanciamento residuo dei confondenti. La stratificazione potrebbe essere anche applicata basandosi direttamente sui confondenti, tuttavia, se questi ultimi sono numerosi, il numero di strati necessari cresce esponenzialmente (Cochran, 1968). Utilizzando invece il *propensity score* possiamo ottenere gli stessi risultati con una procedura molto più semplice. La stratificazione basata sul *propensity score* è un metodo che si può applicare facilmente ed è semplice da comprendere.

Tuttavia, è necessario che il modello con cui si stima il *propensity score* sia correttamente specificato affinché le distribuzioni delle covariate siano simili tra i trattati e i non trattati in ogni strato, e di conseguenza sia possibile ottenere un buon bilanciamento.

Per eseguire la stratificazione, i soggetti vengono ordinati secondo i loro valori di *propensity score* e, poi, suddivisi tramite dei valori soglia. Comunemente si creano 5 strati di uguale dimensione utilizzando i quintili del *propensity score*. Infatti è dimostrato che in questo modo si elimina approssimativamente il 90% della distorsione dovuta ai confondenti osservati quando si stima un effetto del trattamento lineare (Rosenbaum e Rubin, 1984 basandosi su un risultato di Cochran, 1968). Aumentando il numero degli strati la distorsione si riduce ulteriormente, ma maggiore è il numero degli strati, minore è la riduzione della distorsione che ne consegue (Cochran, 1968).

Il passo successivo è quello di stimare l'effetto del trattamento in ognuno degli strati tramite la differenza media delle risposte tra trattati e non trattati, ottenendo quindi delle stime strato-specifiche. Per ottenere una stima dell'effetto del trattamento complessiva si esegue una media ponderata delle stime strato-specifiche. I pesi scelti differiscono in base a quale misura d'effetto si intende stimare. Per la stima dell'ATE (1.1) si utilizzano come pesi il numero di osservazioni per strato, mentre per stimare l'ATT (1.2) si utilizza il numero di trattati per strato

$$\widehat{\text{ATE}} = \sum_{k=1}^K \left(\frac{N_k}{N} \right) \widehat{\tau}_k,$$

$$\widehat{\text{ATT}} = \sum_{k=1}^K \left(\frac{N_{T_k}}{N_T} \right) \widehat{\tau}_k,$$

dove N_k è il numero di osservazioni nello strato k -esimo, N_{T_k} è il numero di soggetti trattati nello strato k -esimo, N è il numero complessivo di osservazioni nel campione, N_T è il numero di soggetti trattati nel campione, $\widehat{\tau}_k$ è

pari alla stima di effetto strato-specifica e K è il numero di strati (Imbens, 2004; Abdia et al., 2017). Anche la varianze strato-specifiche possono essere similmente aggregate tra gli strati. In particolare, se si vuole stimare l'ATE, la relativa stima della varianza sarà

$$\widehat{\mathbf{Var}}(\widehat{\text{ATE}}) = \sum_{k=1}^K \left(\frac{N_k}{N}\right)^2 \widehat{\mathbf{Var}}(\widehat{\tau}_k),$$

mentre se si sceglie come misura l'ATT si avrà

$$\widehat{\mathbf{Var}}(\widehat{\text{ATT}}) = \sum_{k=1}^K \left(\frac{N_{T_k}}{N_T}\right)^2 \widehat{\mathbf{Var}}(\widehat{\tau}_k)$$

(Abdia et al., 2017). Anche per la stratificazione è possibile eseguire una regressione della risposta sui confondenti qualora sia necessario correggerne ulteriormente lo sbilanciamento tra il gruppo dei trattati e quello dei non trattati e ridurre quindi la distorsione delle stime.

1.3.3 Propensity score covariate adjustment

Un altro metodo di applicazione del *propensity score* è il *covariate adjustment*, in cui esso viene utilizzato come variabile esplicativa in un modello di regressione in cui la variabile indipendente è la risposta di interesse Y e le variabili dipendenti sono il trattamento T e il *propensity score* stimato $\widehat{e}(X)$ appunto. La scelta di quale sia la tipologia di modello di regressione da adattare dipende dalla variabile risposta: se è continua si utilizza un modello di regressione lineare, mentre se è dicotomica o discreta è indicato l'utilizzo di un'opportuna forma di modello lineare generalizzato. La stima dell'effetto del trattamento si ottiene poi tramite il coefficiente di regressione associato al trattamento stimato adattando il modello. Tale parametro corrisponde alla differenza delle medie della risposta se si adatta un modello lineare, mentre è pari al rapporto delle quote (*odds ratio*) se il modello stimato è una

regressione logistica (Austin, 2011b). In questo caso la stima della varianza dell'effetto del trattamento risulta quindi facilmente calcolabile e corrisponde alla stima della varianza del coefficiente di regressione stimato tramite il modello. Se si vuole specificatamente ottenere una stima di ATE e ATT, si possono utilizzare i seguenti stimatori (Imbens, 2004)

$$\widehat{\text{ATE}} = \frac{1}{N} \sum_{i=1}^N \{\mathbf{E}[Y | T = 1, \widehat{e}(x_i)] - \mathbf{E}[Y | T = 0, \widehat{e}(x_i)]\},$$

$$\widehat{\text{ATT}} = \frac{1}{N_T} \sum_{i:t_i=1} \{\mathbf{E}[Y | T = 1, \widehat{e}(x_i)] - \mathbf{E}[Y | T = 0, \widehat{e}(x_i)]\},$$

in cui x_i è il vettore delle covariate osservate per l' i -esima unità su un totale di N soggetti, N_T è il numero di soggetti trattati nel campione e t_i vale 1 se l' i -esima unità ha ricevuto il trattamento e 0 altrimenti. Se la regressione adattata ai dati è un modello lineare, allora le due quantità corrispondono alla stima del coefficiente relativo alla variabile T , come detto prima. Ma, se si utilizza un modello lineare generalizzato, è necessario utilizzare la differenza dei valori predetti con $T = 1$ e $T = 0$ per tutti i valori di *propensity score* calcolati nel campione (per la stima dell'ATE) o nel gruppo dei trattati (per la stima dell'ATT). In questo caso la stima della varianza può essere ottenuta con la tecnica del *bootstrap* (Yu et al., 2021; Efron e Tibshirani, 1993). Se per esempio si utilizza la regressione logistica, lo stimatore che si ottiene è

$$\widehat{\text{ATT}} = \frac{1}{N_T} \sum_{i:t_i=1} \left(\frac{e^{\beta_0 + \beta_1 + \beta_2 e(x_i)}}{1 + e^{\beta_0 + \beta_1 + \beta_2 e(x_i)}} - \frac{e^{\beta_0 + \beta_2 e(x_i)}}{1 + e^{\beta_0 + \beta_2 e(x_i)}} \right).$$

A differenza di ciò che accade con i due metodi visti in precedenza, con il *propensity score covariate adjustment* il bilanciamento dei confondenti e la stima dell'effetto del trattamento avvengono nello stesso momento. Con questa tecnica è però necessario dare una precisa forma alla relazione tra la risposta e il trattamento.

Sotto opportune condizioni, i risultati che si ottengono con il *propensity score covariate adjustment* sono analoghi a quelli che si otterrebbero adattando un modello di regressione in cui le variabili dipendenti sono i confondenti X e il trattamento T (Rosenbaum e Rubin, 1983). Tuttavia, questo metodo risulta vantaggioso perché è possibile dividere l'analisi in due passi: inizialmente per stimare il *propensity score* si adatta un modello che può essere anche molto complesso senza che ci si preoccupi della sovrapparametrizzazione, poiché l'unico scopo è ottenere una buona stima; successivamente è sufficiente utilizzare per la stima dell'effetto del trattamento un modello più semplice che sia facilmente interpretabile e di cui si possa valutare la bontà di adattamento (D'Agostino, 1998).

1.3.4 Inverse probability of treatment weighting

Il metodo di *Inverse Probability of Treatment Weighting* (IPTW, Rosenbaum, 1987) si basa sulla costruzione di un nuovo campione in cui le distribuzioni delle covariate sono indipendenti dall'assegnazione al trattamento e di conseguenza sono bilanciate tra gli individui trattati e quelli non trattati (Austin, 2011b). Tutti i soggetti sono inclusi nel campione, ma sono pesati con pesi che dipendono dal loro valore di *propensity score*. Questo metodo può essere paragonato alla tecnica di campionamento in cui si utilizzano dei pesi per rendere un campione rappresentativo di una certa popolazione (Horvitz e Thompson, 1952). In questo caso i pesi servono per determinare l'importanza di ogni unità per la stima dell'effetto del trattamento. Il peso che viene attribuito al soggetto i -esimo è

$$w_i = \frac{t_i}{\widehat{e}(x_i)} + \frac{1 - t_i}{1 - \widehat{e}(x_i)},$$

da cui si ottengono i seguenti pesi rispettivamente per i trattati e i non trattati: $1/\widehat{e}(x_i)$ e $1/[1 - \widehat{e}(x_i)]$. Ciò significa che ogni soggetto viene pesato con l'inverso della probabilità di ricevere il trattamento che ha effettivamente ricevuto (Austin, 2011b). L'effetto del trattamento medio valutato sulla popolazione (ATE) può essere quindi stimato tramite la differenza di due medie pesate

$$\widehat{\text{ATE}} = \frac{1}{N} \sum_{i=1}^N \frac{y_i t_i}{\widehat{e}(x_i)} - \frac{1}{N} \sum_{i=1}^N \frac{y_i (1 - t_i)}{1 - \widehat{e}(x_i)},$$

con y_i pari al valore osservato di Y per l' i -esima unità. In questa espressione i pesi utilizzati non hanno somma pari 1, pertanto è stata proposta una versione modificata in cui la somma dei pesi è effettivamente 1 (Imbens, 2004)

$$\widehat{\text{ATE}} = \frac{\sum_{i=1}^N \frac{y_i t_i}{\widehat{e}(x_i)}}{\sum_{i=1}^N \frac{t_i}{\widehat{e}(x_i)}} - \frac{\sum_{i=1}^N \frac{y_i (1 - t_i)}{1 - \widehat{e}(x_i)}}{\sum_{i=1}^N \frac{(1 - t_i)}{1 - \widehat{e}(x_i)}}.$$

Per una stima dell'ATT è necessario utilizzare dei pesi differenti, ovvero

$$w_i = t_i + \frac{(1 - t_i)\widehat{e}(x_i)}{1 - \widehat{e}(x_i)}.$$

Lo stimatore che si ottiene in questo caso è quindi

$$\widehat{\text{ATT}} = \frac{\sum_{i=1}^N y_i t_i}{\sum_{i=1}^N t_i} - \frac{\sum_{i=1}^N \frac{y_i (1 - t_i) \widehat{e}(x_i)}{1 - \widehat{e}(x_i)}}{\sum_{i=1}^N \frac{(1 - t_i) \widehat{e}(x_i)}{1 - \widehat{e}(x_i)}}, \quad (1.4)$$

(Imbens, 2004). Questi non sono gli unici stimatori che si possono scegliere, infatti altre versioni sono state proposte da Lunceford e Davidian (2004). Per calcolare una stima della varianza di questi stimatori si raccomanda l'utilizzo del *bootstrap* (Imbens, 2004), ma per quanto riguarda lo stimatore dell'ATE si può ottenere una forma esplicita della varianza se il *propensity score* è stimato tramite regressione logistica (Lunceford e Davidian, 2004).

Un aspetto a cui prestare attenzione quando si utilizza questo metodo è la grandezza dei pesi. Infatti, in presenza di valori molto bassi di *propensity score* nei trattati, o molto alti nei non trattati, si otterranno di conseguenza dei pesi molto elevati che possono portare ad un aumento della varianza delle stime e ad una distorsione dei risultati. Questo problema può essere risolto eliminando dal campione i soggetti con valori di *propensity score* troppo estremi (Lee, Lessler e Stuart, 2011). Inoltre è importante che le stime dei valori di *propensity score* siano accurate, poiché intervengono in modo diretto nella stima dell'effetto del trattamento.

1.4 Valutazione del bilanciamento

Una volta applicati i metodi basati sul *propensity score* descritti, è consigliato verificare che si sia raggiunto un buon bilanciamento dei confondenti tra i soggetti trattati e quelli non trattati. Il modo più diffuso per farlo è utilizzare la differenza assoluta standardizzata delle medie (*Absolute Standardized Mean Difference*, ASMD), che si definisce per ogni confondente X

$$\text{ASMD} = \frac{|\bar{X}_T - \bar{X}_C|}{\sqrt{\frac{S_T^2 + S_C^2}{2}}}, \quad (1.5)$$

in cui \bar{X}_T e \bar{X}_C sono i valori medi della variabile X nel gruppo dei trattati (soggetti con $T = 1$) e dei non trattati (soggetti con $T = 0$) rispettivamente e S_T^2 e S_C^2 sono le varianze della variabile X nel gruppo dei trattati e dei non trattati rispettivamente. Per le covariate dicotomiche la quantità può essere definita nel modo seguente

$$\text{ASMD} = \frac{|\hat{p}_T - \hat{p}_C|}{\sqrt{\frac{\hat{p}_T(1 - \hat{p}_T) + \hat{p}_C(1 - \hat{p}_C)}{2}}}, \quad (1.6)$$

dove \widehat{p}_T e \widehat{p}_C rappresentano le proporzioni campionarie di soggetti con X pari a 1 nel gruppo trattati e in quello dei non trattati rispettivamente. Minore è il valore dell'ASMD, minore è la differenza in media del confondente tra il gruppo dei trattati e dei non trattati e, quindi, maggiore è il suo bilanciamento nel campione. Calcolare questa misura prima e dopo l'applicazione dei metodi sopra citati permette di verificare se questi ultimi siano stati efficaci nel ridurre lo sbilanciamento dei confondenti.

La valutazione del bilanciamento tramite l'ASMD subisce delle modificazioni in base a quale metodo basato sul *propensity score* si utilizza e a quale quantità si sceglie per misurare l'effetto del trattamento (Austin, 2009b; Abdia et al., 2017). Quando si utilizza il *propensity score matching* l'ASMD viene calcolato sul campione dei dati appaiati. Per la stratificazione, invece, per valutare il bilanciamento si calcola la differenza assoluta delle medie in ciascuno degli strati generati e successivamente si ottiene un valore unico tramite una media aritmetica standardizzata. Siano $\bar{X}_{T,k}$ e $\bar{X}_{C,k}$ i valori medi della variabile X nel gruppo dei trattati e dei non trattati rispettivamente calcolati nello strato k -esimo definito dal *propensity score* e sia K il numero totale di strati. Allora la quantità che si ottiene è la seguente

$$\text{ASMD}_{\text{strat}} = \frac{\frac{1}{K} \sum_{k=1}^K |\bar{X}_{T,k} - \bar{X}_{C,k}|}{\sqrt{\frac{S_T^2 + S_C^2}{2}}}.$$

Per quanto riguarda la tecnica dell'IPTW, è necessario sostituire la (1.5) con una sua versione pesata. In particolare, è necessario sostituire \bar{X}_T e \bar{X}_C con $(\sum_{i=1}^N x_i w_i) / (\sum_{i=1}^N w_i)$, in cui w_i è definito come nel paragrafo precedente nelle due varianti associate all'utilizzo dell'ATE o dell'ATT. Infine, quando si utilizza il *propensity score covariate adjustment*, si può utilizzare una specifica versione dell'ASMD detta *Weighted Couditional Standardized Absolute Difference* proposta da Austin (2008a). Per ogni covariata continua X

il primo passo è adattare un modello di regressione lineare definito in questo modo:

$$X = \beta_0 + \beta_1 T + \beta_2 e(X) + \beta_3 T e(X) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \quad (1.7)$$

Successivamente la quantità $\left| \widehat{\beta}_1 + \widehat{\beta}_3 e(x_i) \right| / \widehat{\sigma}$ viene stimata per ogni osservazione, dove $\widehat{\beta}_1$, $\widehat{\beta}_3$ e $\widehat{\sigma}$ sono le stime dei corrispondenti parametri nella (1.7). Infine, attraverso la media di questa quantità rispetto ai soggetti presenti nel campione, si ottiene la misura per confrontare il bilanciamento di X tra trattati e non trattati. Per le variabili dicotomiche la procedura può essere opportunamente adattata. La (1.7) viene sostituita con una regressione logistica; successivamente le differenze standardizzate per variabili dicotomiche (1.6) vengono calcolate tramite i valori predetti per ogni valore di *propensity score* osservato nel campione e, infine, se ne fa una media aritmetica. Oltre ad applicare queste modifiche, se la quantità scelta per misurare l'effetto del trattamento è l'ATT, è necessario sostituire il denominatore della (1.5) (o della (1.6) per covariate dicotomiche) con la deviazione standard della variabile X calcolata nel sottogruppo dei soggetti trattati.

Ci sono varie opinioni su quale siano i valori che l'ASMD dovrebbe assumere per essere considerato buono. In generale si considerano accettabili i valori inferiori a 0.1, per cui lo sbilanciamento residuo può essere considerato trascurabile (Austin, 2011b). In realtà, se il bilanciamento è avvenuto correttamente, ci si aspetterebbe che l'intera distribuzione delle covariate sia simile tra i soggetti trattati e quelli non trattati, e non solo la media. Pertanto, anche confrontare altri momenti delle variabili confondenti o le intere distribuzioni tramite dei grafici può essere utile per valutare l'efficacia delle procedure eseguite. Secondo Rubin (2004), per decidere il modo più opportuno per valutare il bilanciamento è necessario considerare la relazione tra le covariate X e la variabile risposta. Se, per esempio, c'è una relazione ap-

prossimativamente lineare tra la risposta e il logaritmo delle covariate, allora è necessario controllare il bilanciamento attraverso le media dei logaritmi; se nella relazione sono presenti anche termini quadratici allora bisogna verificare che le varianze dei confondenti siano il più possibile simili tra trattati e non trattati.

Controllare che ci sia un sufficiente bilanciamento delle covariate è sicuramente utile per capire se la procedura applicata ha funzionato, tuttavia non è sufficiente per stabilire se le stime dell'effetto del trattamento siano non distorte. Inoltre, sotto certe condizioni può accadere che le stime siano non distorte sebbene non si sia raggiunto un bilanciamento accettabile (Kara et al., 2021).

Oltre al bilanciamento, si può pensare di voler analizzare le stime dei valori di *propensity score*, ma fare una valutazione della loro accuratezza non è necessario. Diventa però fondamentale quando le stime stesse vengono utilizzate per stimare l'effetto del trattamento, come avviene per l'IPTW (Rubin, 2004).

1.5 Confronto tra i quattro metodi basati sul propensity score

Diversi studi sono stati fatti per confrontare le prestazioni dei quattro metodi basati sul *propensity score* proposti: Austin e Mamdani (2006), Austin (2006), Austin (2007), Austin, Grootendorst e Anderson (2007), Austin (2008b), Austin (2009b), Austin (2010), Austin (2013), Austin e Schuster (2016), Abdia et al. (2017), Elze et al. (2017), Franklin et al. (2017), Ma et al. (2021). Molti di essi mostrano che il *matching* elimina una porzione maggiore di distorsione sistematica nelle covariate tra i trattati e i non trat-

tati rispetto al *covariate adjustment* e alla stratificazione (Austin, 2011b; Austin, 2009b; Austin, Grootendorst e Anderson, 2007; Austin e Mamdani, 2006). In alcuni casi il *matching* e l'IPTW hanno risultati simili, ma ci sono alcuni studi che dimostrano la superiorità del *matching* (Austin, 2009b) e altri in cui l'IPTW ha prodotto risultati migliori (Austin, 2010; Austin, 2013; Austin e Schuster, 2016). Lo studio di Lunceford e Davidian (2004) mostra che la stratificazione porta a stime dell'ATE con una distorsione maggiore rispetto a diversi stimatori basati sull'IPTW. Infine, anche se nella maggior parte degli studi il *covariate adjustment* non è il metodo migliore e in alcune circostanze non è adatto per stimare l'effetto del trattamento (Abdia et al., 2017), in altri casi ha mostrato di avere buone prestazioni (Ma et al., 2021; Elze et al., 2017; Franklin et al., 2017).

In generale, il *covariate adjustment* ha come svantaggio il fatto di dover definire una relazione tra *propensity score* e la variabile risposta, che deve essere quindi corretta per portare a buoni risultati. Per gli altri tre metodi è sufficiente che il modello per la stima del *propensity score* sia correttamente specificato (Austin, 2011b). Infine, il *covariate adjustment* e l'IPTW sono più sensibili all'accuratezza delle stime del *propensity score* (Rubin, 2004).

Capitolo 2

Il problema dei dati mancanti

2.1 Introduzione

Si dice che un insieme di dati contiene valori mancanti quando per alcuni soggetti non si ha informazione riguardo ad una o più delle variabili rilevate. Nella pratica questa situazione si verifica spesso e i motivi per cui ciò avviene possono essere diversi. Alcuni esempi sono la mancata risposta dei soggetti ad alcune domande poste all'interno di un questionario o la presenza di errori nella procedura di raccolta dei dati. Molto spesso il problema dei dati mancanti viene semplicemente ignorato, perché la sua risoluzione può risultare complessa, e ci si limita a non considerare le osservazioni che presentano dei valori mancanti. Tuttavia, questo approccio porta nella maggior parte dei casi a risultati scorretti e non affidabili. Il modo migliore per gestire i dati mancanti è senza dubbio la prevenzione, ovvero pianificare gli studi al meglio affinché i dati che si raccolgono siano completi, ma quando questo non è possibile, è necessario individuare un'opportuna strategia per risolvere il problema.

2.2 Nozioni di base

La prima caratteristica con cui descrivere i dati mancanti è il *pattern*, che indica quali sono esattamente i valori mancanti nel dataset. Può mancare un intero record, quando per un soggetto nessuna delle variabili viene osservata, o può succedere invece che solo per una o alcune delle variabili del dataset ci siano dei valori mancanti per alcuni o tutti i soggetti del campione. I dati mancanti, poi, possono essere classificati in tre categorie in base al loro meccanismo, cioè a seconda della probabilità di essere mancanti in relazione al valore dei dati (van Buuren, 2018). In particolare, i dati mancanti si dicono:

- *Missing Completely At Random* (MCAR) se la probabilità che i dati siano mancanti è la stessa per tutti i soggetti, ovvero quando le cause per cui i dati sono mancati non sono legate in alcun modo alle variabili dello studio;
- *Missing At Random* (MAR) se la probabilità che i dati siano mancanti è la stessa all'interno di gruppi definiti da altre variabili osservate e pertanto le cause per cui i dati sono mancati sono legate alle variabili osservate;
- *Missing Not At Random* (MNAR) se la probabilità che i dati siano mancanti è diversa tra i soggetti ed è legata a variabili non osservate.

Il caso di dati MCAR è quello più semplice da trattare, ma spesso è poco realistico. Più plausibili sono gli altri due casi, ma al contempo sono di più difficile gestione, in particolare i dati MNAR. Infatti quando i dati sono MNAR si dice che il meccanismo che genera i dati mancanti è *non-ignorable* e deve essere pertanto considerato e incluso nei modelli che si utilizzano nel caso si voglia procedere ad un'imputazione dei dati. Negli altri due gruppi invece,

in cui la distribuzione delle variabili del dataset è la stessa nei sottogruppi definiti dalla presenza o assenza di dati mancanti, il meccanismo è *ignorable*, con l'ulteriore condizione che i parametri su cui si vuole fare inferenza siano distinti da quelli che regolano il processo generatore dei dati mancanti (Rubin, 1976). Non è possibile però determinare a quale categoria appartengono i dati, perché l'informazione necessaria per farlo è mancante (van Buuren, 2018; Rubin, 1976). È possibile solo stabilire che i dati non sono MCAR quando, per almeno una variabile, la mancanza di informazione è legata a dati osservati di altre variabili.

I dati mancanti possono essere trattati in diversi modi. Quello più semplice è senza dubbio la *complete case analysis*, che consiste nell'escludere dal campione le osservazioni che hanno almeno un valore mancante tra le variabili. Nonostante sia piuttosto comune, la *complete case analysis* non è una buona scelta, poiché molti studi dimostrano che porta a stime distorte in ipotesi MAR e MNAR. Solo per dati MCAR i risultati sono validi, tuttavia anche in questo caso risulta un metodo poco efficiente, perché comporta la perdita di parte dell'informazione contenuta nei dati (Donders et al., 2006; Heijden et al., 2006).

Un'altra tecnica piuttosto semplice è quella del *missing indicator*, in cui si aggiunge ai modelli statistici una variabile dicotomica che indica se il valore di una certa variabile è mancante o meno (Miettinen, 1985; Anderson, Basilevsky e Hum, 1983). Anche questo metodo è utilizzato spesso, nonostante non sembri avere buone prestazioni. La letteratura infatti ne sconsiglia l'uso, poiché può condurre a risultati distorti anche quando i dati sono MCAR (Donders et al., 2006; Greenland e Finkle, 1995; Heijden et al., 2006; Knol et al., 2010).

Un altro gruppo di metodi per risolvere il problema dei dati mancanti è

basato sulla costruzione di modelli statistici per i dati osservati (Little, 1992). Modellando la distribuzione congiunta dei dati è possibile fare inferenza sulle quantità di interesse tramite il metodo della massima verosimiglianza oppure utilizzando l'approccio bayesiano. Quando il meccanismo generatore dei dati mancanti è *non-ignorable*, è necessario includerlo nei modelli di cui sopra.

Infine si possono utilizzare i metodi basati sull'imputazione, in cui i dati mancanti vengono rimpiazzati con dei valori plausibili. Poiché l'imputazione è il metodo utilizzato nel presente lavoro, non si approfondiscono ulteriormente gli altri approcci e si rimanda invece al paragrafo successivo una descrizione più dettagliata delle varie tecniche di imputazione.

2.3 Tecniche di imputazione

L'imputazione, come anticipato, è un metodo per risolvere il problema dei dati mancanti, in cui questi ultimi vengono sostituiti con altri valori. L'imputazione può essere singola o multipla. Nel primo caso per ogni dato mancante viene imputato un solo valore opportunamente scelto e, successivamente, il nuovo dataset completo viene analizzato come se i dati fossero stati effettivamente osservati o raccolti. Esistono diverse strategie per stabilire quali siano i valori da imputare. Il metodo più semplice è l'imputazione della media, in cui i valori mancanti per ogni variabile sono sostituiti con la sua media campionaria. Con questa tecnica, tuttavia, si ottengono stime inconsistenti della varianza (Little, 1992) e si possono ottenere stime distorte anche per dati MCAR (Donders et al., 2006). Infatti, imputando la media, si ha come conseguenza una riduzione della variabilità della distribuzione e, inoltre, le relazioni tra variabili tendono ad attenuarsi. Per questi motivi l'utilizzo di questo metodo non è consigliabile.

Più efficace è invece l'imputazione tramite regressione, detta anche *conditionl mean imputation* (Dagenais, 1973), che consiste nel rimpiazzare i dati mancanti con i valori predetti della variabile incompleta ottenuti dalla regressione sulle variabili osservate. Sebbene abbia prestazioni superiori rispetto all'imputazione della media, anche questo metodo porta ad una riduzione della variabilità della distribuzione dei dati, poiché i valori predetti vengono imputati senza tenere conto della variabilità residua. Si può ovviare a questo problema aggiungendo ai valori predetti una componente casuale residua che permetta di aumentare la variabilità. I valori imputati in questo modo possono essere considerati estrazioni da una distribuzione predittiva. Poiché i parametri di quest'ultima non sono noti ma stimati sulla base dei dati osservati, la procedura può essere ulteriormente migliorata includendo nelle imputazioni l'incertezza dovuta alla stima di tali parametri (van Buuren, 2018). I due principali metodi per fare ciò sono: utilizzare un approccio bayesiano, in cui i parametri sono estratti casualmente dalla loro distribuzione a posteriori, oppure ricorrere alla tecnica del *bootstrap* e stimare i parametri più volte dopo aver fatto un ricampionamento dei dati osservati (Heitjan e Little, 1991). In presenza di dati non normalmente distribuiti la distribuzione predittiva può essere opportunamente adattata, nello specifico utilizzando la classe di modelli lineari generalizzati per dati binari o categoriali (van Buuren, 2018).

Un'altra classe di metodi di imputazione è quella della *hot-deck imputation*, che consiste nel sostituire i valori mancanti di una o più variabili di un'unità, detta ricevente, con i valori osservati da un'altra unità con dati completi, detta donatore, il quale deve essere simile al ricevente rispetto alle caratteristiche osservate (Andridge e Little, 2010). La scelta del donatore può avvenire in diversi modi dando origine a diverse procedure. Si può, per esempio, scegliere una metrica per stabilire la distanza tra le unità e associa-

re ad ogni ricevente il donatore con distanza minore (*nearest neighbour hot deck*). Un'alternativa è invece quella di individuare un insieme di possibili donatori tra i quali scegliere casualmente quello effettivo. Comunemente si creano gruppi di 3, 5 o 10 possibili donatori che siano i più simili al ricevente, così da evitare che la stessa unità venga scelta molte volte come donatore effettivo. L'insieme di donatori può essere anch'esso scelto tramite l'utilizzo di una metrica, oppure classificando le unità in classi (*adjustment cells*) basate sui valori osservati per donatori e riceventi. Un particolare metodo che appartiene a questa classe è il *Predictive Mean Matching* (PMM), proposto per la prima volta da Rubin (1986) e Little (1988). Con questo metodo il donatore o il gruppo di possibili donatori viene scelto minimizzando la differenza tra i valori predetti dei rispondenti e quelli dei non rispondenti per una determinata variabile incompleta. I valori predetti vengono comunemente calcolati tramite una regressione di tale variabile rispetto a quelle osservate, ma in alcune varianti della procedura si utilizzano altri metodi come gli alberi di regressione e classificazione (CART) (per un'introduzione all'imputazione tramite CART si veda Saar-Tsechansky, 2007). Inoltre, come per l'imputazione tramite regressione, è possibile modificare l'algoritmo usando l'approccio bayesiano o la tecnica *bootstrap* per ottenere i valori predetti.

Il PMM, e più in generale l'*hot-deck imputation*, ha il vantaggio di imputare dati realmente osservati, evitando così imputazioni al di fuori del range osservato o valori che non rispettano la vera distribuzione dei dati. Inoltre, il PMM è un metodo robusto rispetto a trasformazioni della variabile incompleta. Si tratta poi di un metodo implicito, poiché l'algoritmo utilizzato sottende un modello per i dati mancanti che non ha bisogno di essere formalmente specificato (Little e Rubin, 2002). Per questo motivo è meno sensibile alla scorretta specificazione del modello rispetto ai metodi visti precedente-

mente (van Buuren, 2018). Un'altra caratteristica a suo favore è il fatto che si adatti sia all'imputazione univariata sia a quella multivariata, perché in quest'ultimo caso l'algoritmo procede imputando una variabile alla volta. Infine, questo metodo richiede che ci siano dei buoni appaiamenti tra riceventi e donatori, e ciò avviene più facilmente in dataset ampi piuttosto che piccoli, in cui potrebbe non avere buone prestazioni (Andridge e Little, 2010).

I metodi finora descritti sono generalmente validi per dati di tipo MCAR e MAR, ma lo stesso non si può dire per dati MNAR con meccanismo *non-ignorable*. In questo caso, infatti, il meccanismo dei dati mancanti deve essere incorporato nel modello di imputazione. Ci sono vari metodi per fare ciò, ma non essendo di interesse per questo studio si rimanda ad altri lavori per un approfondimento (van Buuren, 2018; Little e Rubin, 2002; Little, 2009).

Quale che sia la procedura scelta per eseguire l'imputazione singola, rimane un problema che può rendere l'inferenza non valida: utilizzando l'imputazione singola non viene considerato il fatto che i dati utilizzati per le analisi successive non sono stati veramente osservati, ma imputati. L'incertezza dell'imputazione viene ignorata, e ciò si riflette sugli errori standard che tendono ad essere sottostimati (Donders et al., 2006; Heijden et al., 2006). Per risolvere questo problema, come vedremo nel prossimo paragrafo, si può ricorrere all'imputazione multipla.

2.3.1 L'imputazione multipla

Per includere nelle analisi l'incertezza dovuta all'imputazione è stato proposto il metodo dell'imputazione multipla (Rubin, 1978; Rubin, 1987), in cui per ogni dato mancante vengono prodotti più valori, così da ottenere non solo uno, ma diversi dataset completi. Successivamente le analisi statistiche devono essere svolte su ciascuno di essi per ottenere le stime dei parametri

di interesse e delle relative varianze. La stima finale dei parametri si ottiene tramite la media aritmetica dei valori stimati per ogni dataset, mentre la combinazione delle varianze avviene considerando la variabilità delle stime tra i diversi dataset (van Buuren, 2018). In particolare, sia M il numero di nuovi dataset generati con l'imputazione, $\tilde{\beta}^{(j)}$ la stima del parametro di interesse β per il j -esimo dataset e $\widetilde{\mathbf{Var}}(\tilde{\beta}^{(j)})$ la stima della rispettiva varianza, con $j = 1, \dots, M$. La combinazione delle stime nei diversi dataset è pari a

$$\hat{\beta} = \frac{1}{M} \sum_{j=1}^M \tilde{\beta}^{(j)},$$

e la sua varianza è stimata come

$$\widehat{\mathbf{Var}}(\hat{\beta}) = \mathbf{Var}_W + \frac{M+1}{M} \mathbf{Var}_B,$$

dove le due quantità \mathbf{Var}_W e \mathbf{Var}_B rappresentano rispettivamente la varianza *within* e la varianza *between* e sono definite nel modo seguente:

$$\mathbf{Var}_W = \frac{1}{M} \sum_{j=1}^M \widetilde{\mathbf{Var}}(\tilde{\beta}^{(j)}),$$

$$\mathbf{Var}_B = \frac{1}{M-1} \sum_{j=1}^M \left(\tilde{\beta}^{(j)} - \hat{\beta} \right)^2.$$

La prima quantità spiega la variabilità delle stime all'interno dello stesso dataset, mentre la seconda considera la variabilità tra le stime nei diversi dataset generati tramite l'imputazione.

Per applicare l'imputazione multipla bisogna stabilire quanti valori si vogliono imputare per ogni dato mancante. Teoricamente, più alto è il numero di imputazioni e migliori sono i risultati della procedura. Tuttavia, anche un valore piccolo può portare a buoni risultati e il guadagno che si ottiene aumentando tale valore può non valere l'incremento dell'onere computazionale che ne consegue (van Buuren, 2018; Schafer, 1997; Schafer e Olsen, 1998).

Sulla base di questo ragionamento la letteratura individua 5 come un buon valore per il numero di imputazioni (van Buuren, 2018).

In generale, l'imputazione multipla è superiore all'imputazione singola. Infatti, anche se per dati MAR e MCAR entrambi i metodi portano a stime non distorte, come detto precedentemente con l'imputazione singola si sottostimano le varianze delle stime e pertanto l'inferenza risulta essere non valida.

Nel contesto dell'imputazione multipla, gli specifici metodi di imputazione possono essere scelti tra quelli già descritti per l'imputazione singola. Quando però l'imputazione è multivariata tali tecniche devono essere adattate. Per fare ciò esistono tre strategie principali (van Buuren, 2018). La prima, l'imputazione monotona, consiste nell'imputare i dati mancanti con le tecniche di imputazione univariata per una variabile alla volta in modo sequenziale (Rubin, 1987). Questa strategia è consigliata solo quando il pattern dei dati mancanti è monotono e l'imputazione deve essere fatta ordinando le variabili in modo crescente rispetto alla loro percentuale di valori mancanti. La seconda, invece, consiste nell'estrarre i valori da imputare da una opportuna distribuzione multivariata adattata ai dati. Questo metodo viene definito *joint modeling*. La terza, infine, è la tecnica di *fully conditional specification* (van Buuren et al., 2006; van Buuren, 2007) e comprende il noto algoritmo MICE (*Multiple Imputation by Chained Equations*). Si tratta di una generalizzazione dell'imputazione univariata in cui le variabili vengono imputate una alla volta definendo un modello di imputazione per ognuna di esse. Infatti, invece di utilizzare una distribuzione multivariata, si definiscono le distribuzioni condizionate rispetto alle altre variabili presenti nel modello di imputazione e si estraggono poi i valori da imputare da tali distribuzioni con un processo iterativo. Il principale pregio del *fully conditional*

specification è la sua flessibilità: specificando unicamente le distribuzioni condizionate si possono ottenere modelli multivariati al di fuori delle consuete densità. Inoltre permette l'utilizzo di metodi di imputazione che possono essere difficili da formulare in un modello multivariato (van Buuren, 2007). Poiché ogni distribuzione condizionata deve essere specificata separatamente, il *fully conditional specification* può richiedere molto sforzo per dataset con molte variabili, e, in generale, comporta un carico computazionale maggiore del *joint modeling*. In aggiunta, ci sono poche basi teoriche che dimostrano la qualità delle imputazioni tramite *fully conditional specification*, poiché, dopo aver definito la distribuzione condizionata per ciascuna variabile, non è garantito che esista una distribuzione congiunta teorica corrispondente (van Buuren et al., 2006). Tuttavia, la sua efficacia è stata mostrata empiricamente e rappresenta quindi una valida alternativa al *joint modeling* quando non è possibile individuare una distribuzione multivariata adeguata (van Buuren, 2007). Infatti molti studi di simulazione mostrano che questo metodo porta a stime corrette (Brand, 1999; Raghunathan et al., 2001; Brand et al., 2003; Tang et al., 2005; van Buuren et al., 2006; Horton e Kleinman, 2007; Yu, Burton e Rivero-Arias, 2007).

Multiple imputation by chained equations

Un particolare modo per eseguire l'imputazione multipla tramite *fully conditional specification* è con l'algoritmo MICE (van Buuren e Oudshoorn, 2000; van Buuren e Groothuis-Oudshoorn, 2011). Si tratta di una tipologia di *Markov chain Monte Carlo*, e più nel dettaglio di un *Gibbs sampler* (Gelfand e Smith, 1990; Casella e George, 1992), ovvero una tecnica di simulazione bayesiana che ha l'obiettivo di generare campioni da una determinata distribuzione multivariata utilizzando le distribuzioni condizionate. L'algoritmo è

costituito dai seguenti passi (van Buuren, 2018; Azur et al., 2011):

1. per ogni variabile incompleta si definisce un modello di imputazione costituito da una distribuzione predittiva condizionata ad altre variabili presenti nel dataset opportunamente scelte;
2. per ogni variabile incompleta si esegue un'imputazione iniziale con estrazioni casuali dai dati osservati;
3. si fissa una determinata variabile incompleta e si riportano i suoi dati alla condizione di mancanti;
4. si imputano i dati mancanti secondo il modello definito precedentemente;
5. si ripetono i passi 3 e 4 per ogni variabile incompleta presente nel dataset;
6. si ripetono i passi 3, 4 e 5 per un numero prestabilito di iterazioni.

A differenza di altri metodi di *Markov chain Monte Carlo*, questo algoritmo non necessita di molte iterazioni. La letteratura mostra che anche un numero limitato di iterazioni, compreso tra 5 e 20, può portare a buoni risultati (van Buuren et al., 2006; van Buuren, 2007; van Buuren, 2018).

Per quanto riguarda il modello di imputazione da definire per le variabili con dati mancanti, la scelta può ricadere tra i metodi già visti precedentemente. In particolare, una buona combinazione sembra essere quella tra MICE e PMM (Marshall et al., 2010; Marshall, Altman e Holder, 2010).

2.4 Utilizzo del propensity score in presenza di dati mancanti

La presenza di dati mancanti è piuttosto diffusa nel contesto degli studi osservazionali. Pertanto, quando si vuole stimare l'effetto di un trattamento, il problema dei dati mancanti e quello del confondimento possono convivere. Se ciò accade è necessario combinare le tecniche di aggiustamento tramite *propensity score* con i metodi per trattare i dati mancanti. Tutti i metodi descritti in questo capitolo possono essere utilizzati, ma non ci sono garanzie sul fatto che conducano a stime corrette dell'effetto del trattamento e che permettano un adeguato bilanciamento dei confondenti tra i soggetti trattati e quelli non trattati. Alcuni studi empirici sono stati realizzati per valutare l'efficacia di questi metodi. In particolare, si è riscontrato che i metodi del *missing indicator* e di imputazione della media portano a stime distorte dell'effetto del trattamento quando si utilizza il *propensity score matching* per tutti i meccanismi di dati mancanti (Choi, Dekkers e Cessie, 2019, Ling et al., 2020). L'utilizzo della *complete case analysis*, invece, può essere efficace in alcune circostanze. Secondo Ling et al. (2020) il metodo funziona quando l'effetto del trattamento medio sull'intera popolazione osservata è uguale a quello dei soggetti con dati mancanti, sia utilizzando il *propensity score matching* che IPTW. Questa circostanza, tuttavia, non può essere verificata ed è spesso poco realistica. Per meccanismi di dati mancanti più complessi, ma sempre di tipo *ignorable*, l'imputazione multipla sembra avere prestazioni migliori (Penning De Vries e Groenwold, 2017; Hill, 2004; Ling et al., 2020; Choi, Dekkers e Cessie, 2019). Poco si può dire invece per meccanismi *non-ignorable*.

In letteratura sono stati proposti due principali approcci per l'applica-

zione dell'imputazione multipla nel contesto dell'analisi tramite *propensity score*: l'approccio *within* e quello *across*. Nel primo approccio, dopo aver eseguito l'imputazione multipla ed aver ottenuto M dataset completi, si eseguono le analisi già descritte nel primo capitolo per ognuno di essi, ovvero si esegue la stima dei valori di *propensity score* per ogni soggetto e si applica uno dei quattro metodi basati su tali stime, ottenendo così M stime dell'effetto del trattamento. Successivamente, si possono combinare le varie stime tramite media aritmetica come presentato nel Paragrafo 2.3.1. Il secondo approccio, invece, prevede di stimare il *propensity score* per ogni soggetto negli m campioni e successivamente di ottenere un unico valore di *propensity score* per unità facendo la media rispetto ai diversi campioni di dati imputati. Infine, si ottiene la stima dell'effetto del trattamento basandosi sui *propensity score* medi precedentemente calcolati. Il confronto tra questi due approcci è oggetto di diversi studi: Hill (2004), Mitra e Reiter (2016), Penning De Vries e Groenwold (2017), Granger, Sergeant e Lunt (2019), Ling et al. (2020), Leyrat et al. (2019). Tra di essi solo Mitra e Reiter (2016) mostrano che l'approccio *across* comporta una minore distorsione rispetto all'approccio *within*, e ciò a discapito invece delle varianze, che risultano inferiori con quest'ultimo. Inoltre Penning De Vries e Groenwold (2017) mostrano come i risultati ottenuti da Mitra e Reiter siano dovuti alla violazione di una delle assunzioni alla base dei metodi basati sul *propensity score*, ovvero l'assunzione di positività della probabilità di ricevere il trattamento per tutti i soggetti, e al modello di imputazione utilizzato. L'approccio *within* risulta superiore negli studi condotti da Penning De Vries e Groenwold (2017), Granger, Sergeant e Lunt (2019) e Leyrat et al. (2019) per dati di tipo MAR e MCAR (che sono i meccanismi per cui l'imputazione multipla risulta valida). Inoltre, Leyrat e colleghi forniscono dimostrazioni teoriche a supporto del fatto che il *propen-*

sity score medio che si ottiene con il metodo *across* non è un vero *balancing score*, cosa che è invece dimostrata per i *propensity score* che si usano nei diversi dataset con l'approccio *within*. Lo stimatore che si ottiene quindi con quest'ultimo approccio risulta consistente, mentre lo stesso non vale per l'altro approccio. Negli altri studi non si notano differenze rilevanti tra i due approcci.

Capitolo 3

Una prima valutazione empirica

3.1 Introduzione

Come si è visto nei capitoli precedenti, la letteratura fornisce vari risultati sul confronto tra i metodi per trattare il confondimento basati sul *propensity score* e sul confronto tra diversi metodi per gestire il problema dei dati mancanti nel contesto dell'analisi tramite *propensity score*. Tuttavia, poco si sa su quali siano le migliori combinazioni tra i metodi basati sul *propensity score* e le tecniche per gestire i dati mancanti. Per questo motivo si è deciso di condurre uno studio di simulazione che permetta di fare un confronto empirico tra le varie combinazioni di cui sopra. Nell'impostare tale studio si è fin da subito notato che ci sono moltissimi elementi che possono influenzare la performance delle tecniche di aggiustamento basate sul *propensity score*, e che si possono riscontrare delle differenze sui risultati in base al ruolo ricoperto dalle variabili concomitanti e, di conseguenza, trarre delle conclusioni sbagliate. Alcuni degli aspetti che potrebbero condizionare i risultati sono:

- lo schema causale, ovvero le relazioni che intercorrono tra le variabili;

- l'effetto esercitato dalle covariate sul trattamento e sulla risposta, per esempio l'effetto può essere di tipo lineare, quadratico, con interazioni, ecc.;
- il meccanismo di generazione dei dati mancanti: i dati possono essere MCAR, MAR e MNAR e, negli ultimi due casi, il coinvolgimento delle variabili nel modello che descrive il meccanismo può variare;
- il numero di variabili con valori mancanti;
- le tipologie di variabili con valori mancanti, ovvero continue, discrete, dicotomiche, ecc.;
- il ruolo delle variabili con valori mancanti nello schema causale: una variabile incompleta può essere un confondente, un predittore della sola risposta, un predittore del solo trattamento, può essere il trattamento stesso o la risposta;
- il modello di imputazione: può essere correttamente specificato o no, e si può stabilire quali tra le variabili osservate intervengono nel processo di imputazione;
- la misura per valutare l'effetto del trattamento: ATT, ATE, differenza dei rischi, rapporto delle quote, ecc.

La soluzione ideale sarebbe considerare tutte le possibili alternative per ognuno degli aspetti elencati e combinarli tra loro per creare diversi schemi di simulazione e confrontare le prestazioni dei metodi di utilizzo del *propensity score* combinati con i metodi di gestione dei dati mancanti in ognuno di essi. Ciò comporterebbe però un lavoro enorme, difficilmente realizzabile in un unico studio. Si è deciso quindi di trovare un compromesso: su alcune delle

caratteristiche della simulazione si sono operate delle scelte, mentre su altre si sono considerate le diverse possibilità e si sono creati così diversi studi di simulazione.

L'obiettivo principale dello studio è confrontare le prestazioni delle possibili combinazioni tra i quattro metodi di utilizzo del *propensity score* descritti nel Capitolo 1 e le diverse modalità di applicazione dell'imputazione multipla. Infatti, come si è visto nel Capitolo 2, l'imputazione multipla è la tecnica più accreditata sia nel contesto dell'analisi tramite *propensity score* che in altri contesti.

3.2 Disegno della simulazione

Inizialmente si è deciso di valutare se la tipologia delle variabili che sono parzialmente osservate e il ruolo che ricoprono nello schema causale possono influenzare i risultati. Il modello di imputazione deve necessariamente subire delle modifiche a seconda che le variabili bersaglio siano continue, discrete o dicotomiche, e pertanto anche la bontà dell'imputazione può variare. Inoltre, la rilevanza delle covariate nelle analisi cambia in base al ruolo che ricoprono nello schema causale. In particolare ci si aspetta che se un confondente presenta valori mancanti si possano riscontrare più problematiche nelle stime rispetto al caso in cui la/e variabile/i parzialmente osservate siano predittori della sola risposta o del solo trattamento. Per verificare queste ipotesi si sono combinate queste due caratteristiche dei dati e si sono delineati 9 diversi scenari, ognuno con un relativo schema causale. Per non creare incomprensioni nelle spiegazioni successive, d'ora in poi si userà la parola "simulazione/i" per riferirsi ai 9 diversi scenari che costituiscono lo studio empirico, i cui dettagli vengono ora presentati.

La prima fase della simulazione è la generazione dei dati. Si è considerato un campione di numerosità 1000 costituito da 4 covariate $X = (X_1, X_2, X_3, X_4)$, di cui X_1 e X_2 sono continue, X_3 è dicotomica e X_4 è discreta ordinale. Le distribuzioni delle covariate sono:

- $X_1 \sim N(1, 4)$;
- $X_2 \sim N(5, 3)$;
- $X_3 \sim \text{Be}(0.6)$;
- X_4 assume i valori $(1, 2, 3, 4)$ con probabilità $(0.3, 0.2, 0.4, 0.1)$.

A partire dalle covariate si sono generate due variabili dicotomiche: la prima rappresenta il trattamento T e l'altra la risposta Y . Le relazioni tra le variabili nelle 9 simulazioni sono rappresentate in altrettanti diagrammi causali riportati nella Figura 3.1.

L'effetto delle covariate sul trattamento è stato definito secondo 4 scenari nel modo seguente:

Scenario A additivo e lineare;

Scenario B additivo e non lineare: alcuni predittori hanno un effetto quadratico;

Scenario C lineare e non additivo: si considera un'interazione tra alcuni predittori;

Scenario D non additivo e non lineare: si considerano sia effetti quadratici che interazioni.

Per la risposta, invece, si è previsto un effetto additivo e lineare delle covariate e del trattamento. Inizialmente si sono considerati un modello generale per il

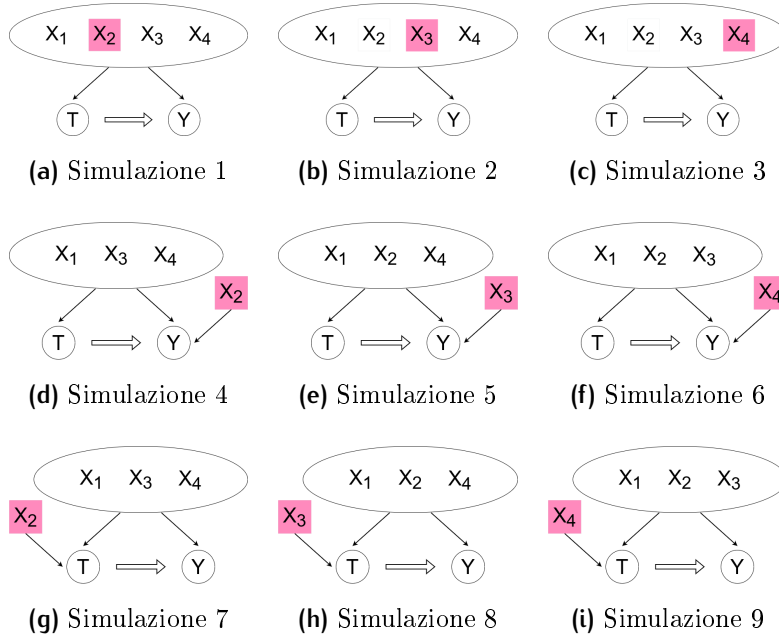


Figura 3.1: Diagrammi causali delle 9 simulazioni. Le variabili su sfondo rosso presentano valori mancanti.

trattamento per ognuno degli scenari e un modello per la risposta. Siano π_T e π_Y le probabilità che si verifichino gli eventi $T = 1$ e $Y = 1$ rispettivamente.

I modelli considerati per il trattamento secondo i 4 scenari sono:

Scenario A $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4,$

Scenario B $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_2^2 + \alpha_6 X_4^2,$

Scenario C $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_7 X_1 X_3 + \alpha_8 X_2 X_4,$

Scenario D $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_2^2 + \alpha_6 X_4^2 + \alpha_7 X_1 X_3 + \alpha_8 X_2 X_4,$

con $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8) = (0.8, -0.25, 0.6, 0.4, 0.05, -0.25, 0.4, -0.15).$

Il modello per la risposta è invece il seguente:

$$\text{logit}(\pi_Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \gamma T,$$

con $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (-1, 0.3, -0.25, 0.35, 0.5)$ e $\gamma = -0.7$.

Successivamente si sono attuate delle variazioni sui modelli per poter rispettare i diversi schemi causali. In particolare, alcuni parametri sono stati fissati a 0 nelle simulazioni in cui non sono presenti le relative variabili concomitanti e negli scenari in cui le variabili o i termini associati vengono considerati. Si mostra di seguito quali parametri sono stati annullati per ognuna delle simulazioni eseguite:

Simulazione 1 nessuno;

Simulazione 2 nessuno;

Simulazione 3 nessuno;

Simulazione 4 $\alpha_2 = \alpha_5 = \alpha_8 = 0$;

Simulazione 5 $\alpha_3 = \alpha_7$;

Simulazione 6 $\alpha_4 = \alpha_6 = \alpha_8 = 0$;

Simulazione 7 $\beta_2 = 0$;

Simulazione 8 $\beta_3 = 0$;

Simulazione 9 $\beta_4 = 0$.

Riassumendo, la generazione dei dati avviene simulando casualmente i valori delle X per ogni soggetto e generando poi i valori di T e Y secondo una distribuzione Bernoulli con parametri rispettivamente π_T e π_Y , i quali si ottengono dai modelli appena specificati. Le distribuzioni delle prevalenze di soggetti con $Y = 1$ e $T = 1$ nei diversi campioni generati per ognuna delle 9 simulazioni sono riportati nelle Figure 3.2 – 3.4.

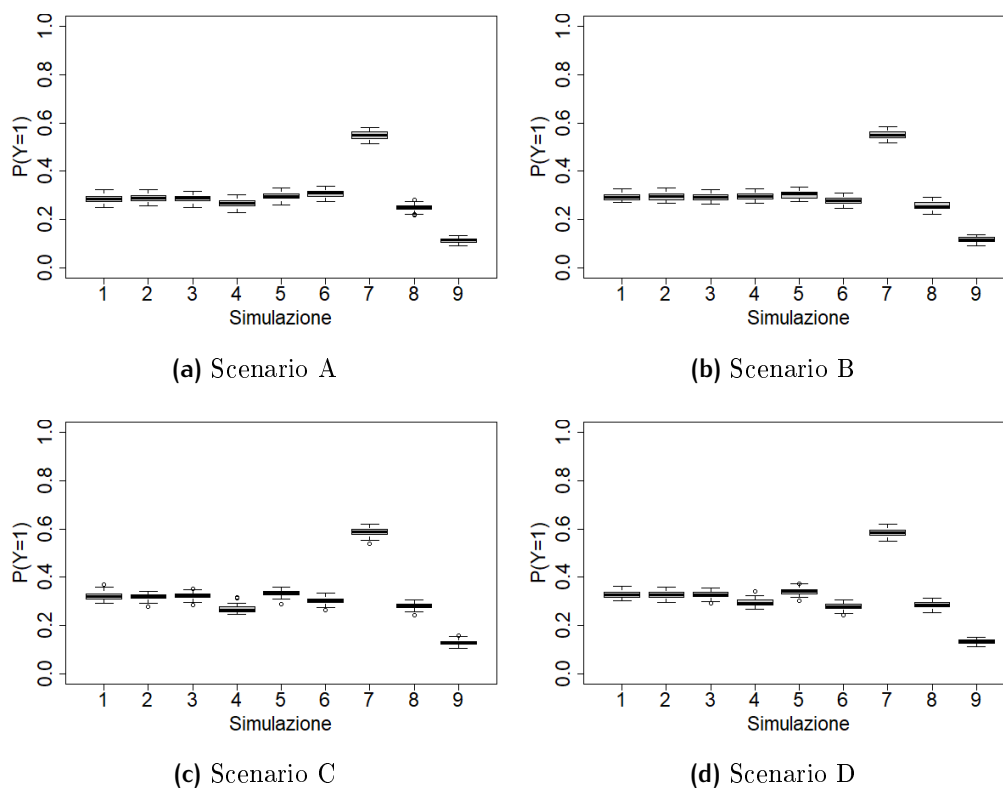


Figura 3.2: Distribuzione della prevalenza di soggetti con $Y = 1$ nei diversi campioni generati per le 9 simulazioni.

Nella seconda fase si sono eliminati dal dataset generato alcuni valori, così da creare un dataset con dati mancanti. Per i dati mancanti si è previsto un meccanismo di tipo MAR e si è utilizzato un modello tale per cui tutte le variabili incluse nel modello intervengono con lo stesso peso nel determinare se i valori della variabile che si vuole amputare siano mancanti o meno. Nel modello sono state considerate solo le esplicative, mentre le variabili trattamento e risposta sono state omesse, perché si è immaginato che esse non siano coinvolte nel meccanismo di generazione dei dati mancanti. La procedura di amputazione si basa sul calcolo di un punteggio definito dalla combinazione lineare dei valori osservati per le covariate complete. A questo

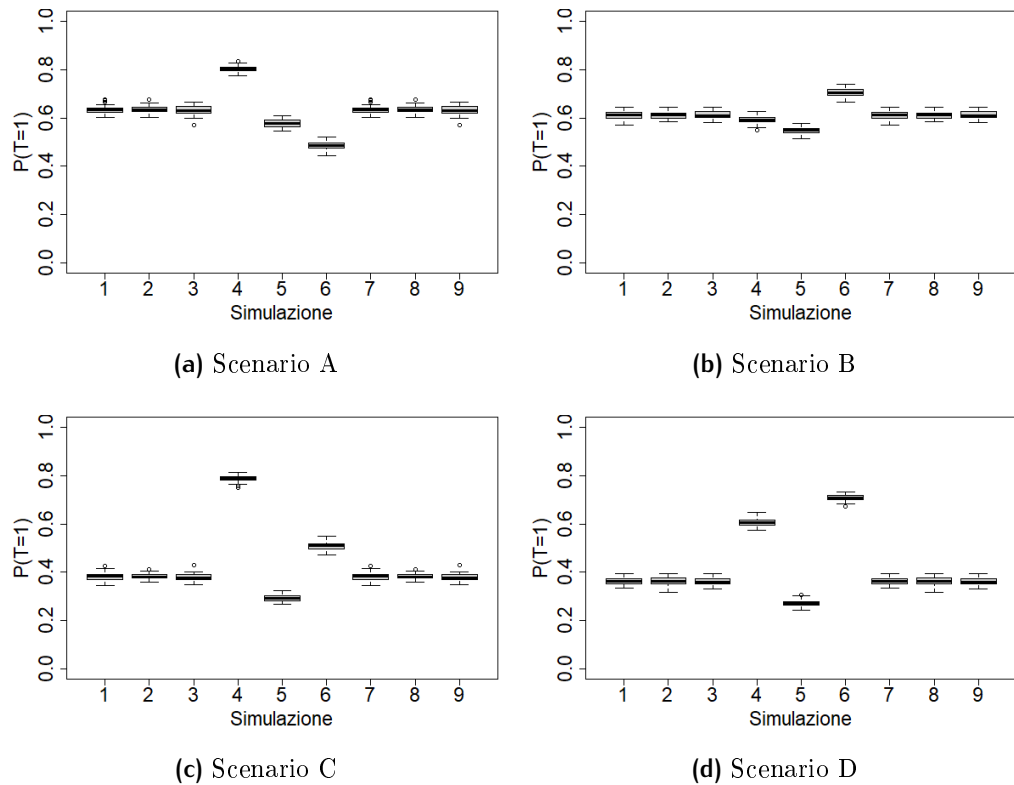
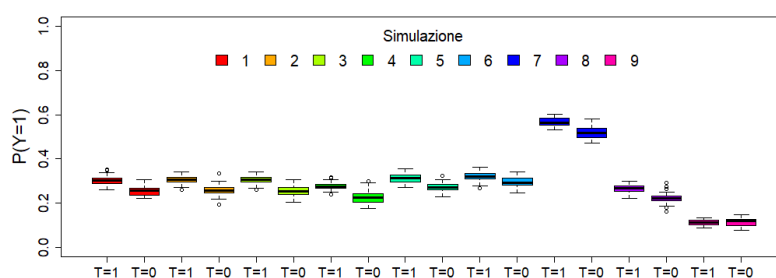
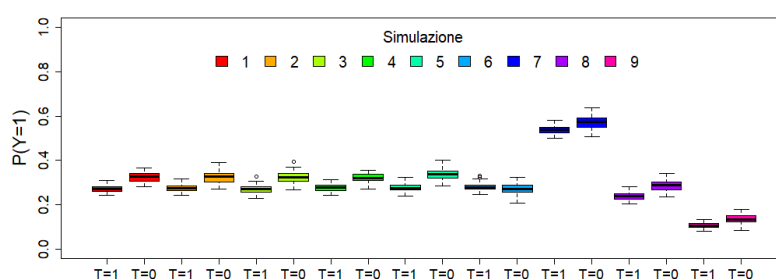


Figura 3.3: Distribuzione della prevalenza di soggetti con $T = 1$ nei diversi campioni generati per le 9 simulazioni.

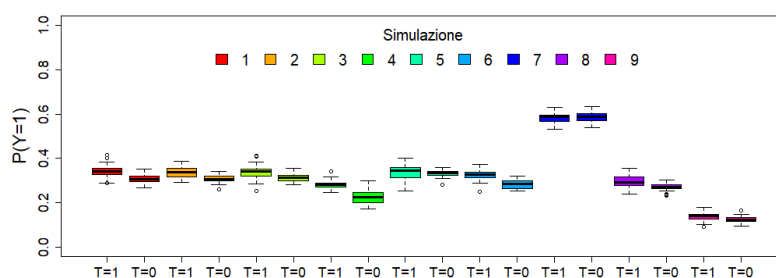
punteggio si applica poi una distribuzione di probabilità asimmetrica con coda allungata verso sinistra, ovvero tale per cui i casi con un punteggio elevato hanno una maggiore probabilità di essere mancanti. Per maggiori dettagli sulla procedura si rimanda a Schouten, Lugtig e Vink (2018, Capitolo 3), che descrive il meccanismo alla base della funzione 'ampute' (*ampute: Generate missing data for simulation purposes*) del pacchetto 'mice' del software R, che è stata utilizzata in questa fase dello studio. La proporzione di dati mancanti è 0.4 ed è da intendersi rispetto al numero di osservazioni complessive, ovvero il 40% delle osservazioni ha almeno un dato mancante. Come si è mostrato nei diagrammi causali (Figura 3.1), si è amputata una



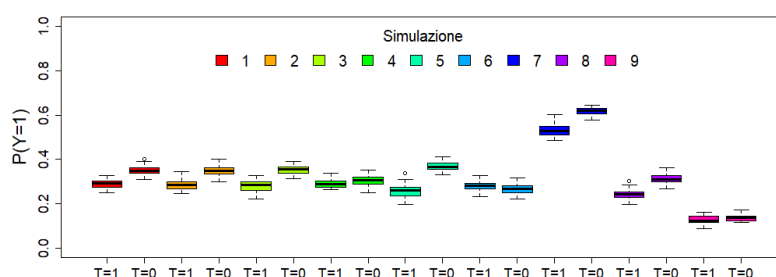
(a) Scenario A



(b) Scenario B



(c) Scenario C



(d) Scenario D

Figura 3.4: Distribuzione della prevalenza di soggetti con $Y = 1$ nei diversi campioni generati nei sottogruppi dei trattati ($T = 1$) e dei non trattati ($T = 0$) per le 9 simulazioni.

sola variabile per ogni simulazione. Di conseguenza il 40% dei valori di tale variabile risulta mancante.

3.3 Trattamento dei dati mancanti

La fase successiva è quella del trattamento dei dati mancanti. Si è scelto come metodo principale quello dell'imputazione multipla e si è utilizzato l'approccio *within* (vedasi spiegazione a pagina 41) per le analisi successive. Si è eseguita la tecnica MICE con tre distinti metodi di imputazione. Due di essi sono basati sul PMM, ma applicati con algoritmi diversi. In particolare, oltre alla procedura comune si è utilizzata una variante detta *Weighted Predictive Mean Matching* (WPMM). Il terzo metodo è l'imputazione tramite regressione, che consiste in una regressione lineare con approccio bayesiano per la variabile continua, una regressione logistica con approccio bayesiano per quella dicotomica e un modello a quote proporzionali per quella ordinale. Oltre a questi metodi si è eseguita anche l'imputazione singola della media, essendo questo un metodo molto utilizzato. Per eseguire le procedure si è utilizzata la funzione 'mice' dell'omonimo pacchetto del software R con i metodi 'pmm', 'midastouch', 'norm', 'logreg', 'polr' e 'mean'. Per i dettagli sulle procedure si veda van Buuren e Groothuis-Oudshoorn (2011) e *Package 'mice'* (2021). Nell'imputazione multipla sono stati generati 5 dataset completi ed eseguite 20 iterazioni. Nel modello di imputazione sono state inserite tutte le covariate, mentre il trattamento e l'outcome sono stati esclusi. Pertanto, il modello risulta essere correttamente specificato.

Oltre alle tecniche basate sull'imputazione si è voluto eseguire la *complete case analysis* (CCA) e, inoltre, le analisi sono state svolte anche sul campione privo di dati mancanti, ovvero prima di eseguire l'amputazione. In questo

modo si può valutare quanto la presenza di dati mancanti influisca sulle stime e l'effetto dell'imputazione sui risultati della simulazione.

3.4 Analisi dei dati

L'ultima fase riguarda le analisi vere e proprie dei dati dopo aver eseguito l'imputazione. Queste vengono ripetute per i 4 scenari considerati e per ognuno dei metodi di imputazione utilizzati.

Per prima cosa si sono stimati i *propensity score* per ogni individuo utilizzando la regressione logistica. Si è voluto simulare quanto più possibile un'analisi reale, in cui non si conosce quali siano le variabili che hanno effetto sul trattamento, né la forma di tale effetto. Pertanto, indipendentemente dallo scenario, il modello logistico adattato comprende tutte le covariate X , che sono inserite con un effetto lineare. In questo modo si possono valutare gli effetti di una scorretta specificazione del modello di stima del *propensity score*.

Successivamente si è stimato l'effetto del trattamento utilizzando come misura d'effetto l'ATT e si sono calcolati i relativi errori standard. Si sono applicati i 4 metodi di aggiustamento del confondimento basati sul *propensity score* con le metodologie descritte nel Paragrafo 1.3. In particolare, si sono eseguiti:

- *matching* uno a uno con ripetizione utilizzando la tecnica *nearest neighbour* con una tolleranza di 0.2 utilizzando la funzione 'Match' del pacchetto 'Matching' di R (Sekhon, 2011; Package 'Matching' 2022);
- stratificazione con 5 strati ottenuti tramite i quintili dei *propensity score* stimati;

- *covariate adjustment* con le specifiche già descritte nel Paragrafo 1.3.3;
- IPTW utilizzando lo stimatore (1.4).

Per ogni simulazione vengono eseguite 50 repliche. Così come per la generazione iniziale dei dati, anche la procedura di amputazione dei dati ha una componente aleatoria che genera variabilità nei risultati. Per poter distinguere nei risultati la variabilità dovuta ai diversi campioni generati da quella dovuta alle diverse amputazioni che possono essere fatte, per ogni dataset simulato si sono generati 10 diversi dataset amputati e le analisi sono state eseguite su ognuno di essi per ogni replica. In questo modo per ognuna delle 50 repliche eseguite si sono ottenute 10 stime dell'ATT per ogni combinazione di metodi di imputazione e di aggiustamento del confondimento utilizzati e per ogni scenario, e altrettanti errori standard. Al fine di valutare la bontà delle stime si sono calcolati i veri valori dell'ATT sul campione simulato prima di eseguire l'amputazione. Poiché tali valori dipendono dal dataset generato, i valori di ATT differiscono per ogni replica e per ogni scenario. Si sono ottenuti quindi 4 valori di ATT per ognuna delle 50 repliche.

3.5 Risultati

Per ogni combinazione di schema causale, scenario, metodo di trattamento del confondimento e metodo di gestione dei dati mancanti, siano ATT_r i veri valori di ATT, $\widehat{ATT}_{r,i}$ i valori di ATT stimati e $ES_{r,i}$ i relativi errori standard, con $r = 1, \dots, 50$ e $i = 1, \dots, 10$. Per la valutazione dei risultati si sono calcolate le seguenti misure:

- la media dei veri valori di ATT rispetto alle 50 repliche

$$\overline{ATT}_{veri} = \frac{1}{50} \sum_{r=1}^{50} ATT_r;$$

- la media delle stime dell'ATT rispetto alle 50 repliche e alle 10 procedure di amputazione fatte per ognuna di esse

$$\overline{\text{ATT}}_{\text{stimati}} = \frac{1}{50} \sum_{r=1}^{50} \left(\frac{1}{10} \sum_{i=1}^{10} \widehat{\text{ATT}}_{r,i} \right)$$

(nel caso delle analisi eseguite sui dati non amputati la media è fatta solo rispetto alle 50 repliche);

- la media dei relativi errori standard ottenuta nello stesso modo

$$\overline{\text{ES}} = \frac{1}{50} \sum_{r=1}^{50} \left(\frac{1}{10} \sum_{i=1}^{10} \widehat{\text{ES}}_{r,i} \right);$$

- la media delle distorsioni relative

$$\overline{\text{Distorsione}}_{rel} = \frac{1}{50} \sum_{r=1}^{50} \left(\frac{\frac{1}{10} \sum_{i=1}^{10} \widehat{\text{ATT}}_{r,i} - \text{ATT}_r}{\text{ATT}_r} \right);$$

- la media degli errori quadratici medi

$$\overline{\text{EQM}} = \frac{1}{50} \sum_{r=1}^{50} \sqrt{\frac{1}{10} \sum_{i=1}^{10} (\widehat{\text{ATT}}_{r,i} - \text{ATT}_r)^2}.$$

Queste quantità sono riportate nelle Tabelle 3.7 – 3.15.

Si segnala che in alcune circostanze il metodo di trattamento del confondimento basato sulla stratificazione ha generato problemi nella procedura di stima. Alcuni degli strati che si ottenevano con tale metodo erano composti solo da soggetti non trattati o solo soggetti trattati. Ciò ha impedito di calcolare la stima dell'ATT per tale metodo nelle circostanze in cui il problema si è verificato. Quando ciò è avvenuto solo per alcune delle 10 repliche dell'amputazione all'interno di uno stesso campione generato, è stato comunque possibile ottenere la stima complessiva dell'ATT per tale campione ignorando le repliche i tali per cui non è stato possibile ottenere la stima $\widehat{\text{ATT}}_{r,i}$. Al

contrario, per le ripetizioni in cui il problema è insorto per tutti i 10 campioni amputati non si è potuto ottenere la stima dell'ATT corrispondente, e i valori $\overline{ATT}_{stimati}$ sono stati calcolati ignorando tali ripetizioni. Il problema potrebbe essere risolto riducendo il numero degli strati, ma non ci sono garanzie che questo basti. Ciò che è avvenuto denota il fatto che non sempre la stratificazione è l'approccio giusto per trattare il confondimento. Nel presente studio, in particolare, il problema si è verificato negli Scenari A e C della Simulazione 4, nello Scenario D della Simulazione 6, in generale negli Scenari C e D quando si è utilizzata la *complete case analysis* e in poche altre circostanze. Nel caso delle Simulazioni 4 e 6, il problema è spiegato dal fatto che la probabilità di essere trattati risulta abbastanza elevata (si veda la Tabella 3.3). Negli altri casi, tuttavia, tale probabilità ha valori moderati. Il problema sembra perciò essere dovuto alla combinazione tra il metodo della stratificazione e la *complete case analysis*. Nella Tabella 3.1 è riportato il numero esatto di repliche in cui si è verificato questo problema per ogni metodo, scenario e simulazione.

Le stime dell'ATT ottenute per ogni replica sono caratterizzate da diverse componenti di variabilità:

- quella intrinseca ai dati;
- quella dovuta all'amputazione dei dati;
- quella dovuta all'imputazione dei valori mancanti;
- l'incertezza della stima.

La variabilità campionaria può essere valutata osservando la dispersione dei veri valori di ATT calcolati nei 50 dataset generati. A tal fine si riportano i coefficienti di variazione dei valori ATT_r nella Tabella 3.1. Come si può

Tabella 3.1: Numero di repliche della stratificazione in cui in almeno uno strato non erano presenti soggetti trattati o non trattati. Le repliche totali sono 500 in tutti i casi ad eccezione dell'analisi sui dati non amputati (dati completi). Qualora avvenga, si riporta tra parentesi il numero di ripetizioni su 50 in cui il problema si verifica in tutti i 10 dataset amputati corrispondenti.

Scenario	Gestione dei dati mancanti	Simulazione								
		1	2	3	4	5	6	7	8	9
A	Dati completi	0	0	0	3	0	0	0	0	0
	CCA	0	1	0	28 (3)	0	0	0	1	0
	Media	0	0	0	30 (3)	0	0	0	0	0
	PMM	0	0	0	38 (3)	0	0	0	0	0
	WPMM	0	0	0	36 (3)	0	0	0	0	0
	Regressione	0	0	0	37 (3)	0	0	0	0	0
B	Dati completi	0	0	0	0	0	0	0	0	0
	CCA	0	0	0	0	0	0	0	0	0
	Media	0	0	0	0	0	0	0	0	0
	PMM	0	0	0	0	0	0	0	0	0
	WPMM	0	0	0	0	0	0	0	0	0
	Regressione	0	0	0	0	0	0	0	0	0
C	Dati completi	0	0	0	12	0	0	0	0	0
	CCA	10	7	10	70	18	3	10	13	10
	Media	1	0	0	118 (11)	0	0	1	0	0
	PMM	0	0	0	129 (12)	0	0	0	0	0
	WPMM	0	0	0	125 (10)	0	0	0	0	0
	Regressione	0	0	0	123 (11)	0	0	0	0	0
D	Dati completi	0	0	0	0	0	1	0	0	0
	CCA	9	13	9	0	7	3	9	7	9
	Media	0	0	0	0	0	10 (1)	0	0	0
	PMM	0	0	0	0	0	13 (1)	0	0	0
	WPMM	0	0	0	0	0	11 (1)	0	0	0
	Regressione	0	0	0	0	0	12 (1)	0	0	0

Tabella 3.2: Coefficienti di variazione dei veri valori di ATT nelle 50 repliche per ogni simulazione ed ogni scenario.

Scenario	Simulazione								
	1	2	3	4	5	6	7	8	9
A	0.009	0.008	0.008	0.009	0.008	0.010	0.009	0.009	0.016
B	0.011	0.011	0.012	0.010	0.011	0.009	0.009	0.012	0.021
C	0.009	0.009	0.010	0.009	0.012	0.009	0.014	0.011	0.017
D	0.016	0.013	0.015	0.010	0.016	0.009	0.013	0.016	0.023

osservare, i valori sono tutti al di sotto dello 0.023, pertanto si può concludere che l'entità della variabilità campionaria sia esigua. Inoltre, per avere un'idea di quanto quest'ultima incida sulle stime dell'ATT, si è calcolato il rapporto tra la deviazione standard dei veri ATT rispetto alle 50 repliche e quella dei valori $\overline{\text{ATT}}_r = \frac{1}{10} \sum_{i=1}^{10} \widehat{\text{ATT}}_{r,i}$ per tutte le combinazioni di metodi utilizzati e per ogni scenario e simulazione. L'intervallo di variazione dei valori ottenuti per tale rapporto nelle diverse situazioni è (0.0016, 0.0724). Ciò conferma il fatto la variabilità delle stime è dovuta solo in minima parte alla variabilità della quantità stessa che deve essere stimata.

Per quanto riguarda la variabilità dovuta all'amputazione dei dati, questa è stata controllata grazie al fatto che per ogni replica sono stati generati 10 diversi dataset incompleti. Per distinguere le due componenti della variabilità rimanenti, ovvero quella data dall'imputazione e quella generata dall'incertezza della stima, si utilizzano dei grafici di dispersione che confrontano le medie delle stime rispetto alle 10 ripetizioni della procedura di amputazione, cioè i valori $\overline{\text{ATT}}_r = \frac{1}{10} \sum_{i=1}^{10} \widehat{\text{ATT}}_{r,i}$ per $r = 1, \dots, 50$, con le stime degli ATT ottenute dai dataset completi prima di eseguire l'amputazione. Ci si

aspetta che la dispersione dei valori \overline{ATT}_r ottenuti quando si è utilizzata l'imputazione multipla sia paragonabile a quella delle stime ottenute sui dataset non amputati, poiché la variabilità dovuta all'imputazione è regolata dalla procedura stessa di imputazione multipla, in cui le stime finali sono ottenute dalla media delle stime ottenute su 5 diversi dataset imputati. Inoltre, al fine di valutare le prestazioni della *complete case analysis* e di confrontarle con quelle dell'imputazione, si sono prodotti altri grafici di dispersione che confrontano le stime degli ATT ottenute con la *complete case analysis* con le stime prodotte dall'analisi dei dati non amputati e con quelle prodotte dopo aver eseguito l'imputazione. I grafici ottenuti per lo Scenario A della prima simulazione sono visibili nelle Figure 3.5 – 3.8. Gli altri sono riportati in Appendice A.

3.6 Discussione

Dai risultati mostrati si può osservare che non ci sono delle specifiche combinazioni che funzionano nettamente meglio o peggio in tutte le circostanze. Inoltre le prestazioni dei diversi metodi non sempre seguono uno schema preciso rispetto allo scenario o ad altre caratteristiche delle simulazioni eseguite.

In generale, le differenze maggiori tra i risultati si riscontrano al variare del metodo di utilizzo del *propensity score* più che al variare del metodo di gestione dei dati mancanti. Da questo possiamo dedurre che la scelta del metodo per trattare il confondimento ha una maggiore rilevanza. Ciò non è del tutto sorprendente se si pensa che l'ipotesi di meccanismo MAR su cui si basano i metodi di imputazione multipla è rispettata e, in aggiunta, nello studio è stata prevista una sola variabile con dati mancanti alla volta. Per-

tanto ci sono i presupposti per ottenere buoni risultati tramite l'imputazione multipla, indipendentemente dalla tecnica con cui viene applicata.

Complessivamente si osserva che la stratificazione tende a produrre risultati peggiori in termini di distorsione relativa, mentre il *covariate adjustment* spesso è il metodo con distorsione inferiore. Tuttavia ciò non può essere generalizzato a tutte le circostanze. Per esempio, nello Scenario C della terza simulazione avviene esattamente l'opposto. Per quanto riguarda invece la dispersione delle stime, le differenze tra i metodi di aggiustamento del confondimento sono costanti al variare dello schema di simulazione e dello scenario. Anche secondo questo parametro di valutazione il *covariate adjustment* sembra essere il metodo migliore. Ciò si evince dai grafici di dispersione e dal fatto che quasi sempre il valore della media degli errori quadratici medi risulta inferiore rispetto agli altri metodi. In particolare, il metodo del *matching* è quello con dispersione maggiore. Ciò è ben visibile nei grafici di dispersione, in cui le nuvole di punti sono molto più sparse rispetto agli altri due metodi, ma anche dai valori dell'errore quadratico medio. Dai grafici di dispersione si può notare, inoltre, una frequente asimmetria delle stime verso sinistra quando si utilizza il metodo dell'IPTW, e, in aggiunta, si osserva in alcuni casi la presenza di valori anomali molti distanti dai veri valori di ATT. Alcuni esempi di quanto detto sono lo Scenario C della Simulazione 4, in cui la distorsione risulta molto elevata a causa di una stima particolarmente inferiore al vero valore di ATT, e lo Scenario D della Simulazione 8. Ciò si verifica anche nelle analisi eseguite sui campioni non amputati. Si osserva, inoltre, che nello Scenario C di tutte le simulazioni, ad eccezione della prima, quando si utilizza l'IPTW in combinazione con la *complete case analysis* l'errore quadratico medio risulta il più elevato di tutte le possibili combinazioni tra i diversi metodi.

Le osservazioni fatte finora sono valide per tutti i metodi utilizzati per gestire i dati mancanti. In particolare, le diverse tecniche di imputazione multipla (PMM, WPMM e l'imputazione tramite regressione) hanno quasi sempre prestazioni molto simili se combinate con uno stesso metodo di trattamento del confondimento, sia in termini di distorsione che di dispersione. Utilizzare, invece, la tecnica dell'imputazione singola della media comporta un lieve aumento della distorsione delle stime in alcuni casi, e risultati analoghi a quelli ottenuti con l'imputazione multipla in altri, come è visibile dai valori dell'errore quadratico medio. Spesso risulta in errori standard inferiori rispetto agli altri metodi, e ciò è in linea con le caratteristiche di questo metodo.

I risultati ottenuti con la *complete case analysis*, invece, sono variabili nei diversi scenari e nelle diverse simulazioni, e non sembra esserci una precisa logica sottostante. Spesso questo metodo mostra risultati in contrasto con gli altri in termini di distorsione relativa. In alcuni casi, quando la distorsione risulta molto alta con tutti i metodi di trattamento dei dati mancanti, la *complete case analysis* ha risultati nettamente superiori. Ciò avviene, per esempio, nella Simulazione 3, per quasi tutti gli scenari e metodi di *propensity score* utilizzati, oppure negli Scenari A e B della Simulazione 1. In effetti, queste fanno parte delle simulazioni con le distorsioni più alte per la maggior parte dei metodi. In altri casi, invece, in cui i metodi di imputazione multipla sono molto efficaci, la distorsione risulta maggiore per la *complete case analysis*, come avviene nello Scenario D delle Simulazioni 5 e 6. Tuttavia questo comportamento non è sempre uniforme per i diversi metodi di *propensity score* utilizzati. Si veda, per esempio, lo Scenario D della Simulazione 8 o lo Scenario C della Simulazione 7. In sintesi, la distorsione delle stime prodotte con il metodo della *complete case analysis* è difficilmente prevedibile. Gli er-

rori quadratici medi delle stime ottenute con quest'ultima, invece, sono quasi sempre superiori, o al più simili, a quelli relativi ai metodi di imputazione.

Si discute ora la variabilità delle stime, cercando di valutare quanto di essa dipenda dalla presenza di dati mancanti grazie all'osservazione dei grafici nelle Figure 3.5 – 3.8 e A.1 – A.140. Come anticipato nel paragrafo precedente, la variabilità incide poco sulla dispersione dei veri valori di ATT, evidenziando il fatto che la variabilità delle stime dipende in misura nettamente maggiore dal processo stesso di stima. Confrontando la dispersione dei valori di ATT stimati dopo l'imputazione con quella degli ATT stimati sui dataset completi prima di eseguire l'amputazione, si osserva che esse sono confrontabili, e ciò sta ad indicare che la variabilità causata dal processo di stima è superiore a quella generata dall'imputazione. Infatti, come detto precedentemente, il metodo di trattamento del confondimento ha più rilevanza rispetto a quello per trattare i dati mancanti. In generale, si osserva un maggiore accordo tra le stime ottenute prima dell'amputazione con quelle ottenute con i metodi di *covariate adjustment* e IPTW che con quelle ottenute con gli altri due metodi basati sul *propensity score*. Questo ad eccezione delle prime tre simulazioni, in cui questi due metodi mostrano per alcuni scenari una distorsione sistematica. Facendo un confronto con gli altri metodi, tuttavia, si nota che in alcuni casi ciò è dovuto al fatto che la distorsione delle stime ottenute con le analisi eseguite prima dell'amputazione è inferiore per i metodi di *covariate adjustment* e IPTW rispetto a *matching* e stratificazione. Alcuni esempio sono lo Scenario C della Simulazione 1 e gli Scenari A e B della Simulazione 3.

Si osserva poi che, in alcune circostanze, la distorsione delle stime è molto elevata anche per le analisi eseguite in assenza di dati mancanti e, a volte, anche superiore a quella ottenuta con le analisi eseguite dopo l'imputazione dei dati. Pertanto sembra che l'imputazione sia efficace nel risolvere il proble-

ma dei dati mancanti e che sia la presenza di variabili confondenti a causare maggior distorsione nelle stime. Si vedano, per esempio, lo Scenario A della Simulazione 2 con il metodo del *covariate adjustment*, lo Scenario C della Simulazione 7 con il metodo del *matching* e la Simulazione 8 in quasi tutti gli scenari e metodi utilizzati.

Per quanto riguarda le differenze tra le 9 simulazioni, si osservano generalmente prestazioni inferiori in termini di distorsione nelle prime tre simulazioni, ovvero quando i dati mancanti sono in una variabile confondente. Spesso ciò non dipende dai metodi utilizzati per trattare il confondimento o i dati mancanti. Questo risultato, come ci si aspettava, indica che il ruolo delle variabili con dati mancanti influisce sulle stime e, in particolare, i risultati peggiorano se tali variabili sono dei confondenti. Nelle altre simulazioni i risultati ottenuti sui campioni non ancora amputati sono spesso molto simili a quelli che si ottengono considerando il problema dei dati mancanti. Sembra quindi che l'effetto della presenza di dati mancanti in variabili che sono predittori unicamente del trattamento o della risposta sia nullo o minimo. Tuttavia le stesse differenze non sono evidenti per quanto riguarda la dispersione e, in particolare, confrontando i valori dell'errore quadratico medio. La tipologia della variabile con dati mancanti, invece, sembra avere meno effetto sui risultati. In particolare, confrontando le ultime sei simulazioni non sembrano esserci delle differenze evidenti a seconda del fatto che la variabile incompleta sia continua, dicotomica o discreta. Come detto, però, in queste simulazioni la presenza dei dati mancanti non sembra influire sui risultati. Se invece si considerano le prime tre simulazioni, complessivamente la distorsione è inferiore nella Simulazione 2, in cui la variabile amputata è dicotomica. Anche questa considerazione, però, non vale per tutti gli scenari e tutti i metodi considerati.

Sulla base di quanto osservato nei risultati si sono prodotti degli schemi riassuntivi con le combinazioni tra i metodi di trattamento del confondimento e le tecniche di gestione dei dati mancanti migliori e peggiori per ognuna delle simulazioni eseguite e ognuno degli scenari considerati. Tali schemi sono riportati nelle Tabelle 3.3 – 3.6. Si può concludere che, nella maggior parte dei casi, la scelta migliore per trattare il confondimento sia utilizzare il metodo di *covariate adjustment*. Sulla scelta del metodo di gestione dei dati mancanti è più difficile fornire delle linee guida. Se ci si basa sulla distorsione relativa la combinazione migliore sembra essere con la *complete case analysis* quando i dati mancanti sono in variabili confondenti e con l'imputazione multipla negli altri. Tuttavia, negli studi reali non è possibile stabilire se quest'ultima condizione sia verificata o meno. Poiché però le differenze che la scelta del metodo di gestione dei dati mancanti generano nei risultati sono maggiori quando i dati mancanti sono in variabili confondenti rispetto agli altri casi, la *complete case analysis* può essere preferibile in generale. Basandosi invece sull'errore quadratico medio, l'imputazione multipla risulta il metodo con prestazioni migliori. Le altre caratteristiche analizzate sembrano avere meno rilevanza nel complesso.

3.7 Un esempio di caso di studio reale

L'indagine empirica è stata strutturata in modo tale da poter indagare le prestazioni dei diversi metodi utilizzati al variare di alcune caratteristiche delle variabili che devono essere analizzate. Lo scopo di questo tipo di studi è quello di fornire delle linee guida per l'utilizzo dei metodi di trattamento del confondimento e di gestione dei dati mancanti nella pratica statistica. Pertanto, si propone un esempio di indagine statistica su dati reali che si

adatti allo schema utilizzato per le simulazioni.

Si consideri uno studio osservazionale retrospettivo in cui si vuole valutare che effetto abbia offrire assistenza per smettere di fumare sulla mortalità in pazienti che hanno avuto un infarto miocardico acuto. La popolazione di riferimento è costituita da soggetti ricoverati per infarto miocardico acuto che sono sopravvissuti fino alla dimissione e per i quali si può testimoniare che al momento fossero fumatori. La variabile trattamento T vale 1 per i soggetti che hanno ricevuto assistenza per smettere di fumare e 0 per gli altri. La variabile risposta Y rappresenta la mortalità entro i tre anni successivi alla dimissione. La variabile X_1 rappresenta la frequenza respiratoria misurata al momento del ricovero, ed è una variabile confondente. Le altre variabili, a seconda dei casi considerati, potrebbero essere sia confondenti che predittori unicamente della risposta o del trattamento. Tuttavia, come spiegato, si immagina di effettuare un'analisi in cui questa informazione non sia nota a priori. Si prendono quindi come esempio delle covariate che siano potenziali confondenti ma che potrebbero rivelarsi non essere determinanti sia del trattamento che della risposta. Le variabili concomitanti considerate sono quindi:

X_2 frequenza cardiaca al momento del ricovero;

X_3 familiarità di malattie coronariche;

X_4 numero di patologie che rappresentano fattori di rischio per l'infarto (cancro, scompenso cardiaco cronico, attacco ischemico transitorio, arteriopatia obliterante periferica).

Per ideare questo esempio si è preso ispirazione dallo studio su cui si basa l'articolo di Austin (2011a).

Tabella 3.3: Schema riassuntivo delle migliori e peggiori combinazioni di metodi di trattamento del confondimento e di gestione dei dati mancanti nelle 9 simulazioni per lo Scenario A.

Variabile incompleta	continua		dicotomica		discreta		
	migliore	peggiore	migliore	peggiore	migliore	peggiore	
confondente	$\overline{\text{Distorsione}}_{rel}$	IPTW CCA	Stratificazione media	Cov. adjustment IM	Matching media	IPTW CCA	Stratificazione IM
	$\overline{\text{EQM}}$	Cov. adjustment media	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment media	Matching CCA
predittore di Y	$\overline{\text{Distorsione}}_{rel}$	Cov. adjustment IM	Stratificazione CCA	IPTW IM	Stratificazione IM	Cov. adjustment IM	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA
predittore di T	$\overline{\text{Distorsione}}_{rel}$	Matching WPMM	Cov. Adjustment CCA	Cov. Adjustment CCA	Stratificazione CCA	Matching WPMM	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA

IM: Imputazione multipla (tutti i metodi). Imputazione: tutti i metodi di imputazione singola e multipla

Tabella 3.4: Schema riassuntivo delle migliori e peggiori combinazioni di metodi di trattamento del confondimento e di gestione dei dati mancanti nelle 9 simulazioni per lo Scenario B.

Variabile incompleta	continua		dicotomica		discreta		
	migliore	peggiore	migliore	peggiore	migliore	peggiore	
confondente	$\overline{\text{Distorsione}}_{rel}$	Matching CCA	IPTW IM	IPTW IM	Stratificazione CCA	IPTW IM	IPTW IM
	$\overline{\text{EQM}}$	Cov. adjustment media	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment media e CCA	Matching CCA
predittore di Y	$\overline{\text{Distorsione}}_{rel}$	Cov. adjustment CCA	IPTW IM	Cov. adjustment CCA	IPTW IM	Cov. adjustment IM	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA
predittore di T	$\overline{\text{Distorsione}}_{rel}$	IPTW IM	Stratificazione media	Matching media	Stratificazione CCA	Matching regressione	Stratificazione media
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA

IM: Imputazione multipla (tutti i metodi). Imputazione: tutti i metodi di imputazione singola e multipla

Tabella 3.5: Schema riassuntivo delle migliori e peggiori combinazioni di metodi di trattamento del confondimento e di gestione dei dati mancanti nelle 9 simulazioni per lo Scenario C.

Variabile incompleta	continua		dicotomica		discreta		
	migliore	peggiore	migliore	peggiore	migliore	peggiore	
confondente	$\overline{\text{Distorsione}_{rel}}$	Cov. adjustment CCA	Stratificazione IM	Cov. adjustment CCA	Stratificazione IM	Stratificazione IM	Cov. adjustment IM
	$\overline{\text{EQM}}$	Cov. adjustment media e CCA	Matching CCA	Cov. adjustment imputazione	IPTW CCA	Stratificazione IM	IPTW CCA
predittore di Y	$\overline{\text{Distorsione}_{rel}}$	Matching CCA	IPTW tutti	Cov. adjustment tutti	Stratificazione CCA	Matching PMM e regressione	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	IPTW CCA	Cov. adjustment imputazione	IPTW CCA	Cov. adjustment imputazione	IPTW CCA
predittore di T	$\overline{\text{Distorsione}_{rel}}$	IPTW media	IPTW CCA	Cov. adjustment media	Stratificazione CCA	Cov. adjustment CCA	Stratificazione media
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	IPTW CCA	Cov. adjustment imputazione	IPTW CCA	Cov. adjustment imputazione	IPTW CCA

IM: Imputazione multipla (tutti i metodi). Imputazione: tutti i metodi di imputazione singola e multipla

Tabella 3.6: Schema riassuntivo delle migliori e peggiori combinazioni di metodi di trattamento del confondimento e di gestione dei dati mancanti nelle 9 simulazioni per lo Scenario D.

Variabile incompleta	continua		dicotomica		discreta		
	migliore	peggiore	migliore	peggiore	migliore	peggiore	
confondente	$\overline{\text{Distorsione}_{rel}}$	Cov. adjustment CCA	Stratificazione tutti	Cov. adjustment CCA	Stratificazione CCA	Cov. adjustment CCA	IPTW IM
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment media	Matching CCA
predittore di Y	$\overline{\text{Distorsione}_{rel}}$	Cov. adjustment media	Stratificazione CCA	Matching WPMM	Stratificazione CCA	IPTW regressione	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA
predittore di T	$\overline{\text{Distorsione}_{rel}}$	Cov. adjustment CCA	Stratificazione media	Stratificazione media	IPTW IM	Cov. adjustment IM	Stratificazione CCA
	$\overline{\text{EQM}}$	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA	Cov. adjustment imputazione	Matching CCA

IM: Imputazione multipla (tutti i metodi). Imputazione: tutti i metodi di imputazione singola e multipla

Tabella 3.7: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 1. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Dati completi	CCA	-0.144	-0.150	0.077	0.045	0.055	-0.135	-0.140	0.068	0.035	0.054	-0.148	-0.138	0.078	-0.064	0.068	-0.136	-0.115	0.069	-0.136	0.061
	Meta	-0.144	-0.139	0.080	-0.032	0.084	-0.135	-0.135	0.077	0.004	0.078	-0.148	-0.137	0.080	-0.071	0.086	-0.136	-0.120	0.077	-0.117	0.072
	WPMM	-0.144	-0.133	0.074	-0.075	0.065	-0.135	-0.132	0.057	0.129	0.064	-0.148	-0.114	0.064	-0.230	0.070	-0.136	-0.113	0.069	-0.171	0.063
Matching	CCA	-0.144	-0.129	0.086	-0.098	0.053	-0.135	-0.157	0.075	0.166	0.054	-0.148	-0.097	0.075	-0.347	0.066	-0.136	-0.113	0.076	-0.169	0.057
	Meta	-0.144	-0.129	0.085	-0.104	0.055	-0.135	-0.160	0.075	0.184	0.054	-0.148	-0.098	0.075	-0.340	0.067	-0.136	-0.113	0.076	-0.170	0.058
	WPMM	-0.144	-0.129	0.085	-0.099	0.052	-0.135	-0.157	0.076	0.168	0.055	-0.148	-0.096	0.075	-0.350	0.067	-0.136	-0.112	0.076	-0.178	0.058
Dati completi	CCA	-0.144	-0.137	0.055	-0.045	0.042	-0.135	-0.133	0.049	-0.015	0.041	-0.148	-0.117	0.049	-0.211	0.053	-0.136	-0.112	0.047	-0.180	0.036
	Meta	-0.144	-0.125	0.059	-0.132	0.066	-0.135	-0.126	0.058	-0.061	0.056	-0.148	-0.118	0.057	-0.199	0.066	-0.136	-0.107	0.057	-0.210	0.053
	WPMM	-0.144	-0.122	0.053	-0.132	0.047	-0.135	-0.142	0.048	0.054	0.043	-0.148	-0.086	0.044	-0.420	0.071	-0.136	-0.105	0.047	-0.229	0.041
Stratificazione	CCA	-0.144	-0.114	0.056	-0.205	0.046	-0.135	-0.149	0.049	0.107	0.041	-0.148	-0.076	0.046	-0.483	0.076	-0.136	-0.106	0.048	-0.224	0.040
	Meta	-0.144	-0.114	0.056	-0.205	0.046	-0.135	-0.150	0.049	0.110	0.041	-0.148	-0.076	0.046	-0.485	0.077	-0.136	-0.105	0.048	-0.226	0.040
	WPMM	-0.144	-0.114	0.056	-0.208	0.046	-0.135	-0.149	0.049	0.104	0.042	-0.148	-0.076	0.046	-0.483	0.076	-0.136	-0.105	0.048	-0.227	0.041
Dati completi	CCA	-0.144	-0.154	0.041	0.069	0.028	-0.135	-0.135	0.057	0.005	0.024	-0.148	-0.145	0.040	-0.016	0.036	-0.136	-0.131	0.037	-0.036	0.029
	Meta	-0.144	-0.146	0.049	0.018	0.044	-0.135	-0.127	0.045	-0.056	0.042	-0.148	-0.142	0.052	-0.037	0.052	-0.136	-0.123	0.048	-0.099	0.044
	WPMM	-0.144	-0.139	0.040	-0.030	0.027	-0.135	-0.150	0.036	0.109	0.028	-0.148	-0.106	0.039	-0.286	0.052	-0.136	-0.122	0.038	-0.103	0.030
Overariate adjustment	CCA	-0.144	-0.132	0.041	-0.083	0.029	-0.135	-0.157	0.037	0.162	0.031	-0.148	-0.093	0.039	-0.574	0.060	-0.136	-0.122	0.038	-0.105	0.030
	Meta	-0.144	-0.132	0.041	-0.081	0.029	-0.135	-0.157	0.037	0.162	0.031	-0.148	-0.092	0.039	-0.576	0.060	-0.136	-0.122	0.038	-0.106	0.030
	WPMM	-0.144	-0.132	0.041	-0.084	0.029	-0.135	-0.157	0.037	0.162	0.031	-0.148	-0.092	0.039	-0.574	0.060	-0.136	-0.122	0.038	-0.106	0.030
Dati completi	CCA	-0.144	-0.135	0.068	0.079	0.046	-0.135	-0.144	0.057	0.065	0.048	-0.148	-0.143	0.061	-0.032	0.081	-0.136	-0.118	0.055	-0.135	0.053
	Meta	-0.144	-0.143	0.074	-0.008	0.073	-0.135	-0.135	0.066	0.006	0.065	-0.148	-0.113	0.072	-0.238	0.099	-0.136	-0.116	0.068	-0.148	0.069
	WPMM	-0.144	-0.146	0.070	0.014	0.055	-0.135	-0.155	0.054	0.131	0.049	-0.148	-0.108	0.056	-0.273	0.068	-0.136	-0.112	0.055	-0.177	0.055
IPTW	CCA	-0.144	-0.137	0.071	-0.047	0.054	-0.135	-0.163	0.055	0.210	0.051	-0.148	-0.094	0.058	-0.365	0.071	-0.136	-0.112	0.056	-0.177	0.054
	Meta	-0.144	-0.138	0.071	-0.036	0.055	-0.135	-0.163	0.055	0.210	0.050	-0.148	-0.095	0.059	-0.355	0.072	-0.136	-0.112	0.056	-0.179	0.054
	WPMM	-0.144	-0.137	0.071	-0.046	0.054	-0.135	-0.163	0.055	0.209	0.050	-0.148	-0.094	0.058	-0.367	0.072	-0.136	-0.112	0.056	-0.178	0.054

Tabella 3.8: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 2. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D									
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}					
Matching	Dati completi	-0.144	-0.131	0.079	-0.088	0.070	-0.135	-0.144	0.065	0.069	0.045	-0.148	-0.133	0.075	-0.101	0.063	-0.136	-0.125	0.070	-0.075	0.064	-0.136	-0.117	0.082	-0.141	0.087
	CCA	-0.144	-0.132	0.086	-0.083	0.089	-0.135	-0.131	0.074	-0.030	0.070	-0.148	-0.140	0.088	-0.055	0.091	-0.136	-0.117	0.082	-0.141	0.087	-0.136	-0.116	0.064	-0.142	0.061
	Media	-0.144	-0.117	0.077	-0.190	0.076	-0.135	-0.127	0.064	-0.056	0.053	-0.148	-0.122	0.067	-0.177	0.067	-0.136	-0.120	0.072	-0.113	0.053	-0.136	-0.120	0.072	-0.113	0.053
	PAM	-0.144	-0.122	0.086	-0.156	0.065	-0.135	-0.129	0.073	-0.048	0.044	-0.148	-0.123	0.076	-0.166	0.060	-0.136	-0.120	0.072	-0.113	0.053	-0.136	-0.120	0.072	-0.113	0.053
Stratificazione	WPAM	-0.144	-0.120	0.085	-0.167	0.067	-0.135	-0.129	0.073	-0.043	0.044	-0.148	-0.123	0.077	-0.170	0.059	-0.136	-0.121	0.073	-0.107	0.054	-0.136	-0.121	0.073	-0.107	0.054
	Regressione	-0.144	-0.120	0.085	-0.168	0.066	-0.135	-0.129	0.073	-0.045	0.044	-0.148	-0.123	0.077	-0.169	0.060	-0.136	-0.121	0.073	-0.108	0.052	-0.136	-0.121	0.073	-0.108	0.052
	Dati completi	-0.144	-0.126	0.056	-0.128	0.049	-0.135	-0.125	0.047	-0.078	0.030	-0.148	-0.114	0.049	-0.228	0.046	-0.136	-0.122	0.048	-0.104	0.039	-0.136	-0.122	0.048	-0.104	0.039
	CCA	-0.144	-0.120	0.064	-0.166	0.068	-0.135	-0.118	0.056	-0.125	0.054	-0.148	-0.116	0.061	-0.219	0.063	-0.136	-0.115	0.060	-0.150	0.059	-0.136	-0.115	0.060	-0.150	0.059
Covariate adjustment	Media	-0.144	-0.118	0.056	-0.180	0.054	-0.135	-0.120	0.046	-0.111	0.037	-0.148	-0.107	0.045	-0.279	0.052	-0.136	-0.119	0.045	-0.127	0.041	-0.136	-0.119	0.045	-0.127	0.041
	PAM	-0.144	-0.119	0.057	-0.176	0.052	-0.135	-0.119	0.047	-0.117	0.035	-0.148	-0.106	0.047	-0.287	0.052	-0.136	-0.117	0.046	-0.139	0.040	-0.136	-0.117	0.046	-0.139	0.040
	WPAM	-0.144	-0.120	0.058	-0.169	0.052	-0.135	-0.119	0.047	-0.116	0.035	-0.148	-0.105	0.047	-0.288	0.052	-0.136	-0.117	0.046	-0.137	0.040	-0.136	-0.117	0.046	-0.137	0.040
	Regressione	-0.144	-0.119	0.058	-0.172	0.052	-0.135	-0.119	0.047	-0.117	0.035	-0.148	-0.106	0.047	-0.285	0.052	-0.136	-0.117	0.046	-0.138	0.040	-0.136	-0.117	0.046	-0.138	0.040
IPTW	Dati completi	-0.144	-0.153	0.040	0.059	0.032	-0.135	-0.135	0.036	-0.003	0.024	-0.148	-0.151	0.040	0.023	0.030	-0.136	-0.137	0.039	0.008	0.031	-0.136	-0.137	0.039	0.008	0.031
	CCA	-0.144	-0.150	0.050	0.038	0.047	-0.135	-0.130	0.045	-0.035	0.039	-0.148	-0.151	0.053	0.021	0.046	-0.136	-0.130	0.049	-0.042	0.047	-0.136	-0.130	0.049	-0.042	0.047
	Media	-0.144	-0.147	0.040	0.022	0.031	-0.135	-0.130	0.036	-0.038	0.025	-0.148	-0.137	0.039	-0.076	0.030	-0.136	-0.124	0.037	-0.086	0.032	-0.136	-0.124	0.037	-0.086	0.032
	PAM	-0.144	-0.146	0.041	0.015	0.031	-0.135	-0.129	0.036	-0.045	0.024	-0.148	-0.135	0.040	-0.089	0.030	-0.136	-0.122	0.038	-0.098	0.032	-0.136	-0.122	0.038	-0.098	0.032
IPTW	WPAM	-0.144	-0.146	0.041	0.015	0.031	-0.135	-0.129	0.036	-0.045	0.024	-0.148	-0.135	0.040	-0.088	0.030	-0.136	-0.122	0.038	-0.099	0.032	-0.136	-0.122	0.038	-0.099	0.032
	Regressione	-0.144	-0.146	0.041	0.015	0.031	-0.135	-0.129	0.036	-0.045	0.024	-0.148	-0.135	0.039	-0.088	0.030	-0.136	-0.123	0.038	-0.099	0.032	-0.136	-0.123	0.038	-0.099	0.032
	Dati completi	-0.144	-0.142	0.067	-0.012	0.052	-0.135	-0.141	0.055	0.046	0.034	-0.148	-0.155	0.058	0.051	0.082	-0.136	-0.133	0.061	-0.020	0.050	-0.136	-0.133	0.061	-0.020	0.050
	CCA	-0.144	-0.138	0.076	-0.040	0.073	-0.135	-0.133	0.065	-0.017	0.059	-0.148	-0.140	0.075	-0.055	0.120	-0.136	-0.126	0.074	-0.070	0.077	-0.136	-0.126	0.074	-0.070	0.077
IPTW	Media	-0.144	-0.137	0.068	-0.048	0.058	-0.135	-0.135	0.055	0.001	0.036	-0.148	-0.133	0.061	-0.099	0.055	-0.136	-0.129	0.059	-0.048	0.056	-0.136	-0.129	0.059	-0.048	0.056
	PAM	-0.144	-0.137	0.068	-0.050	0.056	-0.135	-0.134	0.056	-0.009	0.034	-0.148	-0.133	0.062	-0.100	0.053	-0.136	-0.128	0.060	-0.059	0.053	-0.136	-0.128	0.060	-0.059	0.053
	WPAM	-0.144	-0.137	0.068	-0.048	0.057	-0.135	-0.134	0.056	-0.005	0.034	-0.148	-0.133	0.063	-0.101	0.053	-0.136	-0.128	0.060	-0.055	0.054	-0.136	-0.128	0.060	-0.055	0.054
	Regressione	-0.144	-0.137	0.068	-0.050	0.057	-0.135	-0.134	0.056	-0.008	0.034	-0.148	-0.133	0.062	-0.100	0.053	-0.136	-0.128	0.060	-0.059	0.053	-0.136	-0.128	0.060	-0.059	0.053

Tabella 3.9: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 3. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Dati completi	Dati completi	-0.144	-0.147	0.072	0.022	0.066	-0.135	-0.127	0.067	-0.057	0.057	-0.148	-0.148	0.076	0.000	0.064	-0.136	-0.109	0.071	-0.198	0.059
	CCA	-0.144	-0.131	0.083	-0.092	0.091	-0.135	-0.124	0.074	-0.076	0.077	-0.148	-0.139	0.083	-0.061	0.085	-0.136	-0.116	0.079	-0.140	0.081
	Meta	-0.144	-0.130	0.071	-0.098	0.067	-0.135	-0.169	0.059	0.260	0.068	-0.148	-0.163	0.074	0.106	0.072	-0.136	-0.200	0.054	-0.475	0.076
Matching	PMM	-0.144	-0.120	0.081	-0.168	0.061	-0.135	-0.173	0.071	0.287	0.058	-0.148	-0.170	0.085	0.151	0.062	-0.136	-0.204	0.064	-0.506	0.074
	WPMM	-0.144	-0.123	0.080	-0.147	0.061	-0.135	-0.174	0.069	0.297	0.059	-0.148	-0.167	0.083	0.129	0.064	-0.136	-0.205	0.063	-0.513	0.075
	Regressione	-0.144	-0.120	0.081	-0.166	0.062	-0.135	-0.173	0.071	0.285	0.058	-0.148	-0.168	0.085	0.138	0.061	-0.136	-0.203	0.065	-0.495	0.073
Dati completi	Dati completi	-0.144	-0.132	0.053	-0.080	0.049	-0.135	-0.120	0.049	-0.107	0.048	-0.148	-0.123	0.049	-0.168	0.045	-0.136	-0.109	0.048	-0.192	0.041
	CCA	-0.144	-0.123	0.062	-0.143	0.071	-0.135	-0.127	0.057	-0.059	0.061	-0.148	-0.118	0.059	-0.197	0.066	-0.136	-0.109	0.059	-0.197	0.058
	Meta	-0.144	-0.113	0.051	-0.213	0.056	-0.135	-0.156	0.044	0.163	0.048	-0.148	-0.138	0.048	-0.068	0.045	-0.136	-0.178	0.040	-0.313	0.052
Stratificazione	PMM	-0.144	-0.108	0.054	-0.250	0.056	-0.135	-0.163	0.047	0.215	0.049	-0.148	-0.141	0.050	-0.047	0.043	-0.136	-0.189	0.042	-0.396	0.058
	WPMM	-0.144	-0.109	0.054	-0.241	0.055	-0.135	-0.165	0.046	0.226	0.049	-0.148	-0.141	0.050	-0.042	0.043	-0.136	-0.190	0.041	-0.403	0.059
	Regressione	-0.144	-0.109	0.054	-0.243	0.056	-0.135	-0.164	0.047	0.219	0.049	-0.148	-0.141	0.050	-0.043	0.043	-0.136	-0.190	0.042	-0.400	0.059
Dati completi	Dati completi	-0.144	-0.145	0.040	0.006	0.033	-0.135	-0.134	0.036	-0.001	0.034	-0.148	-0.161	0.040	0.088	0.034	-0.136	-0.134	0.039	-0.209	0.036
	CCA	-0.144	-0.141	0.050	-0.017	0.047	-0.135	-0.131	0.045	-0.027	0.043	-0.148	-0.156	0.052	0.058	0.051	-0.136	-0.130	0.050	-0.042	0.050
	Meta	-0.144	-0.128	0.040	-0.111	0.038	-0.135	-0.163	0.036	0.214	0.043	-0.148	-0.178	0.040	0.202	0.044	-0.136	-0.194	0.036	-0.429	0.061
Overariate adjustment	PMM	-0.144	-0.121	0.041	-0.161	0.041	-0.135	-0.174	0.037	0.292	0.049	-0.148	-0.182	0.040	0.231	0.046	-0.136	-0.212	0.036	-0.563	0.078
	WPMM	-0.144	-0.120	0.040	-0.162	0.041	-0.135	-0.174	0.037	0.295	0.049	-0.148	-0.182	0.040	0.232	0.046	-0.136	-0.212	0.036	-0.564	0.078
	Regressione	-0.144	-0.121	0.041	-0.161	0.041	-0.135	-0.174	0.037	0.295	0.049	-0.148	-0.182	0.040	0.232	0.046	-0.136	-0.212	0.036	-0.561	0.078
Dati completi	Dati completi	-0.144	-0.148	0.066	0.028	0.056	-0.135	-0.128	0.054	-0.048	0.048	-0.148	-0.157	0.058	0.051	0.083	-0.136	-0.126	0.059	-0.207	0.048
	CCA	-0.144	-0.142	0.075	-0.014	0.079	-0.135	-0.131	0.063	-0.024	0.063	-0.148	-0.166	0.072	0.125	0.113	-0.136	-0.124	0.071	-0.083	0.071
	Meta	-0.144	-0.131	0.064	-0.088	0.059	-0.135	-0.178	0.056	0.323	0.064	-0.148	-0.159	0.057	0.079	0.045	-0.136	-0.205	0.049	-0.515	0.077
IPTW	PMM	-0.144	-0.124	0.067	-0.136	0.060	-0.135	-0.179	0.057	0.334	0.061	-0.148	-0.162	0.058	0.100	0.046	-0.136	-0.211	0.052	-0.559	0.080
	WPMM	-0.144	-0.125	0.066	-0.130	0.060	-0.135	-0.181	0.057	0.348	0.062	-0.148	-0.163	0.057	0.101	0.045	-0.136	-0.212	0.053	-0.567	0.082
	Regressione	-0.144	-0.125	0.067	-0.134	0.060	-0.135	-0.180	0.057	0.338	0.061	-0.148	-0.163	0.058	0.101	0.046	-0.136	-0.211	0.052	-0.556	0.080

Tabella 3.10: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 4. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A						Scenario B						Scenario C						Scenario D					
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}				
Matching	Dati completi	-0.137	-0.113	0.100	-0.177	0.101	-0.136	-0.148	0.063	0.084	0.054	-0.138	-0.145	0.124	0.045	0.113	-0.138	-0.127	0.076	-0.084	0.065				
	CCA	-0.137	-0.102	0.102	-0.259	0.118	-0.136	-0.140	0.072	0.027	0.074	-0.138	-0.135	0.120	-0.025	0.129	-0.138	-0.125	0.083	-0.095	0.084				
	Media	-0.137	-0.105	0.099	-0.233	0.105	-0.136	-0.142	0.063	0.047	0.056	-0.138	-0.146	0.124	0.052	0.116	-0.138	-0.125	0.077	-0.094	0.071				
	PAM	-0.137	-0.107	0.106	-0.222	0.097	-0.136	-0.142	0.070	0.043	0.049	-0.138	-0.146	0.130	0.058	0.111	-0.138	-0.125	0.083	-0.093	0.065				
	WPAM	-0.137	-0.104	0.105	-0.237	0.096	-0.136	-0.142	0.069	0.045	0.049	-0.138	-0.147	0.129	0.064	0.112	-0.138	-0.127	0.083	-0.082	0.065				
Stratificazione	Regressione	-0.137	-0.105	0.107	-0.234	0.096	-0.136	-0.142	0.070	0.046	0.048	-0.138	-0.145	0.129	0.051	0.112	-0.138	-0.126	0.083	-0.091	0.065				
	Dati completi	-0.137	-0.098	0.080	-0.282	0.079	-0.136	-0.132	0.046	-0.030	0.034	-0.138	-0.129	0.099	-0.067	0.092	-0.138	-0.124	0.057	-0.103	0.047				
	CCA	-0.137	-0.096	0.083	-0.303	0.106	-0.136	-0.127	0.054	-0.063	0.054	-0.138	-0.131	0.096	-0.055	0.104	-0.138	-0.113	0.063	-0.182	0.069				
	Media	-0.137	-0.099	0.080	-0.274	0.081	-0.136	-0.132	0.046	-0.030	0.034	-0.138	-0.124	0.101	-0.099	0.095	-0.138	-0.121	0.057	-0.123	0.048				
	PAM	-0.137	-0.101	0.083	-0.263	0.077	-0.136	-0.132	0.047	-0.032	0.033	-0.138	-0.130	0.103	-0.058	0.089	-0.138	-0.122	0.058	-0.114	0.047				
Covariate adjustment	WPAM	-0.137	-0.101	0.083	-0.264	0.078	-0.136	-0.132	0.047	-0.032	0.033	-0.138	-0.125	0.104	-0.097	0.090	-0.138	-0.122	0.058	-0.115	0.048				
	Regressione	-0.137	-0.100	0.083	-0.270	0.078	-0.136	-0.132	0.047	-0.033	0.033	-0.138	-0.125	0.104	-0.092	0.093	-0.138	-0.122	0.058	-0.115	0.047				
	Dati completi	-0.137	-0.132	0.050	-0.037	0.035	-0.136	-0.141	0.036	0.035	0.025	-0.138	-0.144	0.052	0.042	0.048	-0.138	-0.143	0.038	0.033	0.032				
	CCA	-0.137	-0.126	0.056	-0.078	0.047	-0.136	-0.133	0.045	-0.022	0.040	-0.138	-0.133	0.059	-0.042	0.056	-0.138	-0.132	0.047	-0.047	0.046				
	Media	-0.137	-0.132	0.050	-0.034	0.036	-0.136	-0.140	0.036	0.029	0.025	-0.138	-0.144	0.052	0.043	0.048	-0.138	-0.142	0.039	0.025	0.033				
IPTW	PAM	-0.137	-0.132	0.050	-0.035	0.036	-0.136	-0.141	0.036	0.033	0.025	-0.138	-0.145	0.052	0.045	0.047	-0.138	-0.143	0.039	0.034	0.033				
	WPAM	-0.137	-0.132	0.050	-0.034	0.036	-0.136	-0.140	0.036	0.032	0.025	-0.138	-0.145	0.052	0.045	0.047	-0.138	-0.143	0.039	0.033	0.033				
	Regressione	-0.137	-0.132	0.050	-0.034	0.035	-0.136	-0.141	0.036	0.033	0.025	-0.138	-0.144	0.052	0.044	0.047	-0.138	-0.143	0.039	0.033	0.033				
	Dati completi	-0.137	-0.104	0.088	-0.237	0.078	-0.136	-0.146	0.054	0.077	0.038	-0.138	-0.145	0.114	2.232	0.356	-0.138	-0.133	0.066	-0.037	0.069				
	CCA	-0.137	-0.105	0.091	-0.235	0.104	-0.136	-0.139	0.063	0.024	0.061	-0.138	-0.309	0.110	1.237	0.312	-0.138	-0.130	0.075	-0.060	0.085				
IPTW	Media	-0.137	-0.105	0.087	-0.234	0.079	-0.136	-0.146	0.053	0.072	0.038	-0.138	-0.173	0.114	2.249	0.109	-0.138	-0.132	0.066	-0.048	0.069				
	PAM	-0.137	-0.105	0.088	-0.235	0.079	-0.136	-0.146	0.053	0.074	0.038	-0.138	-0.173	0.114	2.250	0.109	-0.138	-0.133	0.066	-0.038	0.070				
	WPAM	-0.137	-0.105	0.088	-0.234	0.079	-0.136	-0.146	0.053	0.073	0.038	-0.138	-0.173	0.114	2.250	0.109	-0.138	-0.133	0.066	-0.038	0.070				
	Regressione	-0.137	-0.104	0.088	-0.238	0.078	-0.136	-0.146	0.053	0.074	0.038	-0.138	-0.173	0.114	2.248	0.108	-0.138	-0.133	0.067	-0.039	0.070				
	CCA	-0.137	-0.104	0.088	-0.238	0.078	-0.136	-0.146	0.053	0.074	0.038	-0.138	-0.173	0.114	2.248	0.108	-0.138	-0.133	0.067	-0.039	0.070				

Tabella 3.11: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 5. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A						Scenario B						Scenario C						Scenario D					
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}				
Dati completi	Dati completi	-0.145	-0.132	0.073	-0.087	0.057	-0.135	-0.139	0.063	0.028	0.050	-0.146	-0.135	0.060	-0.080	0.046	-0.131	-0.133	0.059	0.015	0.045				
	CCA	-0.145	-0.136	0.082	-0.061	0.083	-0.135	-0.133	0.072	-0.018	0.069	-0.146	-0.134	0.073	-0.085	0.073	-0.131	-0.126	0.070	-0.041	0.067				
	Matchng	-0.145	-0.134	0.073	-0.078	0.059	-0.135	-0.138	0.063	0.023	0.048	-0.146	-0.137	0.060	-0.068	0.055	-0.131	-0.131	0.058	-0.004	0.052				
	WPMM	-0.145	-0.135	0.079	-0.070	0.055	-0.135	-0.141	0.070	0.040	0.041	-0.146	-0.136	0.068	-0.070	0.048	-0.131	-0.130	0.066	-0.009	0.046				
Dati completi	Regressione	-0.145	-0.135	0.080	-0.071	0.055	-0.135	-0.141	0.070	0.041	0.041	-0.146	-0.137	0.068	-0.067	0.047	-0.131	-0.131	0.066	-0.003	0.045				
	CCA	-0.145	-0.127	0.052	-0.125	0.041	-0.135	-0.129	0.046	-0.044	0.035	-0.146	-0.114	0.042	-0.220	0.046	-0.131	-0.124	0.043	-0.054	0.037				
	Matchng	-0.145	-0.119	0.060	-0.181	0.061	-0.135	-0.126	0.054	-0.068	0.050	-0.146	-0.114	0.053	-0.221	0.059	-0.131	-0.119	0.054	-0.096	0.052				
	WPMM	-0.145	-0.129	0.052	-0.112	0.041	-0.135	-0.129	0.046	-0.046	0.036	-0.146	-0.115	0.042	-0.218	0.046	-0.131	-0.124	0.043	-0.055	0.038				
Stratificazione	CCA	-0.145	-0.128	0.053	-0.120	0.040	-0.135	-0.129	0.046	-0.046	0.034	-0.146	-0.115	0.043	-0.217	0.046	-0.131	-0.124	0.044	-0.053	0.038				
	Matchng	-0.145	-0.128	0.053	-0.121	0.040	-0.135	-0.129	0.046	-0.046	0.034	-0.146	-0.115	0.043	-0.217	0.045	-0.131	-0.124	0.044	-0.053	0.038				
	WPMM	-0.145	-0.127	0.053	-0.123	0.040	-0.135	-0.129	0.046	-0.047	0.034	-0.146	-0.115	0.043	-0.217	0.045	-0.131	-0.124	0.044	-0.054	0.038				
	Regressione	-0.145	-0.154	0.038	0.060	0.030	-0.135	-0.139	0.035	0.029	0.022	-0.146	-0.145	0.039	-0.010	0.035	-0.131	-0.132	0.038	0.002	0.033				
Overariate adjustment	CCA	-0.145	-0.150	0.049	0.031	0.044	-0.135	-0.136	0.044	0.003	0.038	-0.146	-0.145	0.050	-0.008	0.048	-0.131	-0.125	0.048	-0.045	0.045				
	Matchng	-0.145	-0.154	0.039	0.062	0.030	-0.135	-0.139	0.035	0.030	0.022	-0.146	-0.145	0.039	-0.010	0.035	-0.131	-0.132	0.038	0.006	0.032				
	WPMM	-0.145	-0.154	0.039	0.063	0.030	-0.135	-0.139	0.036	0.029	0.022	-0.146	-0.145	0.039	-0.009	0.035	-0.131	-0.132	0.038	0.007	0.032				
	Regressione	-0.145	-0.154	0.039	0.063	0.030	-0.135	-0.139	0.036	0.030	0.022	-0.146	-0.145	0.039	-0.009	0.035	-0.131	-0.132	0.038	0.007	0.032				
IPTW	Dati completi	-0.145	-0.144	0.063	-0.007	0.032	-0.135	-0.143	0.052	0.059	0.033	-0.146	-0.164	0.050	0.120	0.062	-0.131	-0.133	0.050	0.013	0.046				
	CCA	-0.145	-0.138	0.073	-0.048	0.071	-0.135	-0.136	0.062	0.008	0.057	-0.146	-0.143	0.065	-0.020	0.083	-0.131	-0.127	0.061	-0.030	0.062				
	Matchng	-0.145	-0.145	0.064	-0.001	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.072	0.049	-0.131	-0.134	0.050	0.017	0.046				
	WPMM	-0.145	-0.145	0.064	-0.002	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.069	0.049	-0.131	-0.134	0.050	0.018	0.046				
Regressione	CCA	-0.145	-0.145	0.064	-0.002	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.069	0.049	-0.131	-0.134	0.050	0.018	0.046				
	Matchng	-0.145	-0.145	0.064	-0.002	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.069	0.049	-0.131	-0.134	0.050	0.018	0.046				
	WPMM	-0.145	-0.145	0.064	-0.002	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.069	0.049	-0.131	-0.134	0.050	0.018	0.046				
	Regressione	-0.145	-0.145	0.064	-0.002	0.053	-0.135	-0.143	0.053	0.057	0.033	-0.146	-0.136	0.051	-0.070	0.049	-0.131	-0.134	0.050	0.018	0.046				

Tabella 3.12: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 6. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Matching	Dati completi	-0.146	-0.142	0.061	-0.027	0.052	-0.137	-0.129	0.085	-0.057	0.068	-0.147	-0.147	0.080	0.001	0.060	-0.138	-0.134	0.101	-0.085	0.096
	CCA	-0.146	-0.142	0.070	-0.024	0.073	-0.137	-0.124	0.087	-0.095	0.093	-0.147	-0.140	0.086	-0.042	0.081	-0.138	-0.125	0.101	-0.094	0.110
	Media	-0.146	-0.140	0.060	-0.041	0.059	-0.137	-0.125	0.085	-0.086	0.076	-0.147	-0.147	0.079	0.003	0.071	-0.138	-0.134	0.101	-0.033	0.101
	PAM	-0.146	-0.143	0.070	-0.021	0.051	-0.137	-0.126	0.096	-0.079	0.071	-0.147	-0.146	0.089	-0.003	0.066	-0.138	-0.131	0.110	-0.053	0.098
Stratificazione	WPAM	-0.146	-0.141	0.068	-0.033	0.053	-0.137	-0.128	0.091	-0.069	0.070	-0.147	-0.149	0.086	0.019	0.064	-0.138	-0.131	0.106	-0.055	0.098
	Regressione	-0.146	-0.143	0.070	-0.018	0.052	-0.137	-0.126	0.096	-0.082	0.070	-0.147	-0.146	0.089	-0.003	0.064	-0.138	-0.131	0.110	-0.054	0.097
	Dati completi	-0.146	-0.130	0.042	-0.105	0.037	-0.137	-0.129	0.064	-0.058	0.063	-0.147	-0.128	0.055	-0.126	0.047	-0.138	-0.122	0.083	-0.118	0.081
	CCA	-0.146	-0.127	0.052	-0.127	0.054	-0.137	-0.119	0.067	-0.131	0.078	-0.147	-0.117	0.062	-0.200	0.068	-0.138	-0.115	0.082	-0.171	0.091
Covariate adjustment	Media	-0.146	-0.129	0.042	-0.110	0.040	-0.137	-0.128	0.064	-0.068	0.062	-0.147	-0.124	0.055	-0.157	0.050	-0.138	-0.124	0.083	-0.108	0.083
	PAM	-0.146	-0.128	0.043	-0.117	0.039	-0.137	-0.127	0.068	-0.077	0.061	-0.147	-0.123	0.057	-0.163	0.051	-0.138	-0.121	0.088	-0.125	0.083
	WPAM	-0.146	-0.128	0.043	-0.119	0.039	-0.137	-0.129	0.065	-0.063	0.062	-0.147	-0.124	0.056	-0.155	0.050	-0.138	-0.122	0.085	-0.118	0.081
	Regressione	-0.146	-0.128	0.043	-0.118	0.040	-0.137	-0.127	0.068	-0.078	0.061	-0.147	-0.122	0.057	-0.167	0.051	-0.138	-0.121	0.089	-0.124	0.082
IPTW	Dati completi	-0.146	-0.147	0.036	0.011	0.029	-0.137	-0.130	0.042	-0.052	0.037	-0.147	-0.158	0.040	0.076	0.030	-0.138	-0.137	0.045	-0.007	0.032
	CCA	-0.146	-0.144	0.046	-0.008	0.042	-0.137	-0.128	0.050	-0.067	0.048	-0.147	-0.156	0.050	0.061	0.047	-0.138	-0.136	0.052	-0.016	0.045
	Media	-0.146	-0.146	0.037	0.001	0.031	-0.137	-0.130	0.042	-0.053	0.037	-0.147	-0.156	0.040	0.065	0.031	-0.138	-0.137	0.044	-0.011	0.033
	PAM	-0.146	-0.146	0.037	-0.000	0.031	-0.137	-0.131	0.042	-0.048	0.037	-0.147	-0.156	0.041	0.065	0.031	-0.138	-0.138	0.045	-0.004	0.033
IPTW	WPAM	-0.146	-0.145	0.038	-0.001	0.031	-0.137	-0.131	0.042	-0.049	0.037	-0.147	-0.157	0.041	0.067	0.032	-0.138	-0.138	0.045	-0.004	0.033
	Regressione	-0.146	-0.145	0.037	-0.000	0.031	-0.137	-0.131	0.042	-0.048	0.037	-0.147	-0.156	0.040	0.066	0.031	-0.138	-0.138	0.045	-0.004	0.033
	Dati completi	-0.146	-0.147	0.051	0.012	0.038	-0.137	-0.128	0.067	-0.071	0.056	-0.147	-0.017	0.068	-0.881	0.154	-0.138	-0.141	0.088	0.014	0.102
	CCA	-0.146	-0.145	0.062	-0.006	0.062	-0.137	-0.131	0.074	-0.047	0.078	-0.147	-0.007	0.076	-0.953	0.167	-0.138	-0.140	0.093	0.007	0.107
IPTW	Media	-0.146	-0.146	0.051	0.002	0.039	-0.137	-0.128	0.066	-0.069	0.057	-0.147	-0.148	0.066	0.011	0.056	-0.138	-0.140	0.090	0.008	0.102
	PAM	-0.146	-0.146	0.052	0.000	0.039	-0.137	-0.128	0.068	-0.065	0.058	-0.147	-0.149	0.067	0.013	0.056	-0.138	-0.139	0.091	0.005	0.101
	WPAM	-0.146	-0.146	0.052	0.001	0.039	-0.137	-0.128	0.067	-0.067	0.057	-0.147	-0.148	0.067	0.012	0.055	-0.138	-0.140	0.090	0.013	0.101
	Regressione	-0.146	-0.146	0.052	0.000	0.039	-0.137	-0.128	0.067	-0.066	0.057	-0.147	-0.149	0.068	0.014	0.056	-0.138	-0.139	0.091	0.001	0.102

Tabella 3.13: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 7. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Matching	Dati completi	-0.141	-0.142	0.076	0.007	0.052	-0.143	-0.140	0.070	-0.021	0.049	-0.138	-0.115	0.077	-0.165	0.060	-0.143	-0.122	0.072	-0.151	0.059
	CCA	-0.141	-0.154	0.085	0.093	0.077	-0.143	-0.151	0.084	0.056	0.076	-0.138	-0.150	0.084	0.085	0.080	-0.143	-0.140	0.084	-0.025	0.074
	Meta	-0.141	-0.142	0.074	0.005	0.059	-0.143	-0.137	0.068	-0.041	0.052	-0.138	-0.139	0.063	0.010	0.052	-0.143	-0.125	0.071	-0.129	0.064
	PMM	-0.141	-0.143	0.082	0.017	0.048	-0.143	-0.136	0.075	-0.047	0.045	-0.138	-0.141	0.070	0.020	0.040	-0.143	-0.124	0.078	-0.134	0.056
	WFPMM	-0.141	-0.141	0.082	-0.001	0.049	-0.143	-0.137	0.074	-0.037	0.046	-0.138	-0.142	0.069	0.028	0.041	-0.143	-0.125	0.077	-0.130	0.056
Stratificazione	Dati completi	-0.141	-0.143	0.081	0.017	0.047	-0.143	-0.136	0.075	-0.045	0.046	-0.138	-0.142	0.070	0.029	0.042	-0.143	-0.123	0.078	-0.138	0.055
	CCA	-0.141	-0.141	0.068	0.002	0.060	-0.143	-0.142	0.067	-0.003	0.057	-0.138	-0.135	0.063	-0.021	0.055	-0.143	-0.126	0.063	-0.121	0.054
	Meta	-0.141	-0.133	0.058	-0.058	0.044	-0.143	-0.129	0.053	-0.097	0.038	-0.138	-0.122	0.047	-0.113	0.040	-0.143	-0.116	0.049	-0.193	0.042
	PMM	-0.141	-0.132	0.059	-0.066	0.042	-0.143	-0.129	0.053	-0.093	0.036	-0.138	-0.127	0.048	-0.082	0.035	-0.143	-0.117	0.050	-0.185	0.041
	WFPMM	-0.141	-0.132	0.059	-0.066	0.042	-0.143	-0.130	0.053	-0.091	0.036	-0.138	-0.127	0.047	-0.076	0.036	-0.143	-0.116	0.050	-0.188	0.041
Overariate adjustment	Dati completi	-0.141	-0.132	0.059	-0.067	0.042	-0.143	-0.129	0.053	-0.095	0.036	-0.138	-0.127	0.048	-0.079	0.035	-0.143	-0.116	0.050	-0.187	0.041
	CCA	-0.141	-0.161	0.043	0.144	0.047	-0.143	-0.150	0.044	0.054	0.042	-0.138	-0.151	0.048	0.093	0.047	-0.143	-0.145	0.049	0.014	0.041
	Meta	-0.141	-0.153	0.033	0.086	0.030	-0.143	-0.140	0.033	-0.019	0.025	-0.138	-0.135	0.033	-0.018	0.033	-0.143	-0.134	0.036	-0.063	0.026
	PMM	-0.141	-0.153	0.033	0.088	0.030	-0.143	-0.140	0.033	-0.019	0.025	-0.138	-0.136	0.033	-0.015	0.032	-0.143	-0.134	0.036	-0.063	0.026
	WFPMM	-0.141	-0.153	0.033	0.089	0.030	-0.143	-0.140	0.033	-0.020	0.025	-0.138	-0.136	0.033	-0.013	0.032	-0.143	-0.134	0.036	-0.063	0.026
IPYW	Dati completi	-0.141	-0.143	0.052	0.016	0.044	-0.143	-0.138	0.049	-0.030	0.041	-0.138	-0.135	0.054	-0.023	0.038	-0.143	-0.116	0.054	-0.187	0.053
	CCA	-0.141	-0.154	0.062	0.092	0.067	-0.143	-0.149	0.062	0.043	0.062	-0.138	-0.105	0.067	-0.236	0.109	-0.143	-0.133	0.069	-0.071	0.070
	Meta	-0.141	-0.148	0.052	0.052	0.046	-0.143	-0.138	0.046	-0.033	0.040	-0.138	-0.138	0.046	0.002	0.046	-0.143	-0.119	0.054	-0.170	0.053
	PMM	-0.141	-0.146	0.053	0.037	0.044	-0.143	-0.140	0.047	-0.017	0.039	-0.138	-0.139	0.047	0.008	0.042	-0.143	-0.119	0.054	-0.169	0.053
	WFPMM	-0.141	-0.146	0.053	0.039	0.045	-0.143	-0.140	0.047	-0.017	0.039	-0.138	-0.141	0.047	0.020	0.042	-0.143	-0.119	0.054	-0.170	0.053
Regressione	-0.141	-0.146	0.053	0.038	0.044	-0.143	-0.140	0.047	-0.019	0.038	-0.138	-0.139	0.047	0.007	0.042	-0.143	-0.119	0.054	-0.169	0.053	

Tabella 3.14: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 8. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Matching	Dati completi	-0.137	-0.118	0.076	-0.137	0.074	-0.128	-0.139	0.063	0.088	0.049	-0.142	-0.121	0.074	-0.152	0.066	-0.129	-0.122	0.067	-0.052	0.059
	CCA	-0.137	-0.118	0.083	-0.139	0.089	-0.128	-0.120	0.071	-0.057	0.070	-0.142	-0.128	0.085	-0.097	0.093	-0.129	-0.113	0.078	-0.121	0.085
	Media	-0.137	-0.116	0.075	-0.155	0.077	-0.128	-0.128	0.062	-0.001	0.053	-0.142	-0.124	0.066	-0.131	0.063	-0.129	-0.126	0.062	-0.017	0.061
	PAM	-0.137	-0.118	0.083	-0.137	0.069	-0.128	-0.128	0.071	0.006	0.043	-0.142	-0.128	0.075	-0.099	0.058	-0.129	-0.131	0.071	0.019	0.053
	WPAM	-0.137	-0.118	0.083	-0.142	0.069	-0.128	-0.129	0.071	0.012	0.044	-0.142	-0.127	0.076	-0.104	0.058	-0.129	-0.132	0.071	0.027	0.053
Stratificazione	Regressione	-0.137	-0.117	0.083	-0.149	0.069	-0.128	-0.129	0.071	0.007	0.044	-0.142	-0.127	0.076	-0.107	0.059	-0.129	-0.131	0.071	0.024	0.052
	Dati completi	-0.137	-0.114	0.054	-0.166	0.050	-0.128	-0.117	0.045	-0.080	0.032	-0.142	-0.107	0.047	-0.248	0.049	-0.129	-0.119	0.047	-0.070	0.040
	CCA	-0.137	-0.108	0.061	-0.209	0.069	-0.128	-0.109	0.053	-0.148	0.052	-0.142	-0.106	0.058	-0.251	0.067	-0.129	-0.113	0.058	-0.121	0.058
	Media	-0.137	-0.113	0.053	-0.179	0.056	-0.128	-0.118	0.044	-0.076	0.037	-0.142	-0.110	0.044	-0.224	0.050	-0.129	-0.127	0.043	-0.008	0.040
	PAM	-0.137	-0.114	0.055	-0.171	0.054	-0.128	-0.118	0.046	-0.078	0.035	-0.142	-0.110	0.045	-0.227	0.048	-0.129	-0.127	0.045	-0.013	0.038
Covariate adjustment	WPAM	-0.137	-0.115	0.055	-0.162	0.054	-0.128	-0.118	0.046	-0.076	0.035	-0.142	-0.110	0.045	-0.227	0.049	-0.129	-0.127	0.045	-0.011	0.038
	Regressione	-0.137	-0.114	0.055	-0.166	0.054	-0.128	-0.118	0.045	-0.076	0.034	-0.142	-0.110	0.045	-0.226	0.048	-0.129	-0.127	0.045	-0.011	0.037
	Dati completi	-0.137	-0.141	0.040	0.025	0.032	-0.128	-0.126	0.036	-0.015	0.023	-0.142	-0.142	0.040	-0.002	0.034	-0.129	-0.134	0.038	0.044	0.035
	CCA	-0.137	-0.136	0.050	-0.010	0.048	-0.128	-0.119	0.044	-0.067	0.039	-0.142	-0.139	0.053	-0.019	0.049	-0.129	-0.126	0.048	-0.019	0.050
	Media	-0.137	-0.141	0.041	0.025	0.032	-0.128	-0.126	0.036	-0.011	0.023	-0.142	-0.140	0.039	-0.016	0.034	-0.129	-0.133	0.037	0.036	0.035
IPTW	PAM	-0.137	-0.141	0.041	0.026	0.032	-0.128	-0.126	0.036	-0.009	0.022	-0.142	-0.139	0.039	-0.019	0.033	-0.129	-0.133	0.037	0.035	0.034
	WPAM	-0.137	-0.141	0.041	0.027	0.032	-0.128	-0.126	0.036	-0.010	0.022	-0.142	-0.139	0.040	-0.019	0.033	-0.129	-0.133	0.037	0.035	0.034
	Regressione	-0.137	-0.141	0.041	0.026	0.032	-0.128	-0.126	0.036	-0.009	0.023	-0.142	-0.139	0.039	-0.019	0.033	-0.129	-0.133	0.037	0.035	0.034
	Dati completi	-0.137	-0.133	0.067	-0.034	0.056	-0.128	-0.135	0.055	0.060	0.036	-0.142	-0.145	0.058	0.022	0.080	-0.129	-0.131	0.060	0.020	0.052
	CCA	-0.137	-0.126	0.076	-0.084	0.075	-0.128	-0.123	0.064	-0.035	0.059	-0.142	-0.129	0.074	-0.089	0.116	-0.129	-0.124	0.073	-0.038	0.078
IPTW	Media	-0.137	-0.134	0.068	-0.023	0.061	-0.128	-0.135	0.055	0.058	0.037	-0.142	-0.139	0.062	-0.021	0.058	-0.129	-0.140	0.058	0.090	0.056
	PAM	-0.137	-0.134	0.069	-0.022	0.059	-0.128	-0.134	0.056	0.051	0.035	-0.142	-0.140	0.062	-0.017	0.056	-0.129	-0.139	0.060	0.085	0.053
	WPAM	-0.137	-0.135	0.069	-0.019	0.060	-0.128	-0.135	0.056	0.055	0.036	-0.142	-0.140	0.063	-0.017	0.056	-0.129	-0.140	0.060	0.090	0.053
	Regressione	-0.137	-0.134	0.069	-0.021	0.060	-0.128	-0.134	0.056	0.052	0.036	-0.142	-0.140	0.063	-0.018	0.056	-0.129	-0.139	0.060	0.085	0.053
	Dati completi	-0.137	-0.134	0.069	-0.021	0.060	-0.128	-0.134	0.056	0.052	0.036	-0.142	-0.140	0.063	-0.018	0.056	-0.129	-0.139	0.060	0.085	0.053

Tabella 3.15: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la Simulazione 9. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Matchng	Dati completi	-0,086	-0,081	0,055	-0,063	0,057	-0,082	-0,079	0,050	-0,039	0,050	-0,101	-0,095	0,060	-0,055	0,063	-0,093	-0,075	0,055	-0,197	0,052
	CCA	-0,086	-0,070	0,063	-0,183	0,076	-0,082	-0,077	0,056	-0,067	0,062	-0,101	-0,087	0,066	-0,130	0,082	-0,093	-0,072	0,061	-0,232	0,071
	Medta	-0,086	-0,084	0,054	-0,023	0,058	-0,082	-0,079	0,045	-0,031	0,050	-0,101	-0,097	0,060	-0,037	0,066	-0,093	-0,076	0,043	-0,180	0,049
	PMM	-0,086	-0,083	0,064	-0,037	0,051	-0,082	-0,081	0,055	-0,017	0,039	-0,101	-0,098	0,070	-0,020	0,058	-0,093	-0,077	0,052	-0,168	0,036
	WPMM	-0,086	-0,085	0,063	-0,010	0,052	-0,082	-0,082	0,054	0,004	0,039	-0,101	-0,094	0,068	-0,068	0,060	-0,093	-0,078	0,052	-0,165	0,036
Stratificazione	Dati completi	-0,086	-0,072	0,038	-0,157	0,046	-0,082	-0,076	0,035	-0,077	0,034	-0,101	-0,079	0,036	-0,210	0,042	-0,093	-0,061	0,035	-0,345	0,041
	CCA	-0,086	-0,065	0,045	-0,247	0,061	-0,082	-0,077	0,041	-0,060	0,048	-0,101	-0,076	0,043	-0,214	0,056	-0,093	-0,065	0,043	-0,205	0,052
	Medta	-0,086	-0,073	0,037	-0,153	0,046	-0,082	-0,075	0,032	-0,086	0,032	-0,101	-0,079	0,035	-0,218	0,043	-0,093	-0,067	0,029	-0,283	0,037
	PMM	-0,086	-0,074	0,040	-0,132	0,044	-0,082	-0,077	0,034	-0,060	0,031	-0,101	-0,078	0,037	-0,228	0,042	-0,093	-0,069	0,031	-0,257	0,032
	WPMM	-0,086	-0,074	0,039	-0,136	0,044	-0,082	-0,077	0,033	-0,058	0,030	-0,101	-0,078	0,037	-0,224	0,042	-0,093	-0,070	0,031	-0,252	0,032
Overtate adjustment	Dati completi	-0,086	-0,073	0,039	-0,128	0,045	-0,082	-0,078	0,034	-0,047	0,031	-0,101	-0,078	0,037	-0,223	0,042	-0,093	-0,070	0,031	-0,253	0,033
	CCA	-0,086	-0,080	0,042	-0,071	0,045	-0,082	-0,082	0,039	-0,005	0,040	-0,101	-0,100	0,049	-0,001	0,045	-0,093	-0,088	0,045	-0,059	0,043
	Medta	-0,086	-0,081	0,034	-0,058	0,032	-0,082	-0,083	0,031	0,010	0,029	-0,101	-0,103	0,038	0,021	0,029	-0,093	-0,086	0,031	-0,080	0,030
	PMM	-0,086	-0,081	0,034	-0,052	0,032	-0,082	-0,083	0,031	0,013	0,029	-0,101	-0,103	0,038	0,023	0,029	-0,093	-0,088	0,032	-0,056	0,028
	WPMM	-0,086	-0,081	0,034	-0,054	0,032	-0,082	-0,083	0,031	0,011	0,029	-0,101	-0,103	0,038	0,022	0,029	-0,093	-0,088	0,032	-0,052	0,028
IPTW	Dati completi	-0,086	-0,081	0,034	-0,053	0,032	-0,082	-0,083	0,031	0,017	0,029	-0,101	-0,103	0,038	0,024	0,029	-0,093	-0,088	0,032	-0,054	0,028
	CCA	-0,086	-0,082	0,036	-0,042	0,036	-0,082	-0,082	0,046	-0,004	0,039	-0,101	-0,101	0,035	0,007	0,063	-0,093	-0,087	0,034	-0,071	0,030
	Medta	-0,086	-0,074	0,060	-0,140	0,069	-0,082	-0,083	0,052	0,015	0,055	-0,101	-0,107	0,066	0,064	0,087	-0,093	-0,082	0,061	-0,124	0,067
	PMM	-0,086	-0,083	0,054	-0,036	0,055	-0,082	-0,085	0,046	0,037	0,043	-0,101	-0,094	0,033	-0,008	0,047	-0,093	-0,085	0,044	-0,083	0,045
	WPMM	-0,086	-0,084	0,056	-0,018	0,054	-0,082	-0,085	0,047	0,035	0,038	-0,101	-0,093	0,035	-0,072	0,045	-0,093	-0,083	0,046	-0,113	0,038
Regressione	Dati completi	-0,086	-0,084	0,056	-0,017	0,054	-0,082	-0,085	0,047	0,038	0,038	-0,101	-0,093	0,034	-0,073	0,045	-0,093	-0,084	0,047	-0,102	0,039
	CCA	-0,086	-0,085	0,056	-0,011	0,055	-0,082	-0,085	0,047	0,040	0,038	-0,101	-0,093	0,035	-0,074	0,045	-0,093	-0,083	0,047	-0,112	0,038

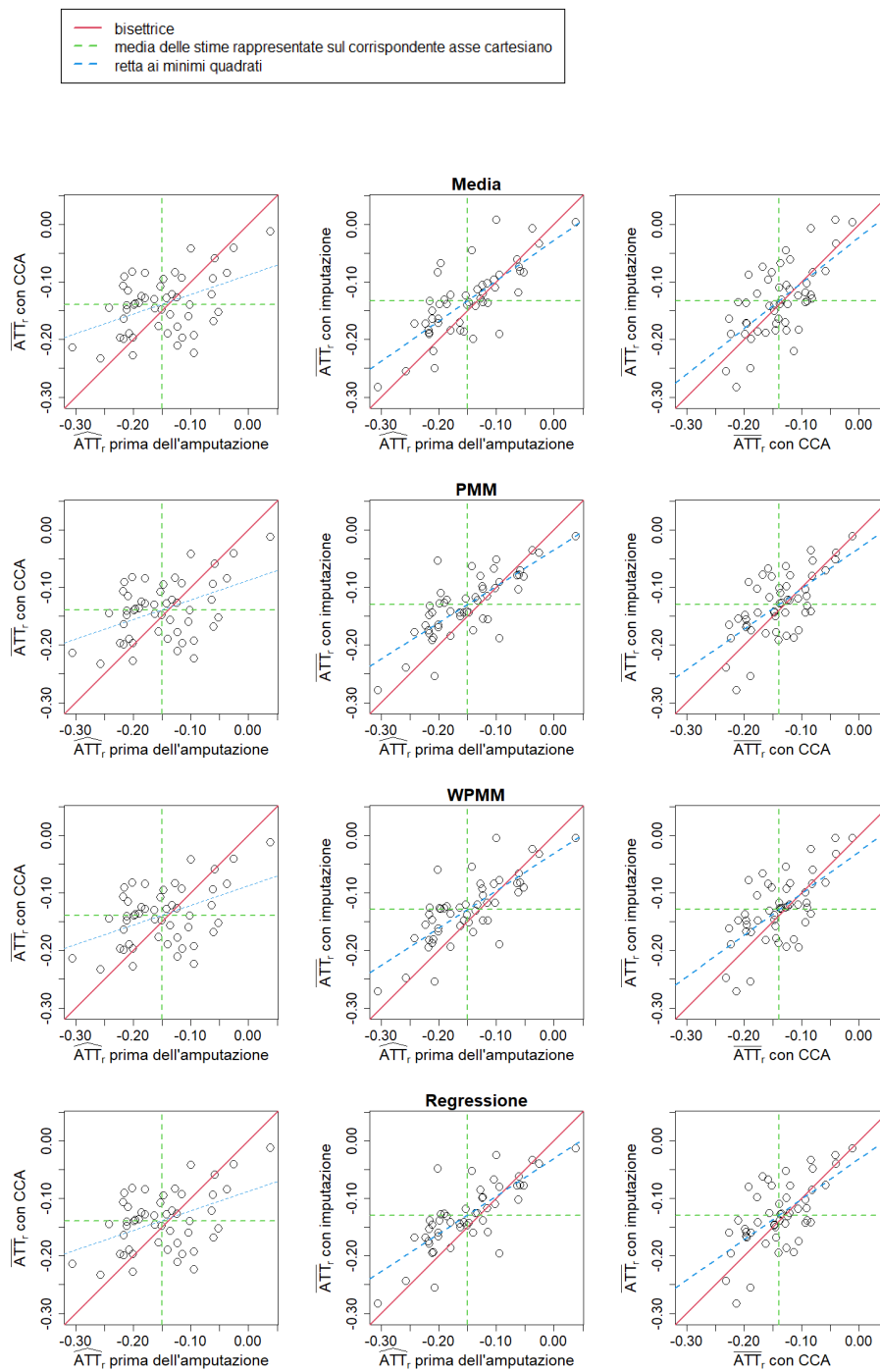


Figura 3.5: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 1.

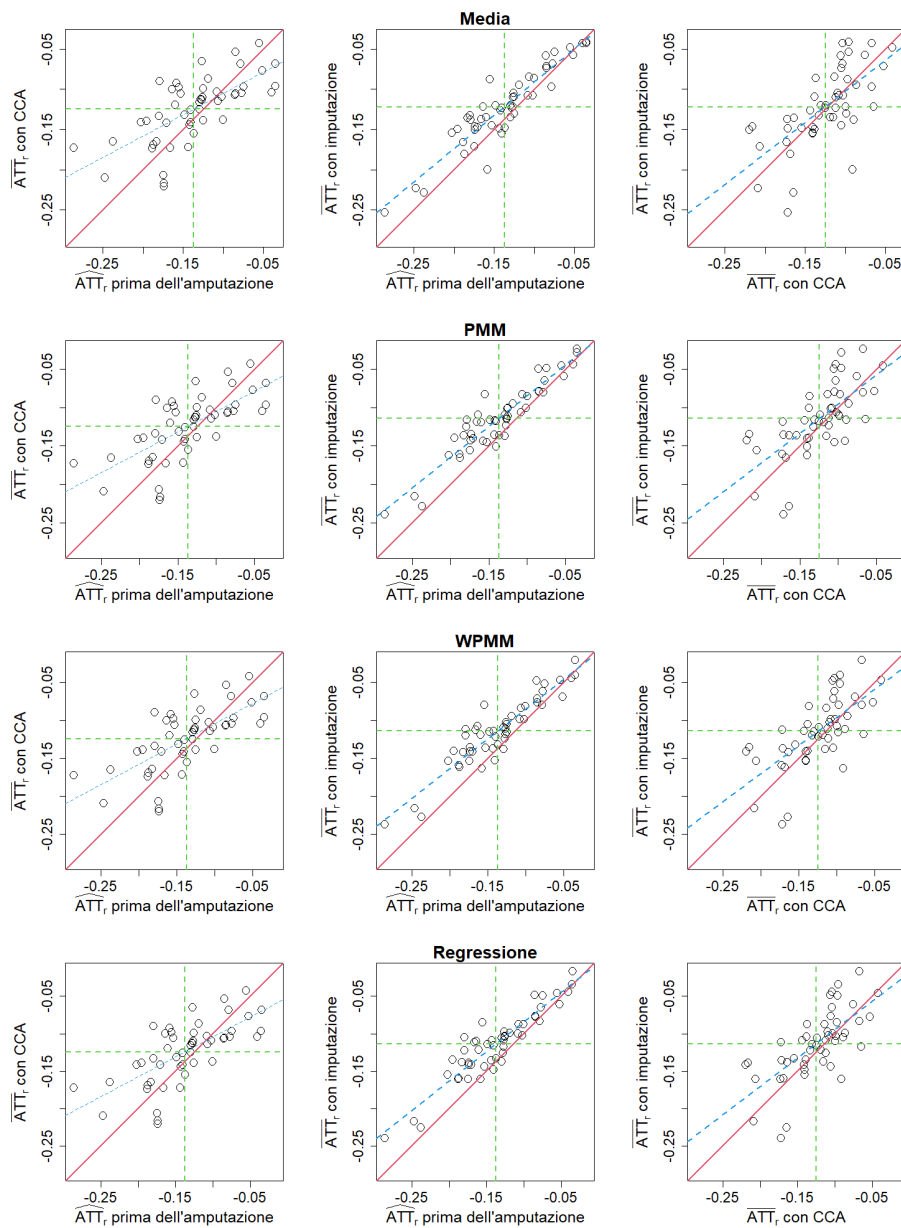


Figura 3.6: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 1. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

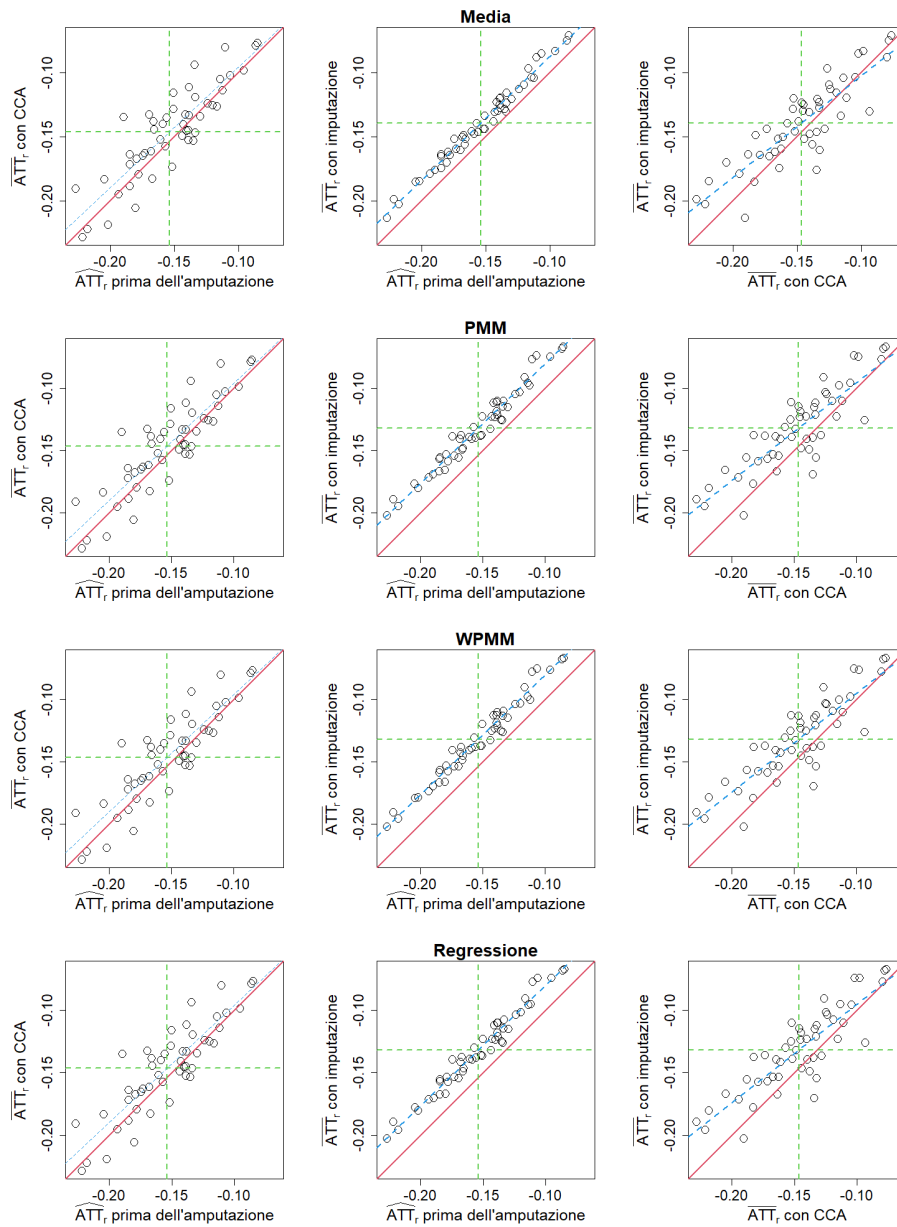


Figura 3.7: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 1.

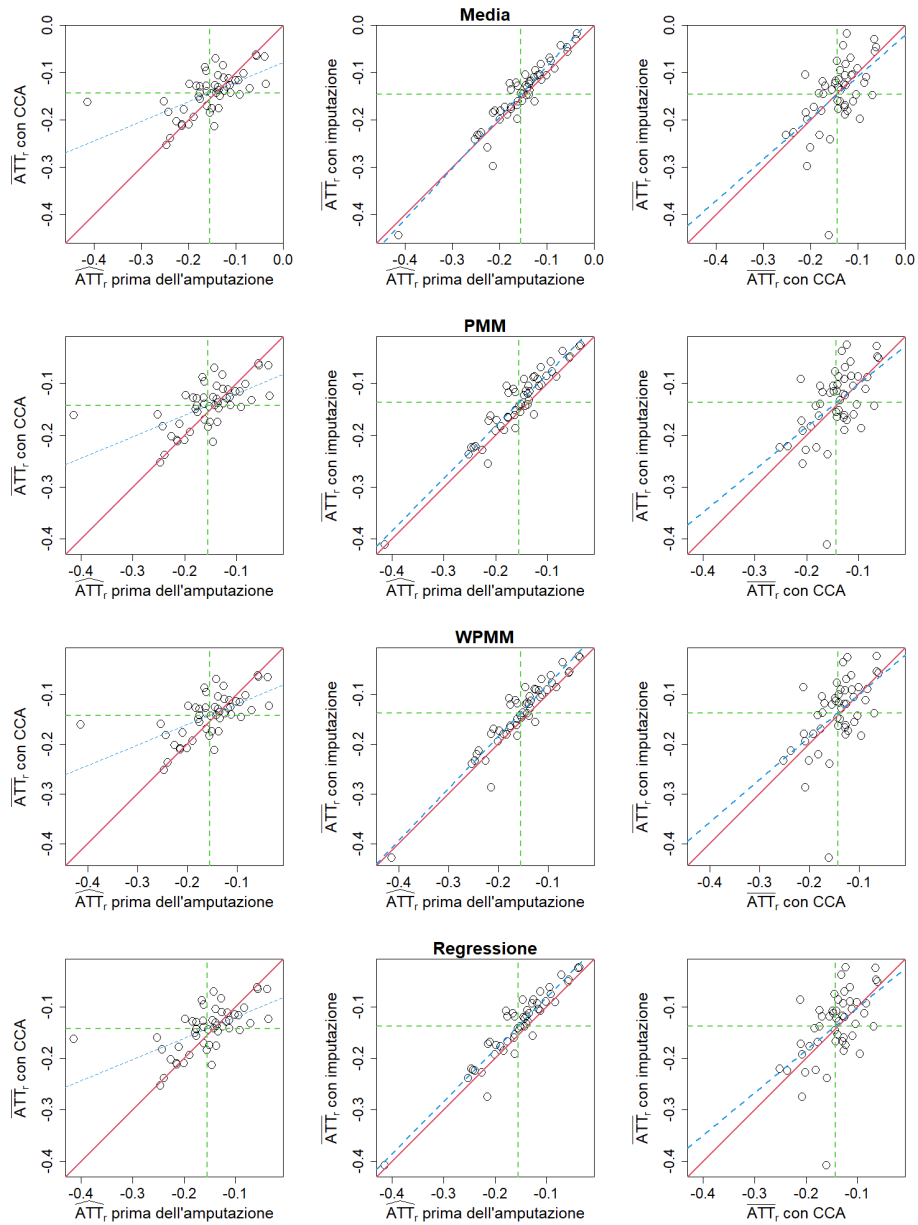


Figura 3.8: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 1.

Capitolo 4

Alcuni scenari realistici

4.1 Introduzione

Nel capitolo precedente si è eseguito un primo studio empirico in cui il disegno della simulazione è stato ideato appositamente per valutare gli effetti che la tipologia delle variabili con dati mancanti e il loro ruolo nello schema causale possono avere sulle prestazioni dei metodi di aggiustamento del confondimento basati sul *propensity score* combinati con le tecniche per trattare i dati mancanti. Successivamente si è voluto eseguire un secondo studio di simulazione per approfondire il confronto tra le combinazioni di tali metodi alla luce di quanto riscontrato nel primo studio. A tal fine si è definito uno scenario più complesso e che si avvicini maggiormente ad una situazione reale. In particolare, si è deciso di aumentare il numero di variabili con dati mancanti e di includere contemporaneamente variabili confondenti, predittori del trattamento e predittori della risposta.

4.2 Disegno della simulazione

Per questo secondo studio empirico si è considerato un campione di numerosità invariata e pari a 1000 costituito da 5 variabili concomitanti $X = (X_1, X_2, X_3, X_4, X_5)$. Tra di esse quattro sono le stesse considerate nello studio del Capitolo 3, mentre una rappresenta una variabile continua aggiuntiva. Le distribuzioni delle covariate sono quindi:

- $X_1 \sim N(1, 4)$;
- $X_2 \sim N(5, 3)$;
- $X_3 \sim N(3, 1)$;
- $X_4 \sim \text{Be}(0.6)$;
- X_5 assume i valori $(1, 2, 3, 4)$ con probabilità $(0.3, 0.2, 0.4, 0.1)$.

Anche le procedure di generazione della variabile risposta Y e del trattamento T rimangono invariate rispetto al capitolo precedente. In particolare, il trattamento è generato a partire dalle covariate secondo i 4 scenari già descritti a pagina 3.2, ad ognuno dei quali è associato il medesimo modello definito nel precedente studio. Si riportano in seguito i modelli per ogni scenario dopo aver aggiornato le variabili concomitanti:

Scenario A $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_4 + \alpha_4 X_5,$

Scenario B $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_4 + \alpha_4 X_5 + \alpha_5 X_2^2 + \alpha_6 X_5^2,$

Scenario C $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_4 + \alpha_4 X_5 + \alpha_7 X_1 X_4 + \alpha_8 X_2 X_5,$

Scenario D $\text{logit}(\pi_T) = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_4 + \alpha_4 X_5 + \alpha_5 X_2^2 + \alpha_6 X_5^2 + \alpha_7 X_1 X_4 + \alpha_8 X_2 X_5,$

con $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8) = (0.8, -0.25, 0.6, 0.4, 0.05, -0.25, 0.4, -0.15)$ e π_T pari alla probabilità che si verifichi l'evento $T = 1$. Nel modello della risposta, invece, uno dei predittori utilizzati nello studio del Capitolo 3 è stato sostituito con la variabile continua aggiuntiva. Il modello risulta quindi:

$$\text{logit}(\pi_Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_3 + \beta_3 X_4 + \beta_4 X_5 + \gamma T,$$

con $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (-1, 0.3, -0.25, 0.35, 0.5)$, $\gamma = -0.7$ e π_Y pari alla probabilità che si verifichi l'evento $Y = 1$. Le distribuzioni delle prevalenze di soggetti con $Y = 1$ e $T = 1$ nei diversi campioni generati per ogni scenario sono riportati nella Figura 4.2.

Dall'assetto delle variabili definito per lo studio di simulazione si ottiene un nuovo schema causale, il quale è mostrato nella Figura 4.1.

Per la fase di amputazione dei dati si è utilizzata la medesima procedura già descritta nel precedente capitolo. In questo caso, però, si è stabilito che i valori mancanti siano presenti in 3 delle covariate, ovvero X_1 , X_3 e X_4 . Pertanto per generare i dati mancanti si sono prima definiti tutti i possibili *pattern* con cui essi possono manifestarsi nelle tre diverse variabili. Ad ognuno di essi è stata poi associata la frequenza con cui si verifica nel dataset con valori mancanti. Nello specifico, ad ogni *pattern* è stata attribuita la stessa frequenza. Successivamente è stato possibile generare i dati mancanti

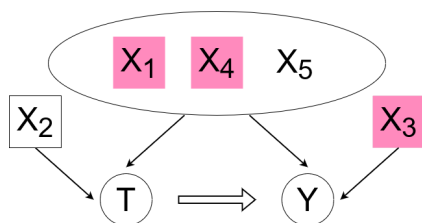


Figura 4.1: Diagramma causale della simulazione. Le variabili su sfondo rosso presentano valori mancanti.

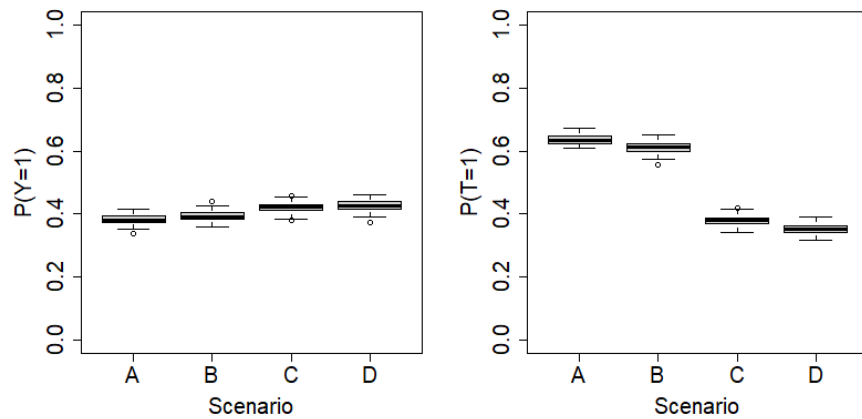
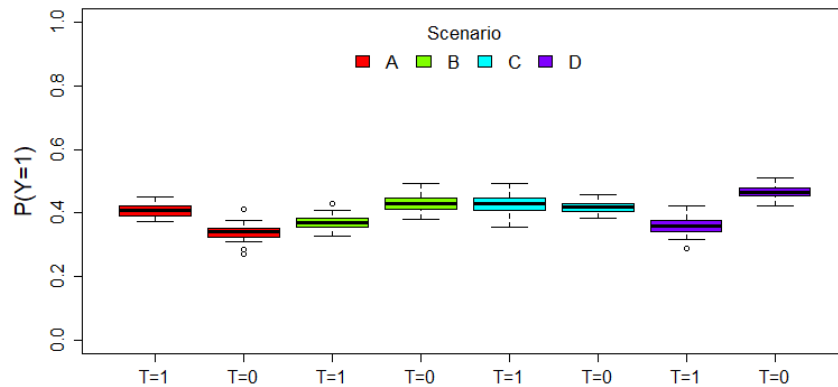
(a) Prevalenza di $Y = 1$ (b) Prevalenza di $T = 1$ (c) Prevalenza di $Y = 1$ nei sottogruppi dei trattati e dei non trattati

Figura 4.2: Distribuzione della prevalenza di soggetti con $Y = 1$ e $T = 1$ nei diversi campioni generati per i 4 scenari.

con la funzione 'ampute' del pacchetto 'mice' del software R. Il meccanismo previsto per i dati mancanti è di tipo MAR, e il modello con cui essi vengono generati coinvolge con ugual peso le variabili concomitanti osservate. Come stabilito nello studio precedente, le variabili trattamento e risposta sono escluse dal meccanismo di generazione dei dati mancanti. Si è definita una proporzione di dati mancanti rispetto alle osservazioni pari a 0.5. Ciò

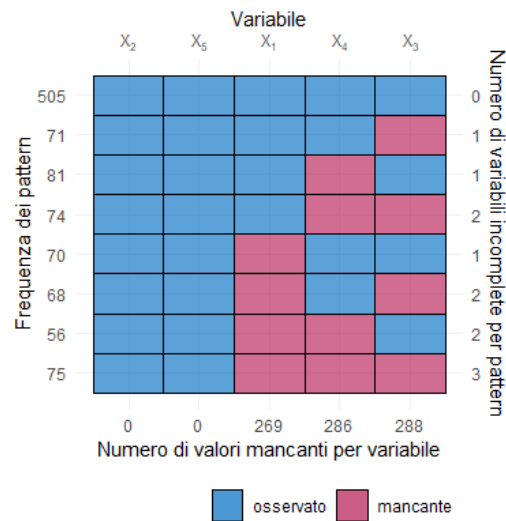


Figura 4.3: Rappresentazione grafica della matrice dei *pattern* dei dati mancanti nel primo dataset amputato generato.

significa che il 50% delle osservazioni presenta dati mancanti secondo uno dei *pattern* possibili. Nella Figura 4.3 si mostra graficamente la matrice dei *pattern* dei dati mancanti associata al primo dataset amputato generato nella simulazione.

Per quanto riguarda la fase di trattamento dei dati mancanti e quella di stima dell'ATT non sono state fatte modifiche rispetto alle procedure svolte nello studio del capitolo precedente e descritte nei Paragrafi 3.4 e 3.3. Inoltre anche in questa simulazione sono stati generati 50 dataset iniziali e per ognuno di essi si sono ottenuti 10 diversi dataset amputati su cui si sono svolte poi le analisi.

4.3 Risultati

Le misure utilizzate per la valutazione dei risultati in ogni scenario e per ogni combinazione di metodi di aggiustamento del confondimento e di

trattamento dei dati mancanti sono riportate nella Tabella 4.1. In aggiunta, nelle Figure 4.4 – 4.19 sono rappresentati i grafici di dispersione analoghi a quelli descritti nel capitolo precedente.

A proposito della variabilità campionaria delle stime, si conferma quanto osservato nel precedente studio di simulazione. Nel presente studio i coefficienti di variazione dei veri valori di ATT per i 4 scenari sono 0.006, 0.008, 0.008 e 0.009. Inoltre il rapporto tra la deviazione standard dei veri valori di ATT e quella dei valori \overline{ATT}_r va da un minimo di 0.0095 e un massimo di 0.0461. La variabilità campionaria risulta quindi trascurabile.

Anche in questa simulazione talvolta non è stato possibile stimare l'ATT tramite il metodo della stratificazione basata sul *propensity score* a causa della presenza di strati costituiti unicamente da soggetti trattati o non trattati. In questo caso, però, tale problema si è verificato unicamente quando è stata utilizzata la tecnica della *complete case analysis*. Più precisamente, il problema è avvenuto negli Scenari A, C e D rispettivamente in 2, 9 e 6 delle 500 repliche totali eseguite. Ciò conferma quanto osservato nel Capitolo 3, ovvero che la combinazione tra la stratificazione e il metodo della *complete case analysis* potrebbe non essere una buona scelta per stimare l'ATT.

4.4 Discussione

I risultati che si ottengono per questo secondo studio di simulazione sono per alcuni aspetti diversi da quelli che si osservano per lo studio precedente, ma le conclusioni che si traggono sono in generale le stesse.

A prima vista si nota un netto peggioramento della distorsione relativa delle stime per tutti gli scenari e tutti i metodi utilizzati. Si osserva poi che, in questo caso, il metodo di gestione dei dati mancanti gioca un ruolo più impor-

tante. Le differenze tra le diverse tecniche di applicazione dell'imputazione multipla (PMM, WPMM e imputazione tramite regressione) continuano ad essere minime e molto inferiori a quelle che si riscontrano tra i metodi di aggiustamento del confondimento basati sul *propensity score*. Tuttavia, in questa simulazione ci sono evidenti differenze tra i metodi di imputazione e la *complete case analysis*: con quest'ultima si ottengono stime molto meno distorte, indipendentemente dallo scenario e dal metodo di applicazione del *propensity score* utilizzato. La differenza tra le distorsioni è leggermente meno accentuata quando si utilizza la stratificazione, ma comunque presente. Anche le stime dell'ATT che si ottengono con la tecnica dell'imputazione singola della media, seppur fortemente distorte, sono generalmente più accurate rispetto a quelle ottenute con l'imputazione multipla. Si notano comunque delle differenze in termini di distorsione a seconda del metodo di trattamento del confondimento che si impiega, ma tali differenze sono irrilevanti se confrontate con quelle che si osservano tra la *complete case analysis* e gli altri metodi di gestione dei dati mancanti. Inoltre, non si osserva una netta superiorità di un metodo di applicazione del *propensity score* in tutte le circostanze.

Alla luce di quanto osservato nel Capitolo 3, i problemi di distorsione che si riscontrano in questo studio sono molto probabilmente dovuti alla presenza di due variabili parzialmente osservate tra i confondenti. I risultati, infatti, sono in parte simili a quelli ottenuti per le prime tre simulazioni dello studio precedente, in cui la variabile con dati mancanti era proprio un confondente. In questo caso, quindi, l'imputazione non sembra essere efficace.

Al contrario di quanto avviene per la distorsione relativa, l'entità degli errori quadratici medi è molto simile a quella osservata nel Capitolo 3 per tutte le combinazioni di metodi considerati. Inoltre, anche in questo secondo

studio di simulazione gli errori quadratici medi sono spesso superiori per le stime prodotte tramite la *complete case analysis* rispetto ai metodi di imputazione. Ciò può essere spiegato dal fatto che nel primo caso la numerosità campionaria è la metà rispetto a quella del secondo. Allo stesso modo si notano stime lievemente più sparse nei grafici di dispersione.

In generale, tra i diversi metodi di trattamento del confondimento, la dispersione delle stime è inferiore per il covariate adjustment e superiore per il matching. Un'eccezione è lo Scenario C, in cui l'errore quadratico medio che si ottiene combinando l'IPTW e la *complete case analysis* è particolarmente elevato. Lo stesso fenomeno si è osservato nel primo studio condotto. Infatti, anche in questo caso, nello Scenario C l'IPTW ha prodotto una stima dell'ATT molto distante dal vero valore. Si può concludere che la dispersione delle stime sia poco influenzata dalla presenza di dati mancanti e, di conseguenza, dal metodo con cui essi vengono trattati, al contrario di quanto accade per la distorsione. A conferma di ciò, le dispersioni delle stime ottenute sui dataset non amputati e su quelli ottenuti dopo l'imputazione sono simili tra di loro, anche se la maggior parte delle volte sono leggermente inferiori nel primo caso.

Questo secondo studio empirico non porta a conclusioni differenti in merito ai diversi metodi basati sul *propensity score*. Tuttavia mette in luce nuovi elementi sui metodi di gestione dei dati mancanti. In particolare, in presenza di variabili con dati mancanti tra i confondenti la *complete case analysis* sembra essere il metodo che porta a minor distorsione, e ciò è ancora più evidente se le variabili incomplete sono più di una. In questa circostanza l'utilizzo dell'imputazione dei dati mancanti conduce invece a stime distorte. Infine, come già osservato nel capitolo precedente, le differenze tra le tecniche di imputazione multipla sono minime, se non assenti.

Tabella 4.1: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{rel}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A				Scenario B				Scenario C				Scenario D							
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{rel}$	\overline{EQM}
Matching	Dati completi	-0.151	-0.153	0.078	0.014	0.000	-0.146	-0.146	0.067	-0.001	0.049	-0.150	-0.146	0.073	-0.030	0.049	-0.145	-0.142	0.062	-0.020	0.048
	CCA	-0.151	-0.150	0.098	-0.012	0.095	-0.146	-0.147	0.083	0.010	0.082	-0.150	-0.147	0.093	-0.025	0.088	-0.145	-0.141	0.082	-0.025	0.079
	Media	-0.151	-0.112	0.068	-0.259	0.068	-0.146	-0.123	0.056	-0.154	0.055	-0.150	-0.103	0.057	-0.316	0.067	-0.145	-0.116	0.051	-0.195	0.054
	PMM	-0.151	-0.078	0.073	-0.482	0.079	-0.146	-0.098	0.063	-0.326	0.055	-0.150	-0.077	0.065	-0.487	0.077	-0.145	-0.084	0.061	-0.418	0.064
	WPMM	-0.151	-0.081	0.073	-0.467	0.077	-0.146	-0.099	0.063	-0.320	0.054	-0.150	-0.076	0.066	-0.468	0.078	-0.145	-0.084	0.060	-0.419	0.064
Stratificazione	Dati completi	-0.151	-0.140	0.059	-0.078	0.041	-0.146	-0.148	0.050	0.016	0.039	-0.150	-0.124	0.050	-0.178	0.042	-0.145	-0.140	0.045	-0.035	0.035
	CCA	-0.151	-0.138	0.077	-0.091	0.074	-0.146	-0.140	0.066	-0.037	0.063	-0.150	-0.125	0.069	-0.171	0.070	-0.145	-0.135	0.064	-0.070	0.057
	Media	-0.151	-0.096	0.050	-0.367	0.065	-0.146	-0.101	0.042	-0.305	0.053	-0.150	-0.085	0.042	-0.482	0.070	-0.145	-0.091	0.039	-0.369	0.057
	PMM	-0.151	-0.073	0.050	-0.521	0.082	-0.146	-0.096	0.043	-0.341	0.054	-0.150	-0.071	0.043	-0.528	0.081	-0.145	-0.088	0.041	-0.390	0.058
	WPMM	-0.151	-0.072	0.050	-0.521	0.082	-0.146	-0.097	0.043	-0.337	0.053	-0.150	-0.071	0.043	-0.529	0.081	-0.145	-0.088	0.041	-0.391	0.058
Covariate adjustment	Dati completi	-0.151	-0.155	0.040	0.022	0.031	-0.146	-0.152	0.035	0.041	0.029	-0.150	-0.148	0.038	-0.015	0.030	-0.145	-0.148	0.036	0.021	0.022
	CCA	-0.151	-0.154	0.055	0.019	0.032	-0.146	-0.149	0.049	0.023	0.047	-0.150	-0.148	0.055	-0.020	0.051	-0.145	-0.144	0.052	-0.007	0.046
	Media	-0.151	-0.109	0.037	-0.282	0.050	-0.146	-0.108	0.034	-0.239	0.042	-0.150	-0.093	0.035	-0.381	0.060	-0.145	-0.095	0.034	-0.341	0.051
	PMM	-0.151	-0.097	0.038	-0.357	0.058	-0.146	-0.098	0.034	-0.329	0.050	-0.150	-0.083	0.036	-0.446	0.068	-0.145	-0.083	0.034	-0.426	0.062
	WPMM	-0.151	-0.097	0.038	-0.360	0.058	-0.146	-0.098	0.034	-0.330	0.050	-0.150	-0.083	0.036	-0.446	0.068	-0.145	-0.083	0.034	-0.428	0.063
IPTW	Dati completi	-0.151	-0.156	0.062	0.032	0.055	-0.146	-0.152	0.049	0.044	0.043	-0.150	-0.168	0.056	0.120	0.083	-0.145	-0.139	0.051	-0.041	0.041
	CCA	-0.151	-0.150	0.084	-0.010	0.089	-0.146	-0.145	0.067	-0.003	0.066	-0.150	-0.146	0.075	-0.030	0.126	-0.145	-0.139	0.070	-0.043	0.072
	Media	-0.151	-0.113	0.057	-0.251	0.063	-0.146	-0.116	0.043	-0.204	0.046	-0.150	-0.105	0.046	-0.299	0.058	-0.145	-0.107	0.039	-0.263	0.048
	PMM	-0.151	-0.087	0.066	-0.425	0.073	-0.146	-0.098	0.048	-0.330	0.054	-0.150	-0.074	0.049	-0.507	0.078	-0.145	-0.083	0.042	-0.427	0.064
	WPMM	-0.151	-0.087	0.066	-0.424	0.073	-0.146	-0.098	0.048	-0.326	0.053	-0.150	-0.074	0.050	-0.507	0.078	-0.145	-0.083	0.042	-0.427	0.064
Regression	-0.151	-0.087	0.066	-0.428	0.074	-0.146	-0.099	0.049	-0.323	0.052	-0.150	-0.073	0.050	-0.512	0.079	-0.145	-0.083	0.042	-0.429	0.064	

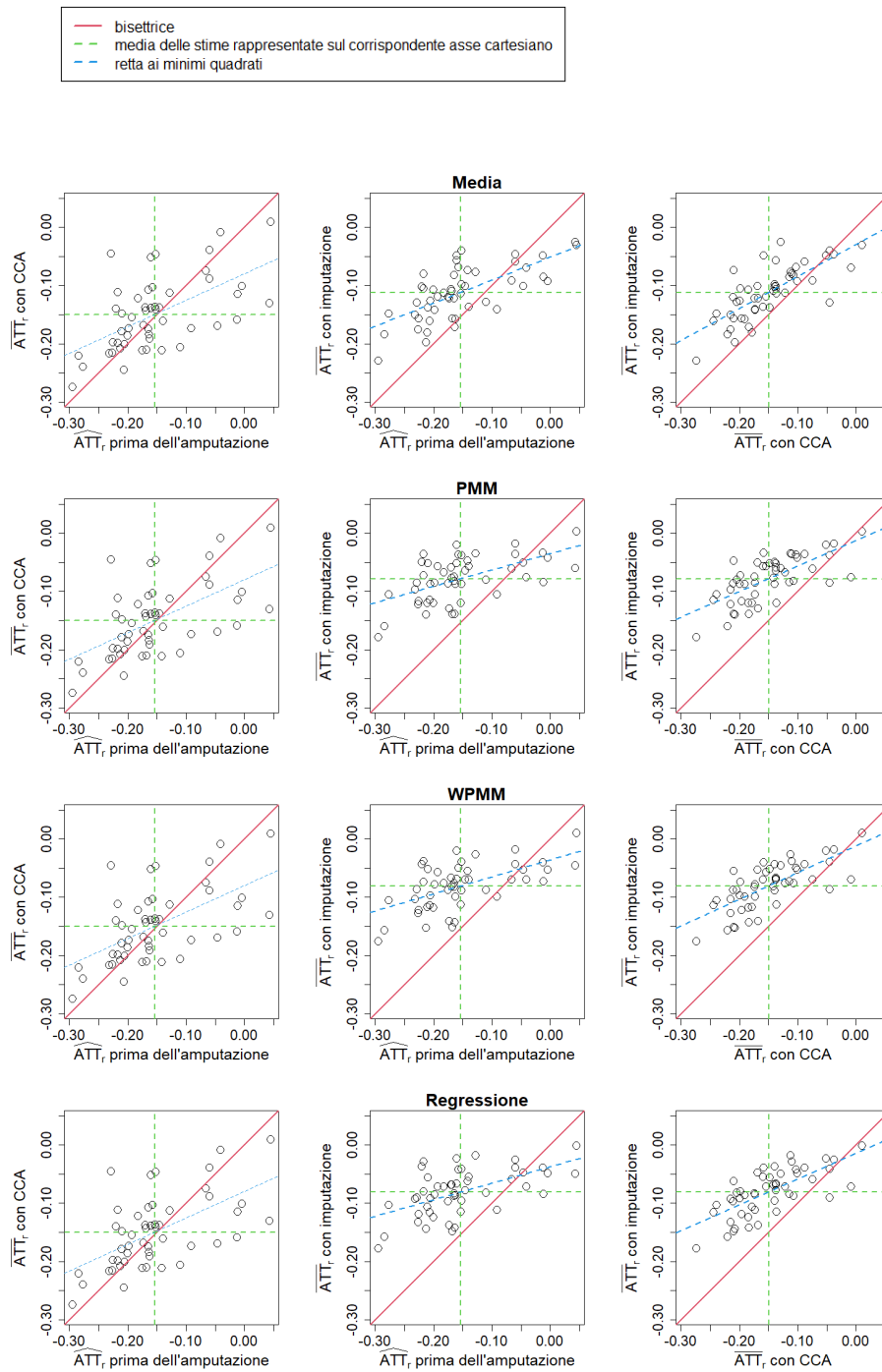


Figura 4.4: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A.

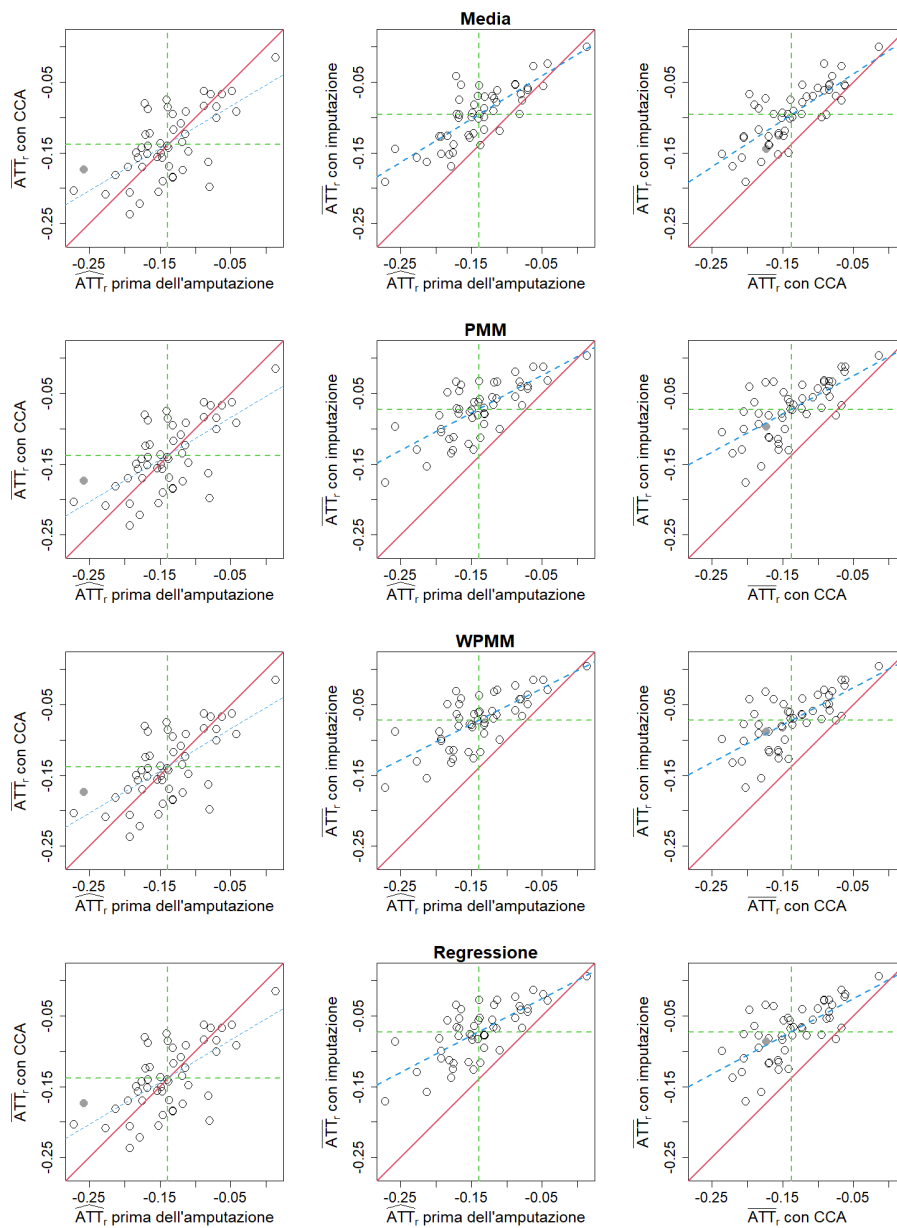


Figura 4.5: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

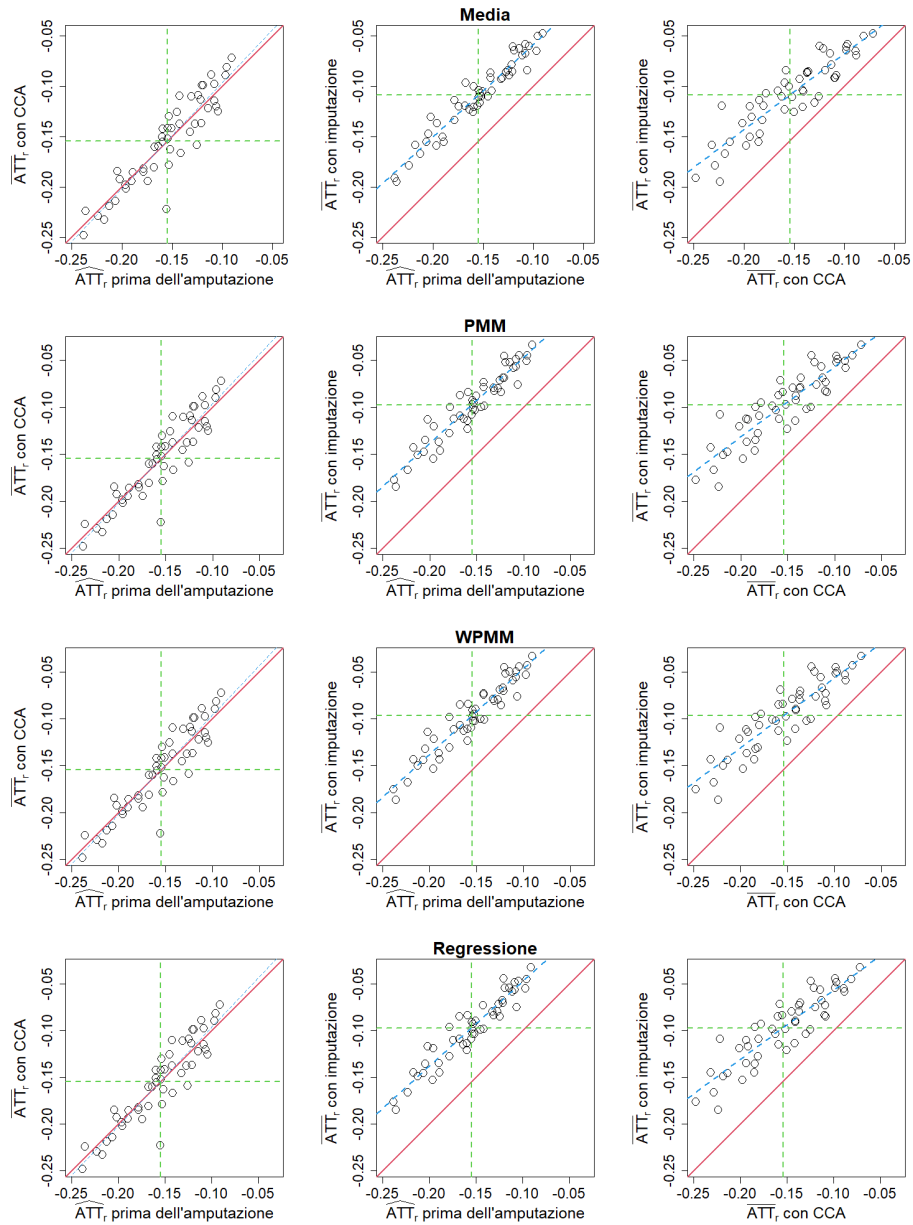


Figura 4.6: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A.

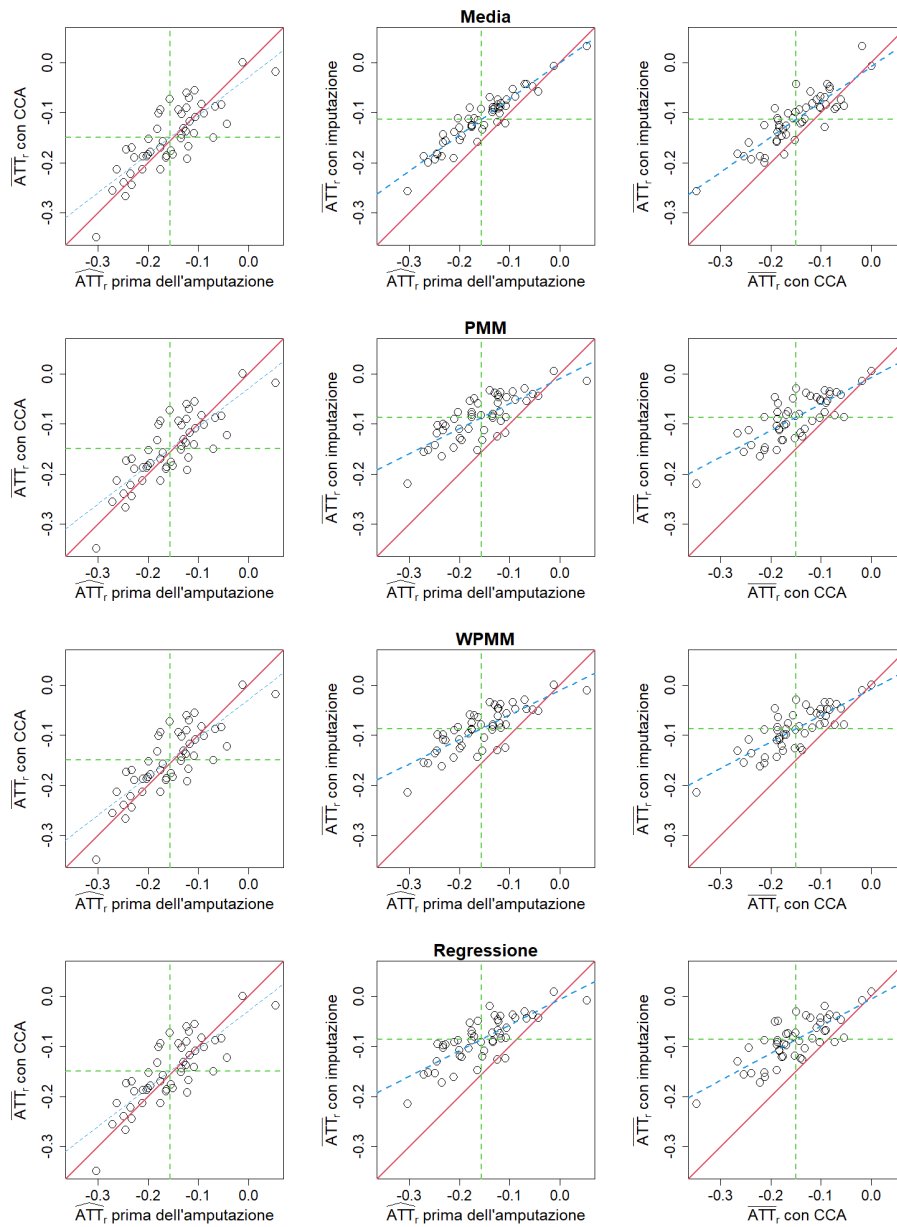


Figura 4.7: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A.

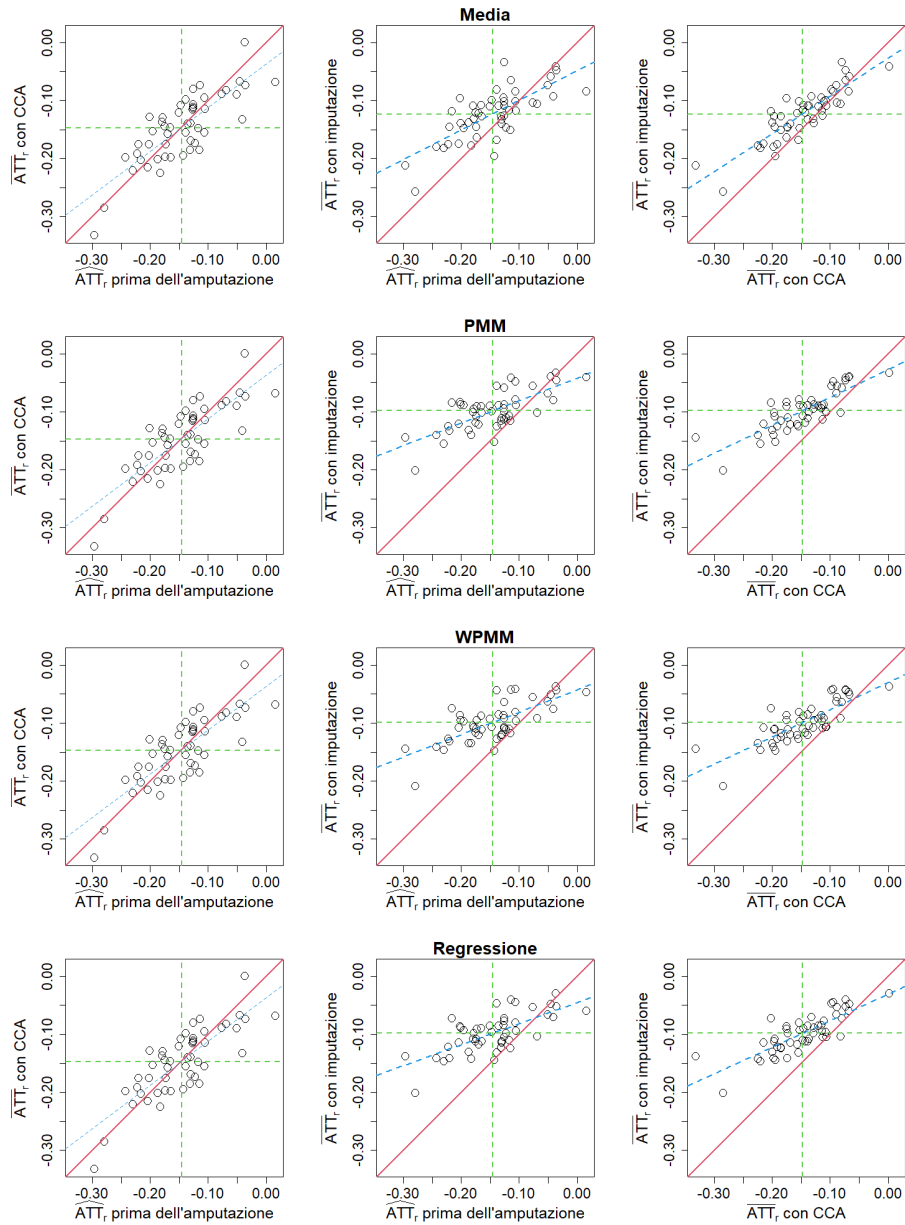


Figura 4.8: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B.

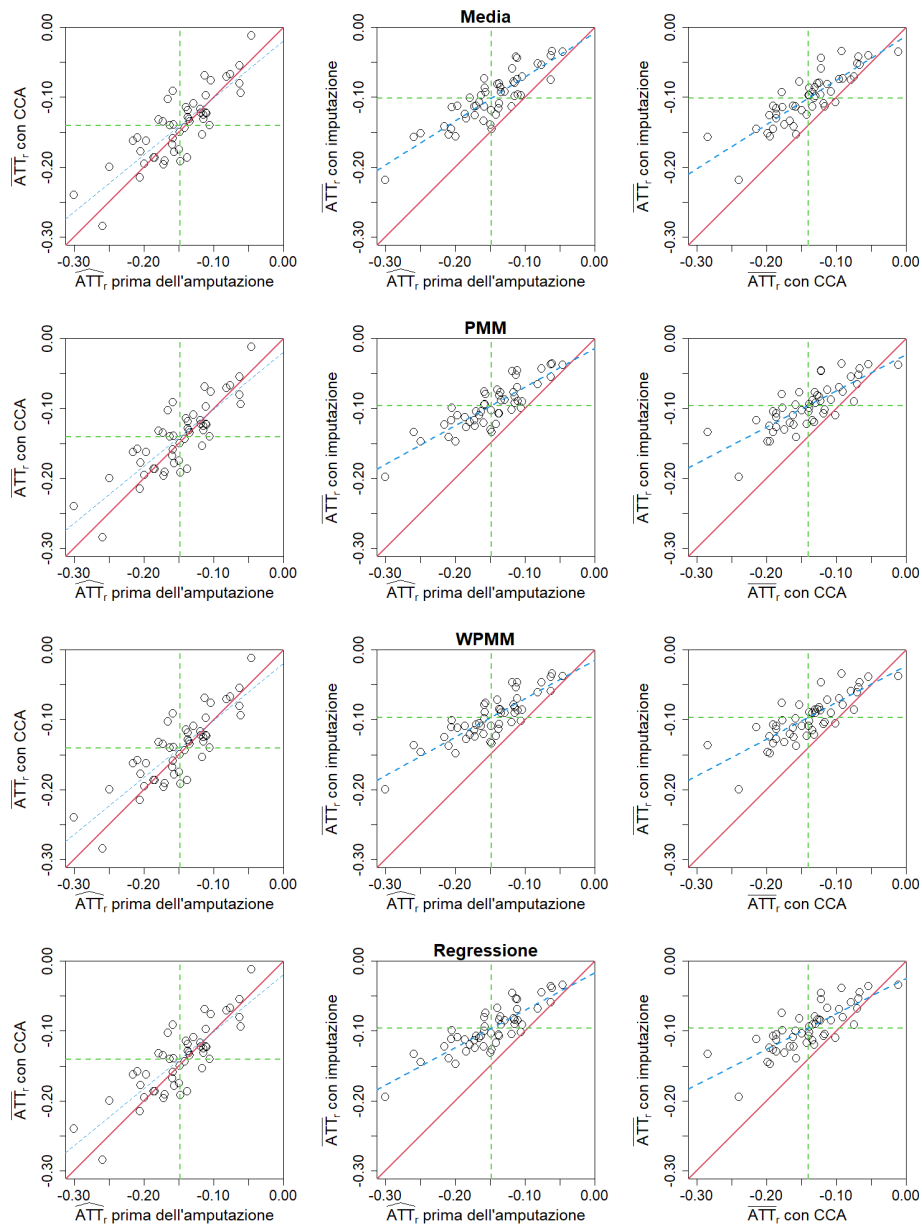


Figura 4.9: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B.

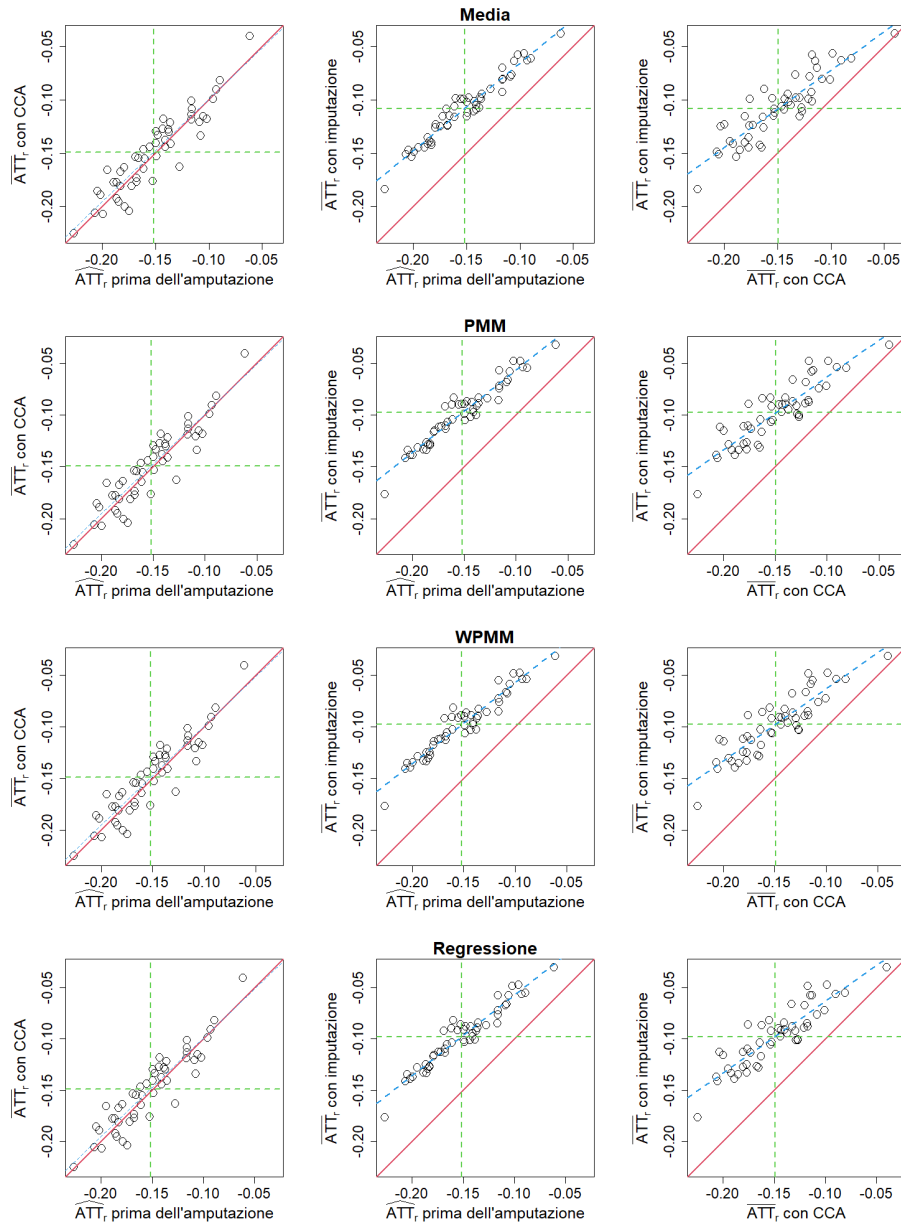


Figura 4.10: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B.

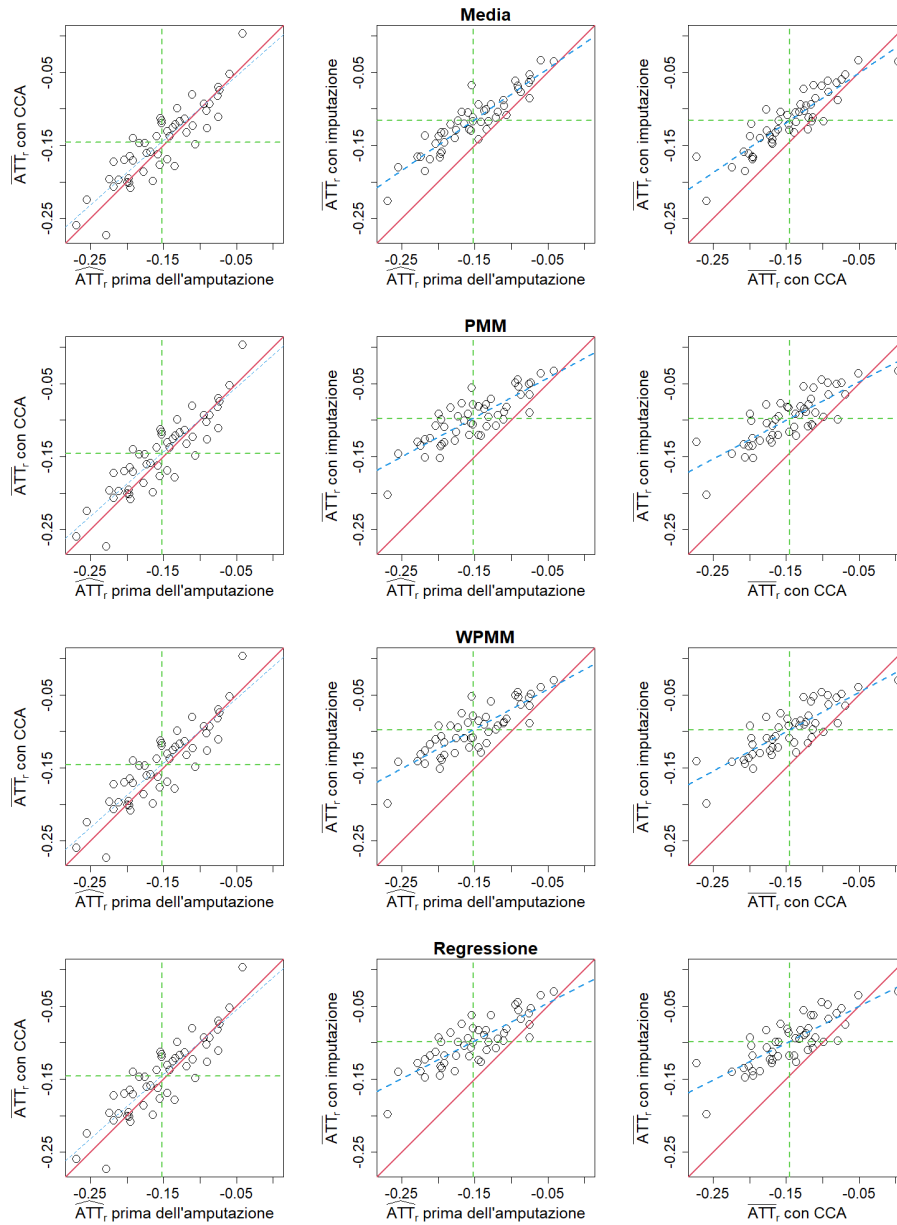


Figura 4.11: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B.

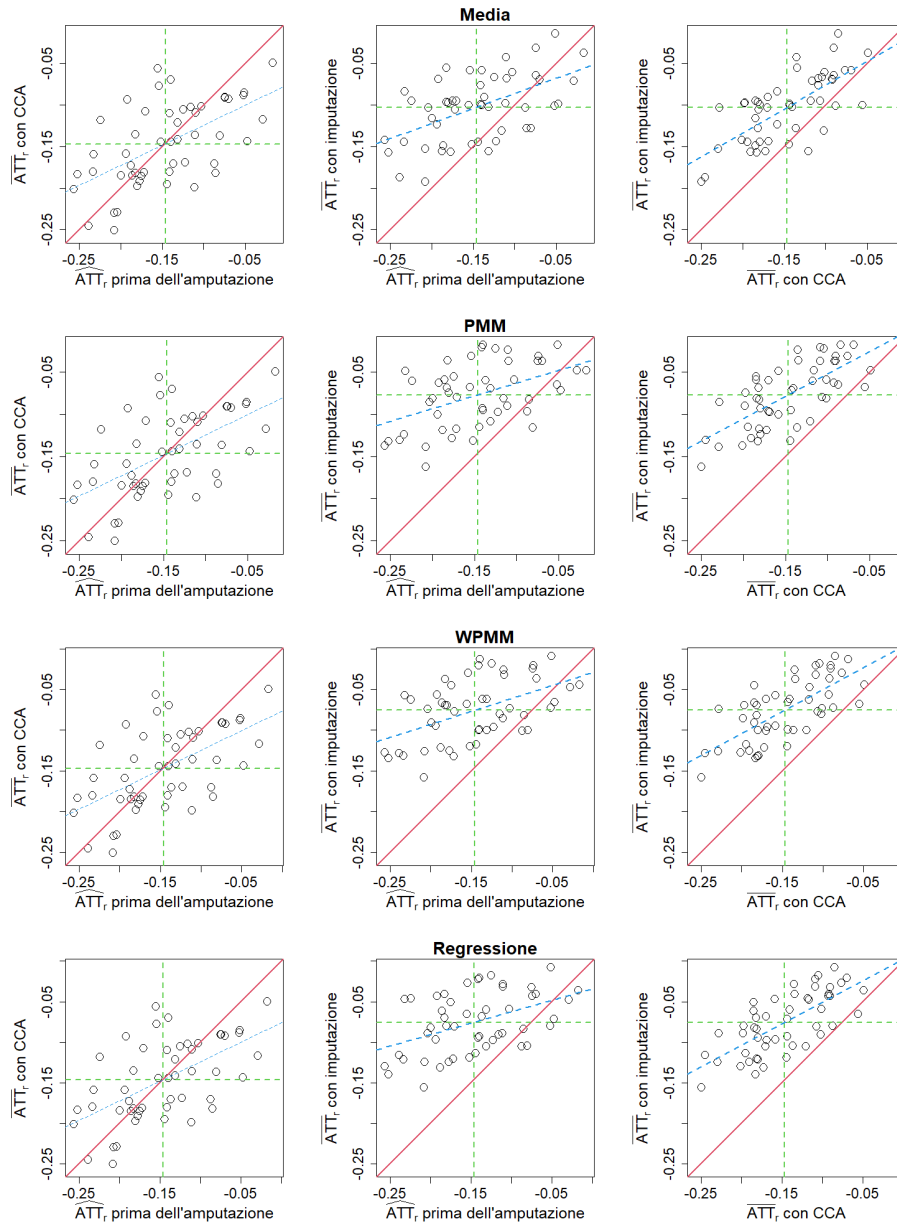


Figura 4.12: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C.

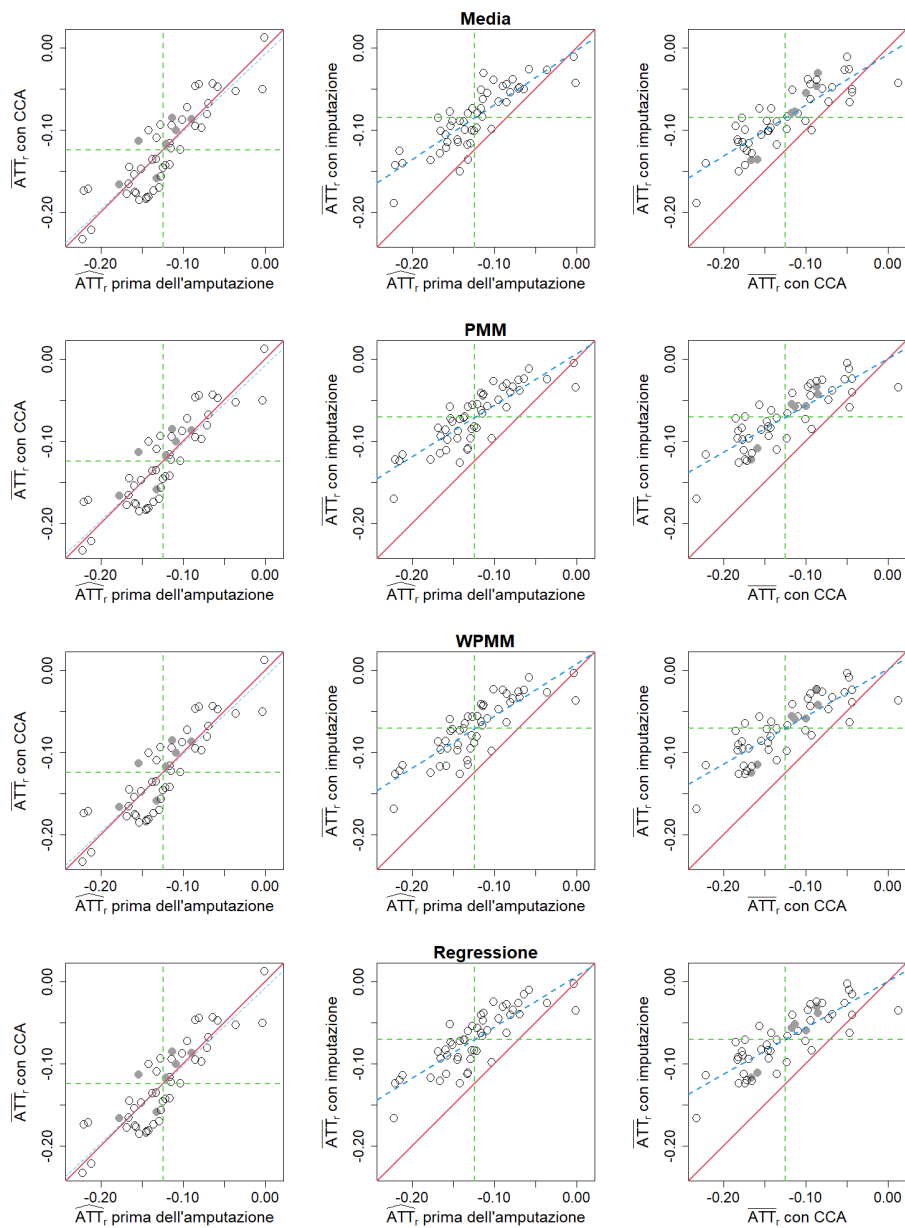


Figura 4.13: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

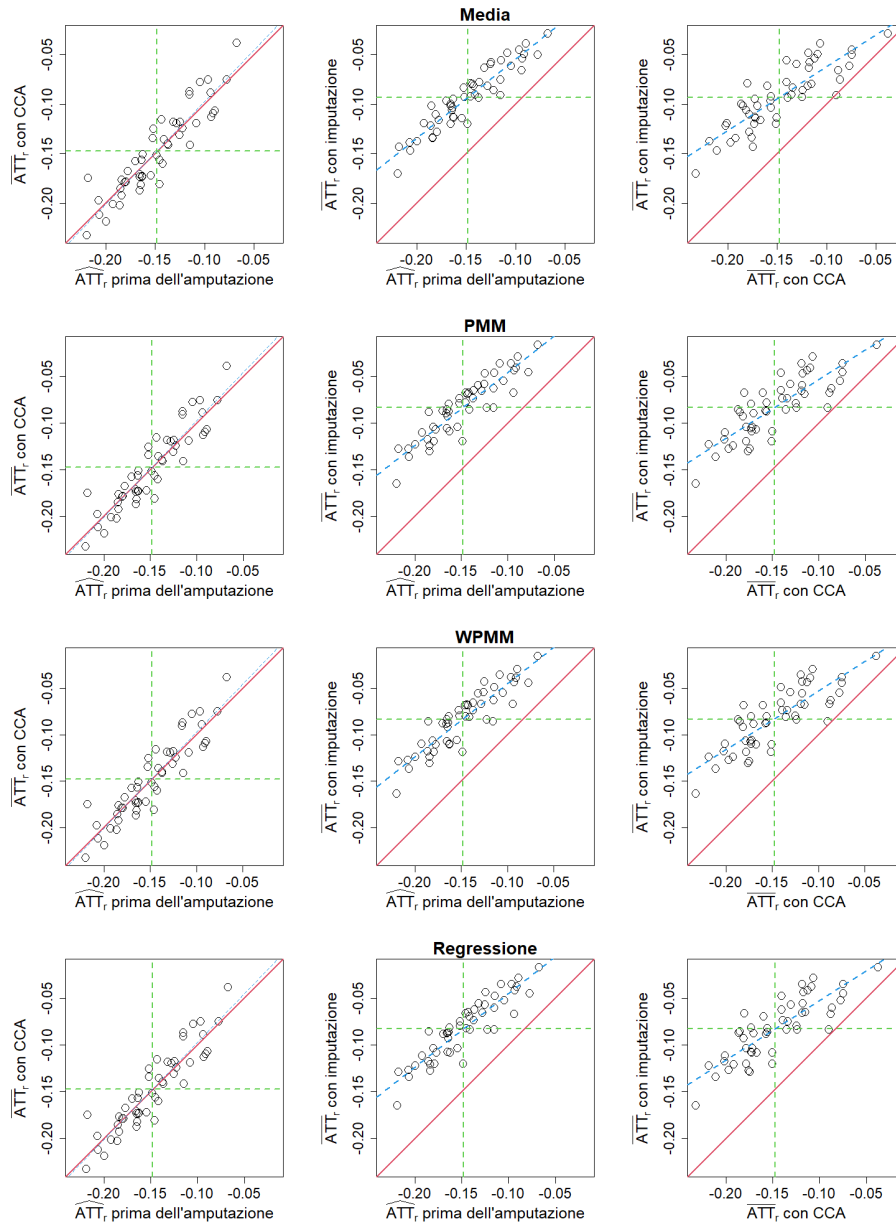


Figura 4.14: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C.

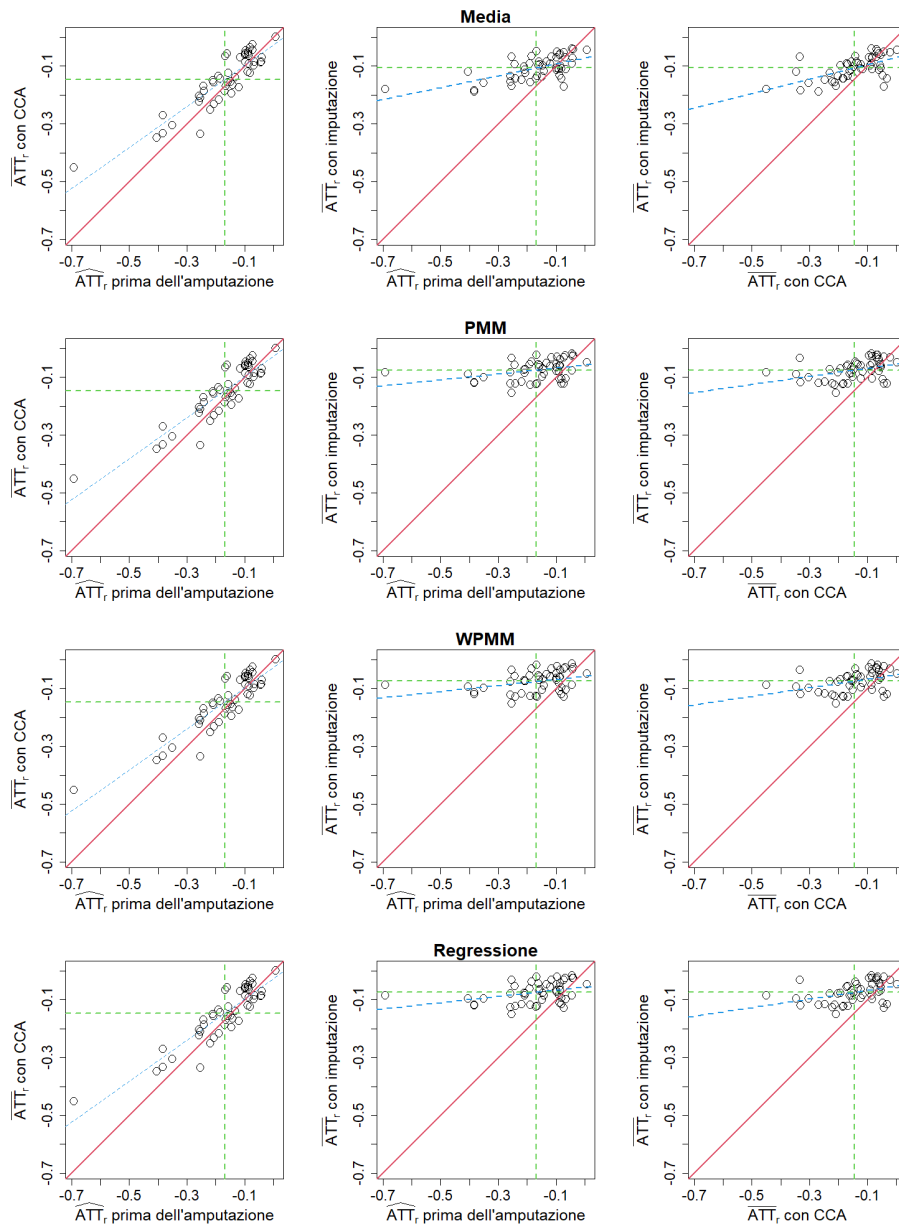


Figura 4.15: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C.

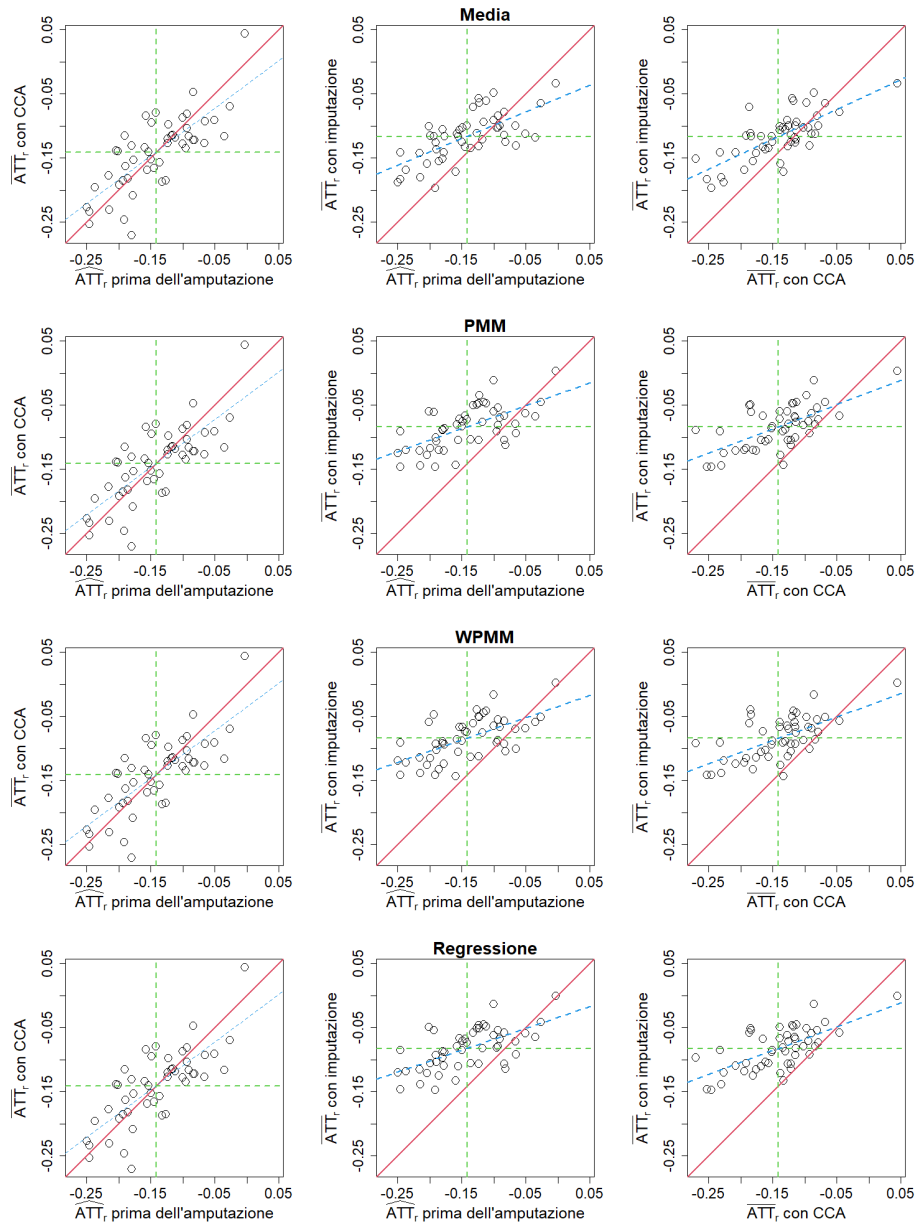


Figura 4.16: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D.

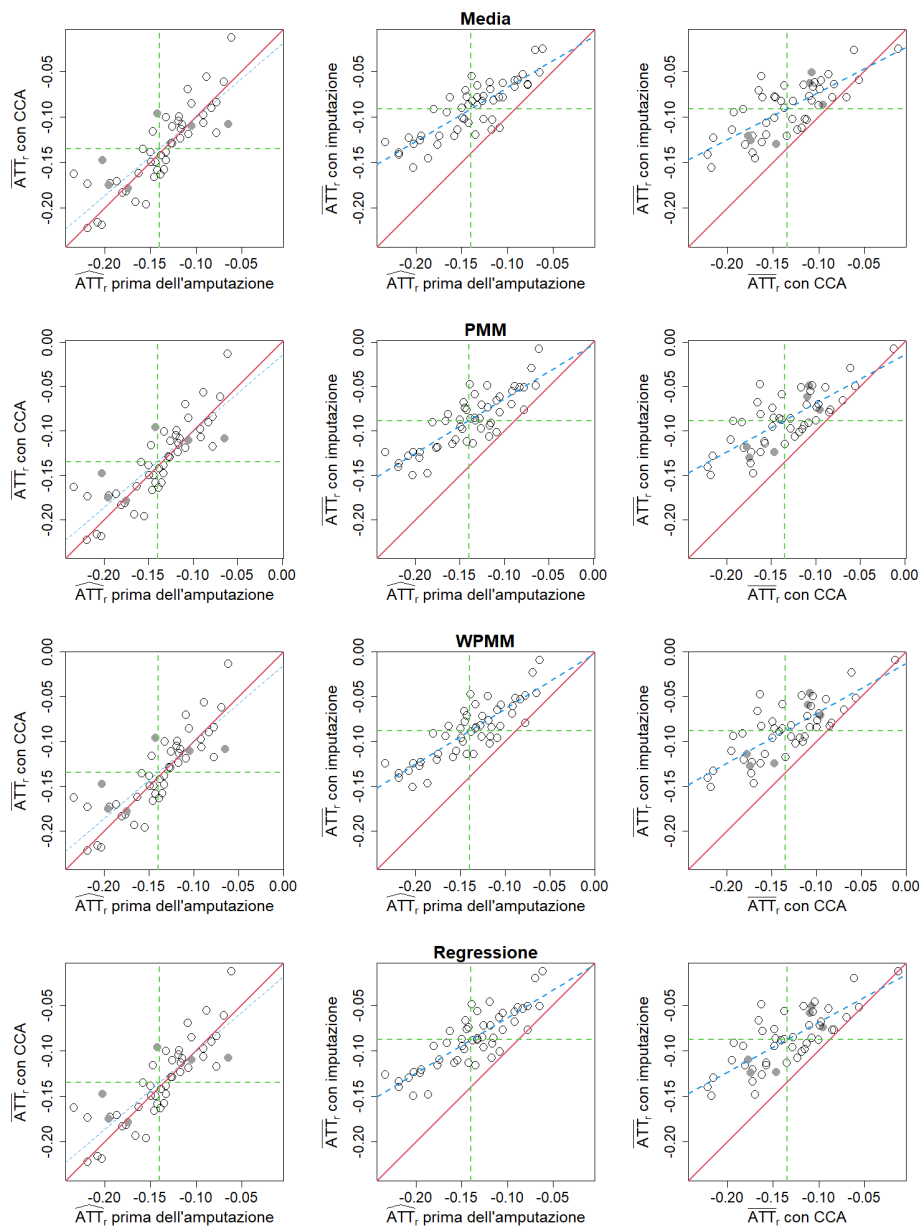


Figura 4.17: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

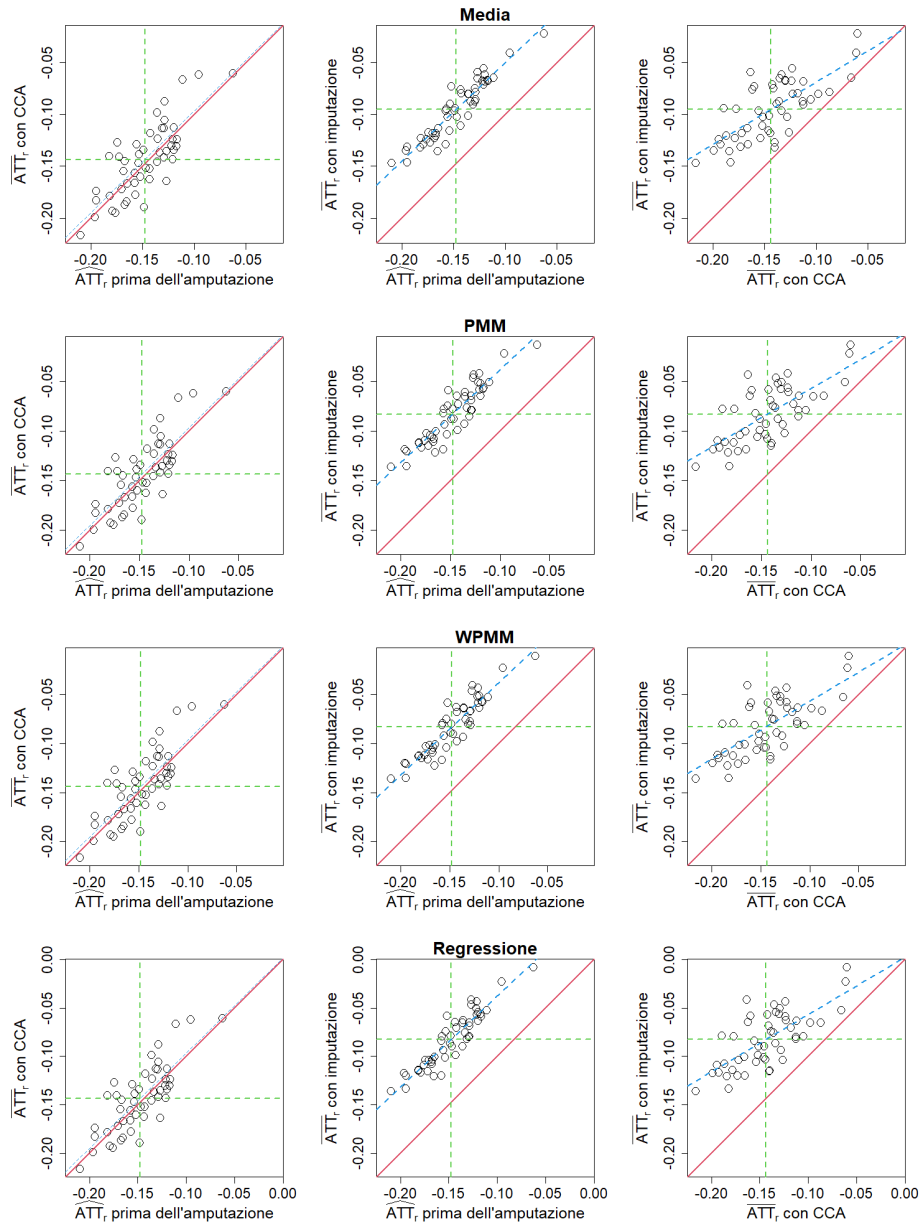


Figura 4.18: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D.

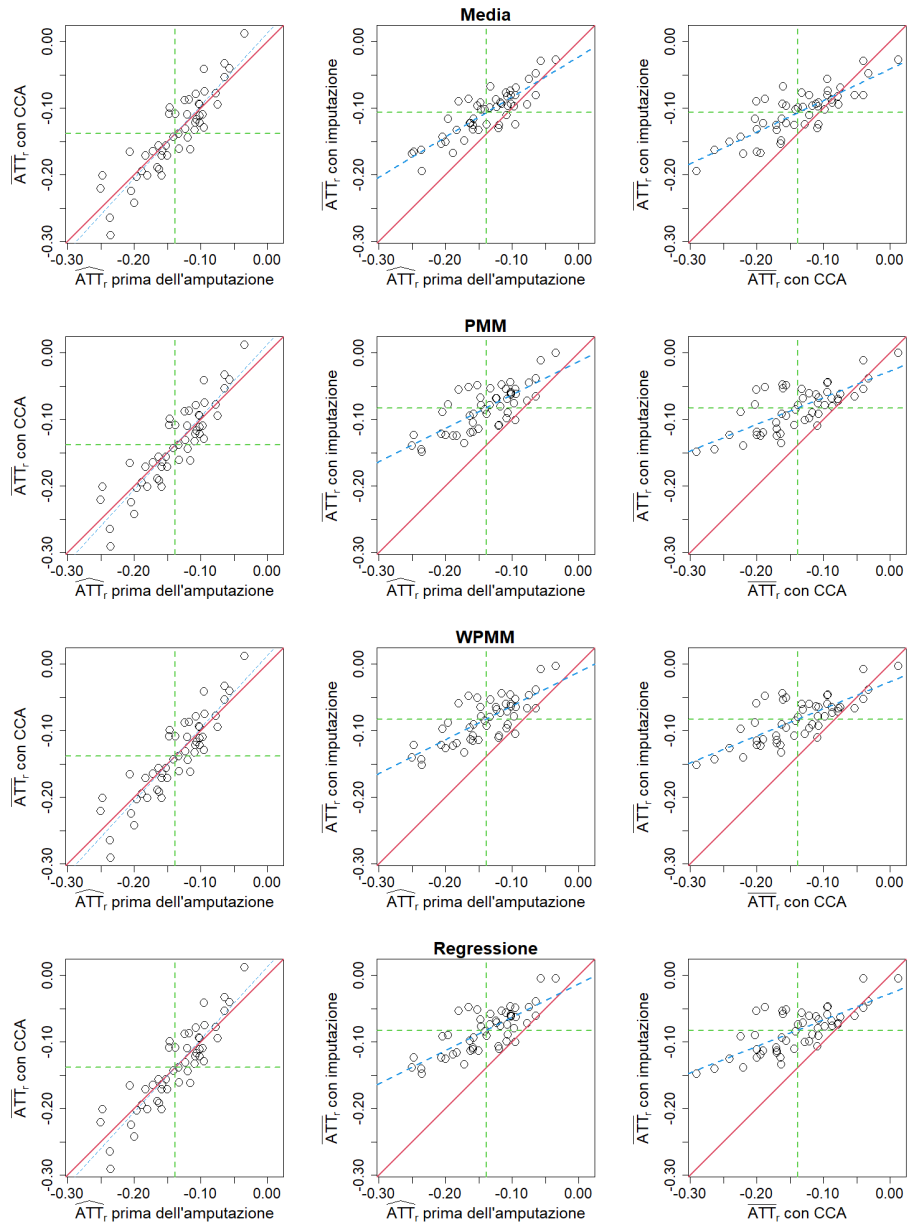


Figura 4.19: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D.

4.5 Due varianti della simulazione

Una volta osservati i risultati ottenuti in questo secondo studio empirico, si è voluto valutare se alcune variazioni nel disegno della simulazione potessero generare differenze nelle stime. Prima si è ripetuta la simulazione diminuendo la numerosità campionaria a 500. Successivamente, è stata fatta una modifica nella procedura di amputazione dei dati. In particolare, si è esclusa la variabile dicotomica X_4 dal modello di generazione dei dati mancanti, cioè si è stabilito che tale variabile non abbia alcun effetto sulla probabilità che i dati delle altre variabili siano mancanti o meno. Il modello di imputazione è rimasto invece invariato e risulta, perciò, scorrettamente specificato. Si riportano nelle Tabelle 4.2 e 4.3 le consuete misure per la valutazione dei risultati. Di seguito si discutono i risultati nei due casi.

4.5.1 Diminuzione della numerosità campionaria

Complessivamente non si notano evidenti differenze nelle stime dell'ATT rispetto alla simulazione precedentemente eseguita. Le differenze di prestazione che si osservano nelle due simulazioni tra i diversi metodi di aggiustamento del confondimento non sembrano seguire una logica, al contrario sembrano casuali. Per quanto riguarda la gestione dei dati mancanti, la *complete case analysis* continua ad avere prestazioni molto superiori a quelle degli altri metodi. Tuttavia si osserva che, a fronte di un leggero aumento della distorsione relativa per le stime ottenute sui dati non amputati e utilizzando la *complete case analysis*, la distorsione per gli altri metodi rimane pressoché invariata. Le differenze tra la *complete case analysis* e gli altri metodi di gestione dei dati mancanti sembrano quindi attenuarsi leggermente. Per quanto riguarda la dispersione, come ci si poteva aspettare si nota un leggero aumento

dell'errore quadratico medio, che risulta più evidente per le stime ottenute con la *complete case analysis*. Sembra quindi che i metodi di imputazione non risentano della diminuzione della numerosità campionaria, al contrario di quanto avviene per la *complete case analysis*.

4.5.2 Variazione del modello di amputazione

Anche in questo caso i risultati non sono particolarmente diversi da quelli ottenuti con la prima simulazione eseguita in questo capitolo. La variazione dei valori dell'errore quadratico medio risulta minima. Nelle distorsioni si notano delle piccole differenze tra le varie combinazioni di metodi, tuttavia tali differenze sembrano essere casuali. Infatti, le conclusioni che si possono trarre sulle prestazioni dei metodi sono le stesse già specificate per la prima simulazione del presente capitolo. Da questi risultati sembra che inserire nel modello di imputazione delle variabili che non sono coinvolte nel meccanismo di generazione dei dati mancanti non porti ad un peggioramento dell'imputazione. Per poter essere certi di questo, però, si dovrebbe fare un ulteriore confronto dei valori imputati e non confrontare unicamente le stime dell'ATT ottenute nelle due diverse simulazioni. Possiamo in ogni caso concludere che le stime dell'ATT non risentono di questa variazione nel meccanismo di amputazione.

Tabella 4.2: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la variante con numerosità pari a 500. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione_{rel}}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione_{rel}}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione_{rel}}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione_{rel}}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione_{rel}}$	\overline{EQM}
Matchung	Dati completi	-0.151	-0.152	0.110	0.006	0.099	-0.146	-0.152	0.090	0.040	0.075	-0.151	-0.152	0.091	0.007	0.065	-0.145	-0.157	0.085	-0.058	0.073
	CCA	-0.151	-0.130	0.125	-0.141	0.141	-0.146	-0.148	0.110	0.014	0.113	-0.151	-0.152	0.114	0.010	0.112	-0.145	-0.142	0.107	-0.201	0.113
	Medta	-0.151	-0.106	0.091	-0.297	0.092	-0.146	-0.110	0.076	-0.247	0.079	-0.151	-0.113	0.076	-0.249	0.078	-0.145	-0.115	0.071	-0.205	0.071
	PMM	-0.151	-0.079	0.104	-0.480	0.093	-0.146	-0.101	0.089	-0.310	0.064	-0.151	-0.090	0.092	-0.400	0.071	-0.145	-0.081	0.087	-0.441	0.075
	WPMM	-0.151	-0.078	0.105	-0.487	0.094	-0.146	-0.102	0.089	-0.302	0.063	-0.151	-0.089	0.092	-0.406	0.072	-0.145	-0.081	0.087	-0.442	0.075
Stratificazione	Dati completi	-0.151	-0.079	0.104	-0.478	0.093	-0.146	-0.102	0.090	-0.302	0.062	-0.151	-0.089	0.092	-0.409	0.071	-0.145	-0.083	0.087	-0.425	0.074
	CCA	-0.151	-0.122	0.089	-0.192	0.088	-0.146	-0.139	0.073	-0.052	0.069	-0.151	-0.138	0.070	-0.084	0.050	-0.145	-0.135	0.066	-0.067	0.049
	Medta	-0.151	-0.117	0.114	-0.224	0.127	-0.146	-0.143	0.099	-0.021	0.101	-0.151	-0.131	0.098	-0.129	0.091	-0.145	-0.137	0.093	-0.053	0.068
	PMM	-0.151	-0.090	0.074	-0.405	0.090	-0.146	-0.105	0.062	-0.280	0.064	-0.151	-0.098	0.059	-0.349	0.067	-0.145	-0.088	0.056	-0.392	0.068
	WPMM	-0.151	-0.073	0.074	-0.518	0.092	-0.146	-0.101	0.064	-0.313	0.059	-0.151	-0.084	0.062	-0.444	0.072	-0.145	-0.085	0.059	-0.413	0.069
Overrate adjustment	Dati completi	-0.151	-0.072	0.075	-0.527	0.093	-0.146	-0.100	0.063	-0.316	0.060	-0.151	-0.084	0.062	-0.442	0.072	-0.145	-0.085	0.059	-0.410	0.069
	CCA	-0.151	-0.150	0.054	-0.008	0.043	-0.146	-0.152	0.050	0.041	0.042	-0.151	-0.157	0.054	0.042	0.044	-0.145	-0.138	0.052	-0.047	0.038
	Medta	-0.151	-0.142	0.077	-0.059	0.069	-0.146	-0.152	0.070	0.036	0.068	-0.151	-0.154	0.078	0.019	0.073	-0.145	-0.135	0.075	-0.069	0.068
	PMM	-0.151	-0.107	0.053	-0.294	0.059	-0.146	-0.113	0.047	-0.225	0.046	-0.151	-0.101	0.050	-0.327	0.060	-0.145	-0.090	0.049	-0.377	0.064
	WPMM	-0.151	-0.097	0.054	-0.355	0.063	-0.146	-0.104	0.049	-0.288	0.051	-0.151	-0.092	0.051	-0.391	0.065	-0.145	-0.079	0.050	-0.456	0.072
IPTW	Dati completi	-0.151	-0.097	0.054	-0.380	0.064	-0.146	-0.103	0.049	-0.293	0.051	-0.151	-0.091	0.052	-0.394	0.065	-0.145	-0.079	0.049	-0.456	0.072
	CCA	-0.151	-0.140	0.080	-0.077	0.087	-0.146	-0.152	0.073	0.038	0.066	-0.151	-0.205	0.073	0.358	0.105	-0.145	-0.140	0.068	-0.038	0.060
	Medta	-0.151	-0.127	0.110	-0.164	0.125	-0.146	-0.151	0.097	0.030	0.105	-0.151	-0.198	0.107	0.313	0.193	-0.145	-0.141	0.096	-0.030	0.098
	PMM	-0.151	-0.104	0.073	-0.310	0.089	-0.146	-0.117	0.062	-0.198	0.065	-0.151	-0.121	0.065	-0.198	0.060	-0.145	-0.104	0.057	-0.283	0.063
	WPMM	-0.151	-0.085	0.088	-0.437	0.090	-0.146	-0.102	0.071	-0.306	0.061	-0.151	-0.086	0.069	-0.429	0.070	-0.145	-0.081	0.061	-0.442	0.073
Regressione	WPMM	-0.151	-0.086	0.088	-0.432	0.089	-0.146	-0.101	0.070	-0.306	0.061	-0.151	-0.086	0.069	-0.429	0.071	-0.145	-0.081	0.061	-0.443	0.073
	Regressione	-0.151	-0.084	0.088	-0.446	0.092	-0.146	-0.101	0.071	-0.309	0.062	-0.151	-0.087	0.070	-0.423	0.070	-0.145	-0.082	0.061	-0.438	0.073

Tabella 4.3: Misure di valutazione della prestazione dei metodi di trattamento del confondimento e di gestione dei dati mancanti per la variante sul modello di amputazione. \overline{ATT}_{veri} : media dei veri valori di ATT rispetto alle 50 repliche. $\overline{ATT}_{stimati}$: media delle stime dell'ATT rispetto alle 50 repliche e alle 10 amputazioni per ognuna di esse. \overline{ES} : media dei relativi errori standard ottenuta nello stesso modo. $\overline{Distorsione}_{el}$: media delle distorsioni relative. \overline{EQM} : media degli errori quadratici medi.

Trattamento del confondimento	Gestione dei dati mancanti	Scenario A					Scenario B					Scenario C					Scenario D				
		\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{el}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{el}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{el}$	\overline{EQM}	\overline{ATT}_{veri}	$\overline{ATT}_{stimati}$	\overline{ES}	$\overline{Distorsione}_{el}$	\overline{EQM}
Matching	Dati completi	-0.151	-0.153	0.080	0.045	0.067	-0.146	-0.147	0.068	0.006	0.054	-0.150	-0.143	0.071	-0.050	0.057	-0.145	-0.145	0.063	0.001	0.055
	CCA	-0.151	-0.153	0.098	0.013	0.097	-0.146	-0.145	0.083	-0.007	0.090	-0.150	-0.159	0.092	0.057	0.087	-0.145	-0.145	0.084	0.003	0.079
	Media	-0.151	-0.121	0.068	-0.202	0.067	-0.146	-0.117	0.056	-0.198	0.062	-0.150	-0.121	0.057	-0.194	0.056	-0.145	-0.123	0.052	-0.147	0.052
	PAM	-0.151	-0.087	0.074	-0.423	0.072	-0.146	-0.097	0.063	-0.336	0.061	-0.150	-0.093	0.065	-0.382	0.063	-0.145	-0.084	0.061	-0.418	0.066
Stratificazione	WPAM	-0.151	-0.085	0.074	-0.437	0.073	-0.146	-0.096	0.063	-0.342	0.062	-0.150	-0.092	0.066	-0.387	0.065	-0.145	-0.084	0.061	-0.420	0.066
	Regressione	-0.151	-0.087	0.073	-0.428	0.072	-0.146	-0.097	0.063	-0.333	0.060	-0.150	-0.093	0.066	-0.384	0.064	-0.145	-0.084	0.061	-0.416	0.065
	Dati completi	-0.151	-0.143	0.060	-0.058	0.052	-0.146	-0.145	0.050	-0.007	0.043	-0.150	-0.139	0.049	-0.073	0.036	-0.145	-0.136	0.046	-0.060	0.041
	CCA	-0.151	-0.140	0.078	-0.074	0.077	-0.146	-0.139	0.066	-0.045	0.066	-0.150	-0.133	0.070	-0.119	0.064	-0.145	-0.133	0.065	-0.080	0.060
Covariate adjustment	Media	-0.151	-0.104	0.050	-0.315	0.063	-0.146	-0.100	0.042	-0.315	0.059	-0.150	-0.101	0.042	-0.331	0.056	-0.145	-0.092	0.040	-0.361	0.059
	PAM	-0.151	-0.080	0.050	-0.473	0.076	-0.146	-0.095	0.043	-0.351	0.059	-0.150	-0.086	0.043	-0.427	0.067	-0.145	-0.087	0.041	-0.388	0.061
	WPAM	-0.151	-0.079	0.050	-0.479	0.076	-0.146	-0.095	0.043	-0.352	0.059	-0.150	-0.086	0.043	-0.428	0.067	-0.145	-0.087	0.041	-0.399	0.061
	Regressione	-0.151	-0.079	0.050	-0.478	0.076	-0.146	-0.095	0.043	-0.351	0.059	-0.150	-0.086	0.043	-0.430	0.067	-0.145	-0.087	0.041	-0.401	0.061
IPTW	Dati completi	-0.151	-0.162	0.039	0.071	0.037	-0.146	-0.153	0.035	0.046	0.031	-0.150	-0.161	0.038	0.067	0.029	-0.145	-0.147	0.037	0.014	0.033
	CCA	-0.151	-0.160	0.054	0.059	0.054	-0.146	-0.148	0.049	0.016	0.045	-0.150	-0.157	0.055	0.043	0.052	-0.145	-0.143	0.052	-0.008	0.049
	Media	-0.151	-0.116	0.037	-0.235	0.046	-0.146	-0.107	0.033	-0.263	0.048	-0.150	-0.106	0.035	-0.294	0.050	-0.145	-0.095	0.034	-0.344	0.056
	PAM	-0.151	-0.105	0.038	-0.306	0.052	-0.146	-0.098	0.034	-0.331	0.054	-0.150	-0.096	0.036	-0.359	0.057	-0.145	-0.082	0.035	-0.431	0.065
IPTW	WPAM	-0.151	-0.105	0.038	-0.308	0.052	-0.146	-0.097	0.034	-0.332	0.054	-0.150	-0.097	0.036	-0.359	0.057	-0.145	-0.082	0.035	-0.432	0.065
	Regressione	-0.151	-0.105	0.038	-0.308	0.053	-0.146	-0.097	0.034	-0.334	0.054	-0.150	-0.097	0.037	-0.359	0.057	-0.145	-0.082	0.035	-0.432	0.065
	Dati completi	-0.151	-0.149	0.060	-0.016	0.062	-0.146	-0.156	0.051	0.066	0.051	-0.150	-0.200	0.059	0.327	0.083	-0.145	-0.143	0.053	-0.013	0.050
	CCA	-0.151	-0.148	0.079	-0.025	0.084	-0.146	-0.150	0.068	0.027	0.072	-0.150	-0.196	0.080	0.300	0.136	-0.145	-0.142	0.071	-0.019	0.071
IPTW	Media	-0.151	-0.113	0.056	-0.255	0.064	-0.146	-0.116	0.044	-0.203	0.053	-0.150	-0.120	0.048	-0.206	0.052	-0.145	-0.111	0.040	-0.232	0.050
	PAM	-0.151	-0.093	0.065	-0.388	0.069	-0.146	-0.098	0.049	-0.332	0.059	-0.150	-0.092	0.050	-0.391	0.063	-0.145	-0.083	0.043	-0.425	0.065
	WPAM	-0.151	-0.092	0.065	-0.391	0.070	-0.146	-0.097	0.050	-0.335	0.059	-0.150	-0.092	0.050	-0.391	0.064	-0.145	-0.084	0.044	-0.422	0.065
	Regressione	-0.151	-0.092	0.065	-0.392	0.069	-0.146	-0.097	0.050	-0.333	0.059	-0.150	-0.092	0.050	-0.392	0.063	-0.145	-0.083	0.043	-0.429	0.065

Conclusioni

Questa tesi è nata con lo scopo di confrontare le prestazioni dei quattro principali metodi per trattare il confondimento basati sul *propensity score* in combinazione con diverse tecniche di imputazione dei dati mancanti. Per fare ciò si sono considerate tutte le combinazioni tra i metodi di *propensity score matching*, stratificazione basata sul *propensity score*, *propensity score covariate adjustment* e *inverse probability of treatment weighting* con le tecniche per la gestione dei dati mancanti basate sulla *complete case analysis*, sull'imputazione singola della media e sull'imputazione multipla in tre diverse variati e si sono delineati degli studi empirici basati sulla simulazione in cui tali combinazioni di metodi sono state confrontate sulla base dell'accuratezza e della precisione delle stime dell'effetto medio del trattamento sui trattati. Nel corso dello studio si è constatato che ci sono molti aspetti che possono influenzare le prestazioni dei metodi di utilizzo del *propensity score*, e non è quindi possibile eleggere come vincitore un unico metodo tra di essi. Tuttavia possiamo concludere che alcuni metodi hanno una tendenza ad avere risultati peggiori nella maggior parte dei casi ed altri, al contrario, sembrano essere più efficaci nel ridurre la distorsione causata dal confondimento. In particolare, la stratificazione basata sul *propensity score* risulta il metodo che produce stime più distorte nella maggior parte dei casi. Al contrario il metodo di *propensity score covariate adjustment* è nella maggior parte dei casi il metodo

con prestazioni migliori per quanto riguarda la distorsione e la dispersione delle stime. Dai risultati ottenuti si può concludere, inoltre, che l'effetto della presenza di dati mancanti è fonte di maggiore distorsione nelle stime quando le variabili parzialmente osservate sono confondenti rispetto a quando sono predittori unicamente del trattamento o della risposta. Un'altra importante considerazione è che il metodo di imputazione dei dati mancanti utilizzato e la tipologia delle variabili incomplete non hanno un evidente effetto sulle stime.

Nonostante la letteratura scientifica comprenda altri studi atti a valutare le prestazioni di diversi metodi di applicazione del *propensity score* in presenza di dati mancanti, il presente lavoro è il primo a confrontare tutte le possibili combinazioni tra i metodi e a indagare come variano le prestazioni in relazione ad altre caratteristiche dei dati. Inoltre, le conclusioni evidenziate in merito all'effetto che il ruolo delle variabili parzialmente osservate e il metodo di imputazione dei dati hanno sulle stime dell'effetto del trattamento risultano innovative e non precedentemente menzionate dalla letteratura.

Lo studio svolto per questa tesi può essere ampliato in studi successivi per valutare se ci siano altri aspetti che possono influenzare le stime dell'effetto del trattamento. Alcuni suggerimenti per approfondire lo studio sono:

- utilizzare altre misure per l'effetto del trattamento, per esempio il rapporto delle quote;
- considerare altri meccanismi di generazione dei dati mancanti;
- confrontare i risultati al variare della proporzione di dati mancanti;
- confrontare i due diversi approcci *between* e *across* per l'analisi dei dati tramite *propensity score* quando si utilizza l'imputazione multipla;

- considerare modelli per la risposta più complessi.

Inoltre, per rendere più solide le conclusioni che si sono potute trarre da questo studio, il passo successivo sarebbe quello di associare al confronto di precisione e accuratezza delle stime una valutazione basata sull'inferenza, conducendo test statistici o producendo intervalli di confidenza. Infine, un altro tipo di indagine può essere fatta confrontando il bilanciamento delle covariate dopo aver applicato i metodi di aggiustamento del confondimento.

Appendice A

Grafici di dispersione

Si riportano di seguito i grafici di dispersione prodotti per la valutazione dei risultati ottenuti nello studio empirico descritto nel Capitolo 3 che non sono stati inseriti nel capitolo stesso. Si ricorda che, per ogni simulazione, scenario e combinazione di metodi, si sono prodotti tre grafici di dispersione che confrontano:

- le stime degli ATT ottenute dai dataset completi prima di eseguire l'amputazione con le medie delle stime degli ATT ottenute con la *complete case analysis*;
- le stime degli ATT ottenute dai dataset completi prima di eseguire l'amputazione con le medie delle stime rispetto alle 10 ripetizioni della procedura di amputazione;
- le medie delle stime degli ATT ottenute con la *complete case analysis* con le medie delle stime rispetto alle 10 ripetizioni della procedura di amputazione.

A.1 Simulazione 1 Scenario B

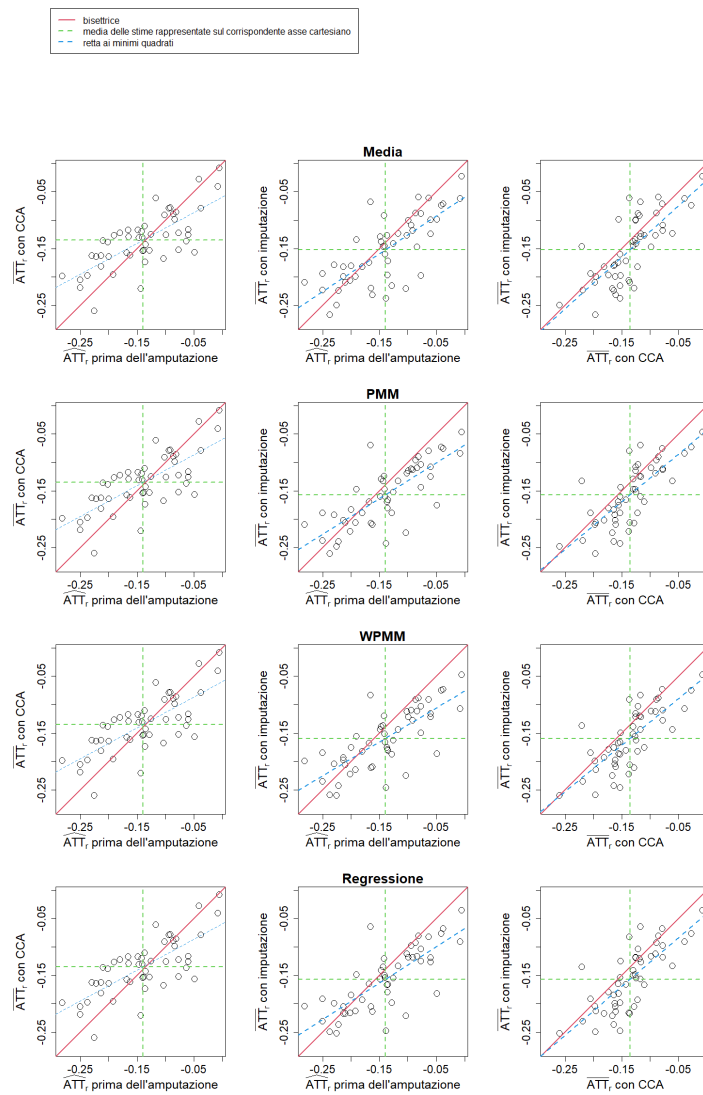


Figura A.1: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 1.

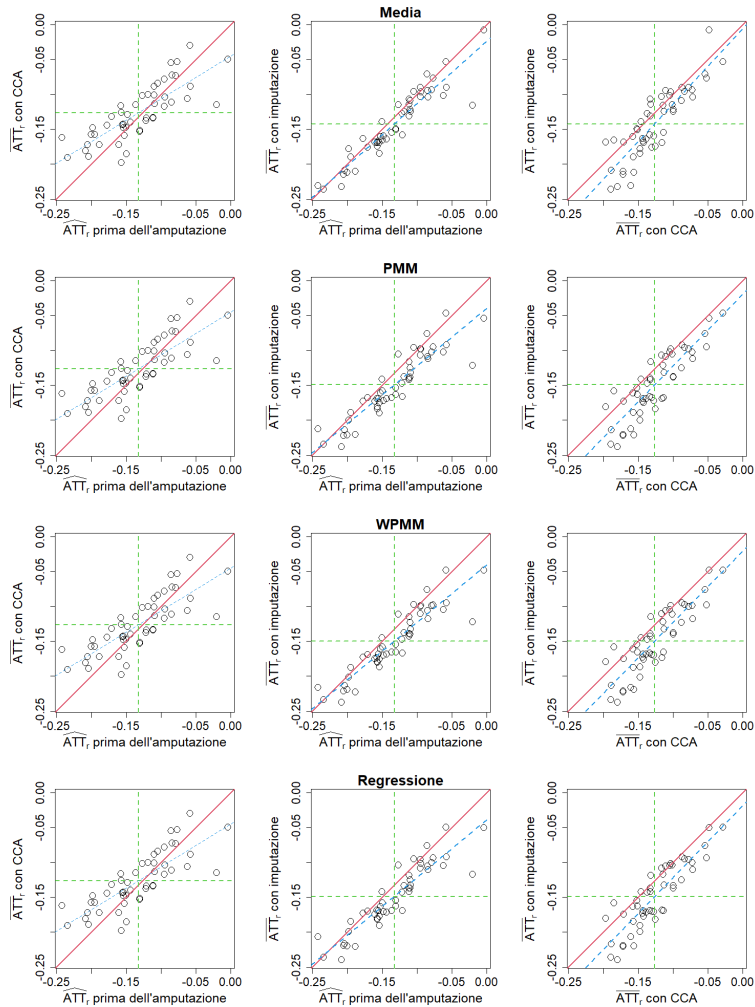


Figura A.2: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 1. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato..

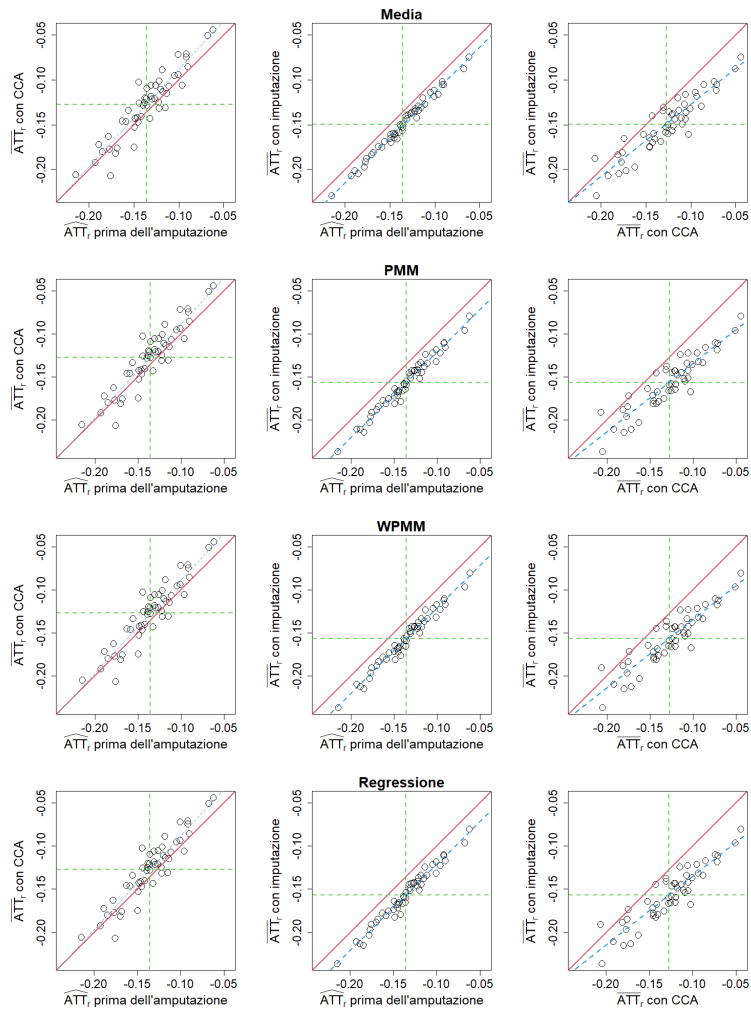


Figura A.3: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 1.

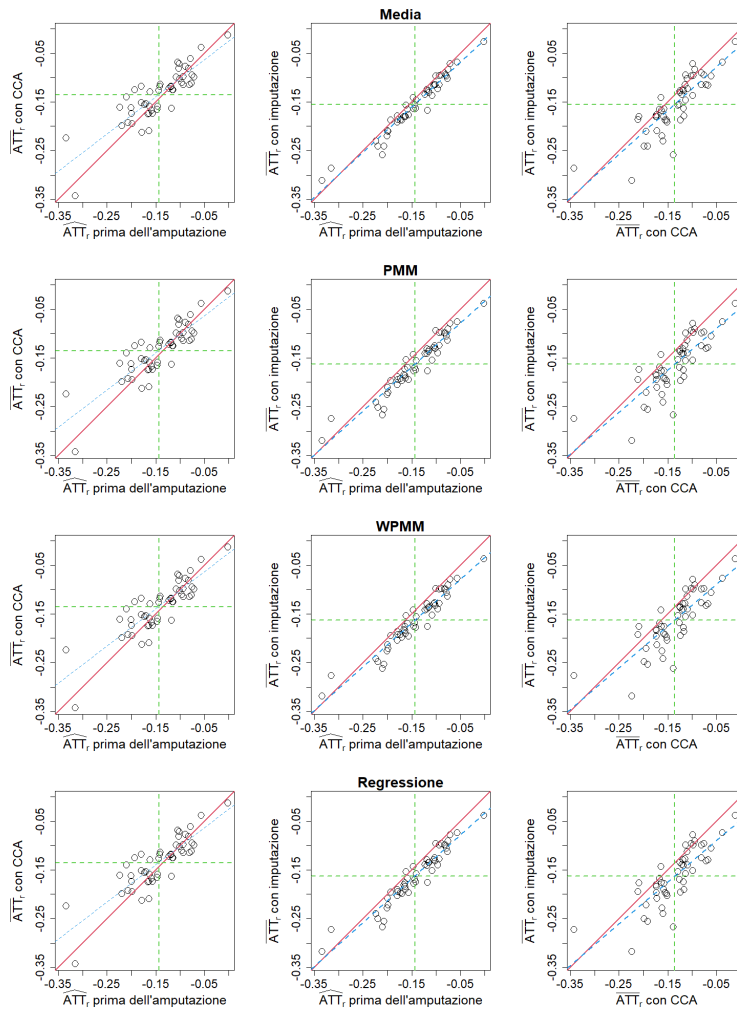


Figura A.4: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 1.

A.2 Simulazione 1 Scenario C

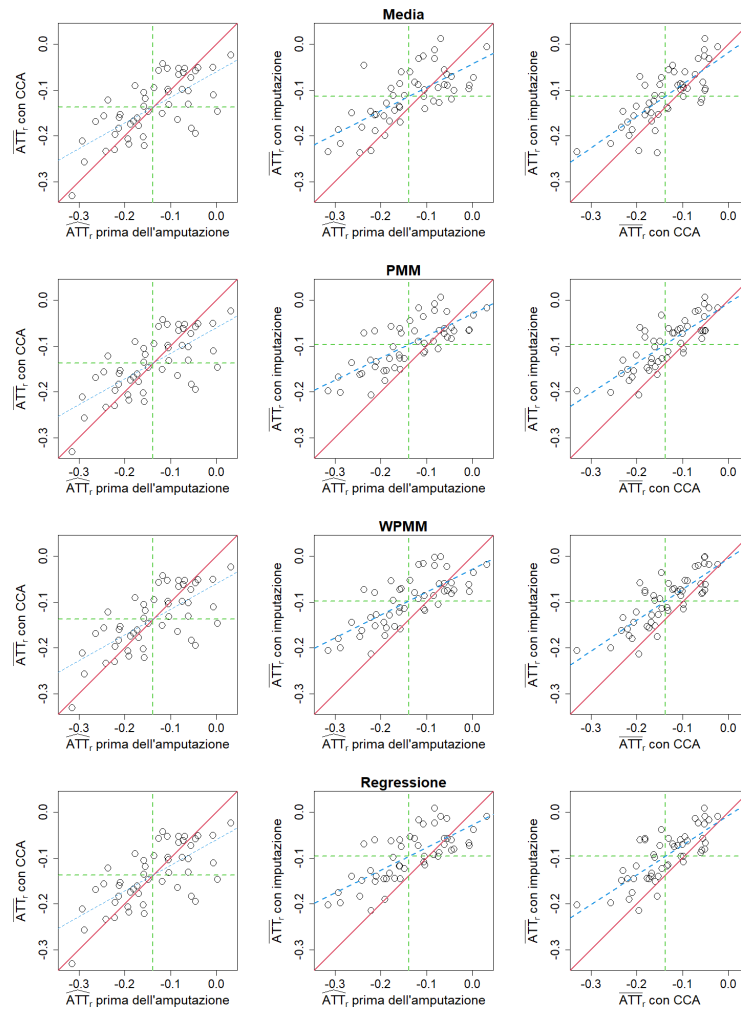


Figura A.5: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 1.

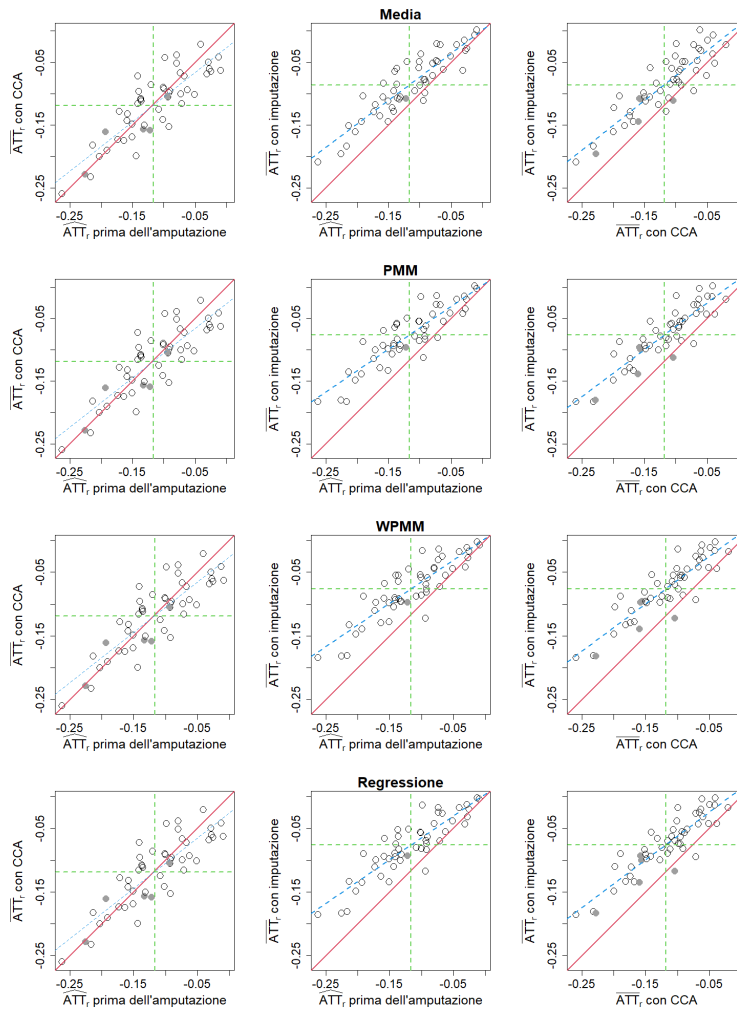


Figura A.6: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 1. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

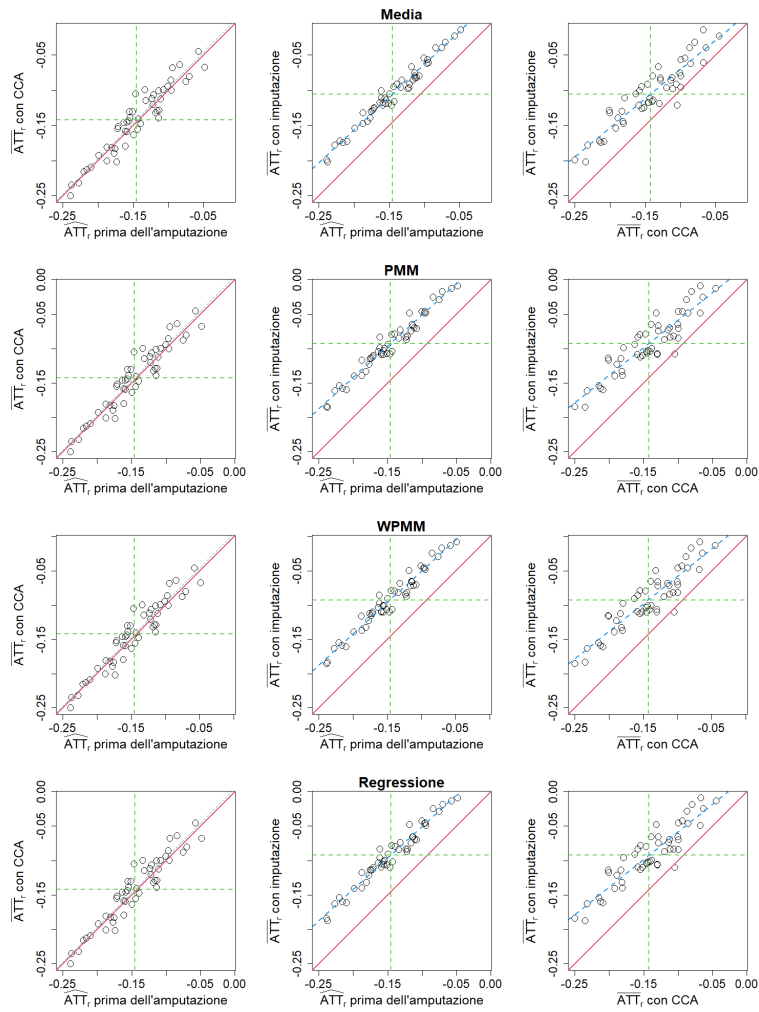


Figura A.7: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 1.

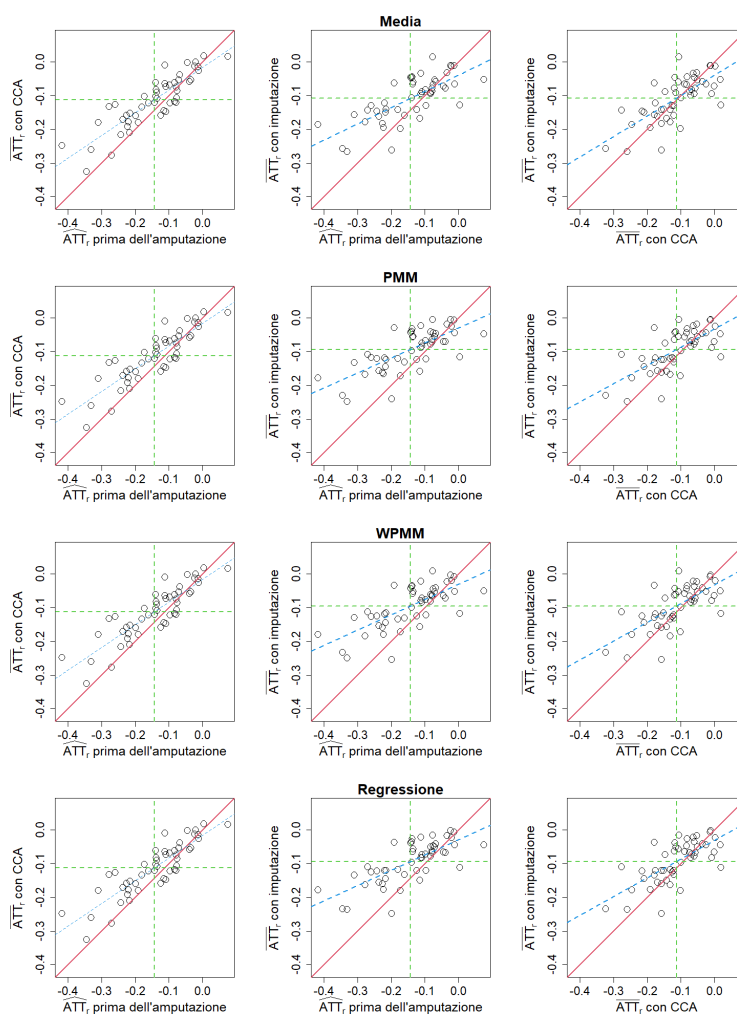


Figura A.8: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 1.

A.3 Simulazione 1 Scenario D

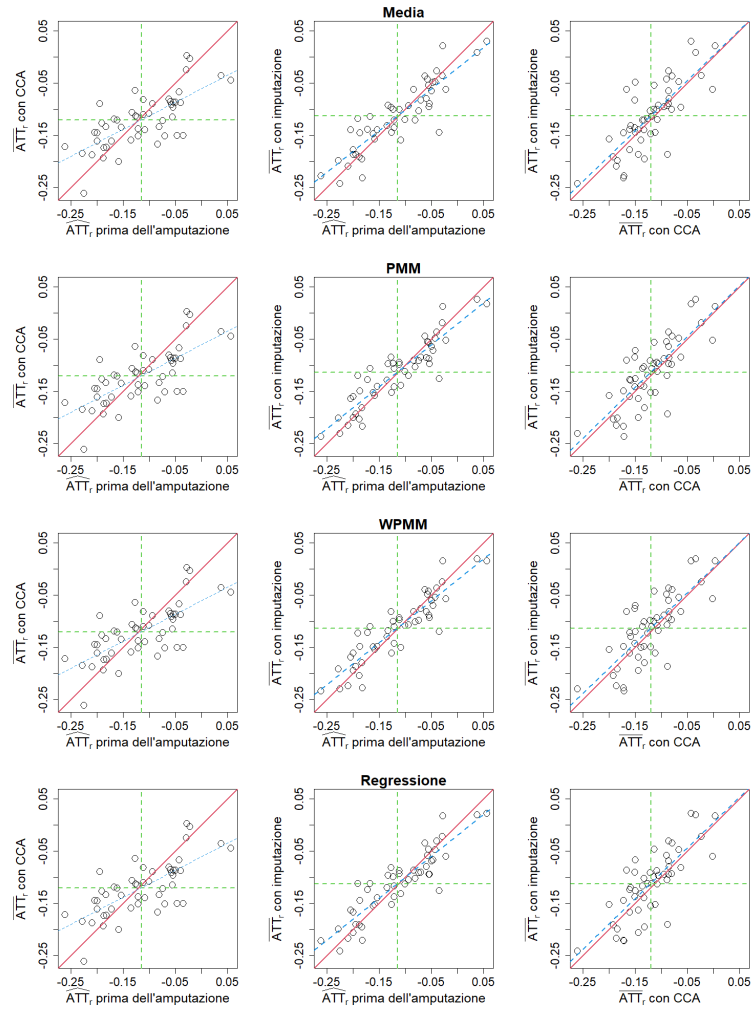


Figura A.9: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 1.

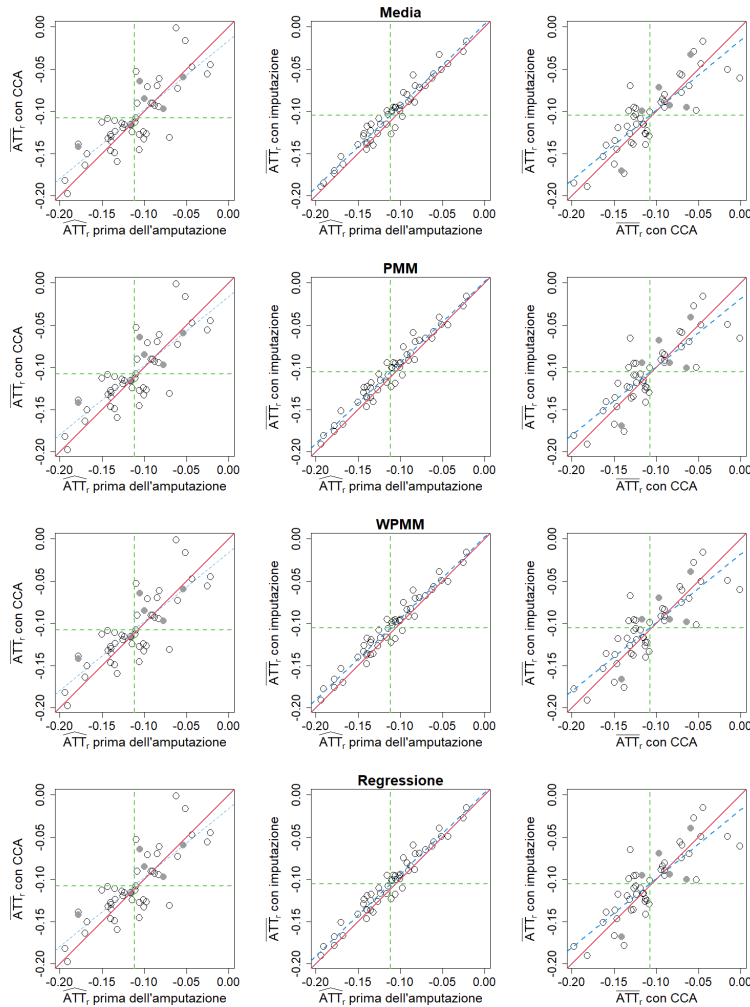


Figura A.10: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 1. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

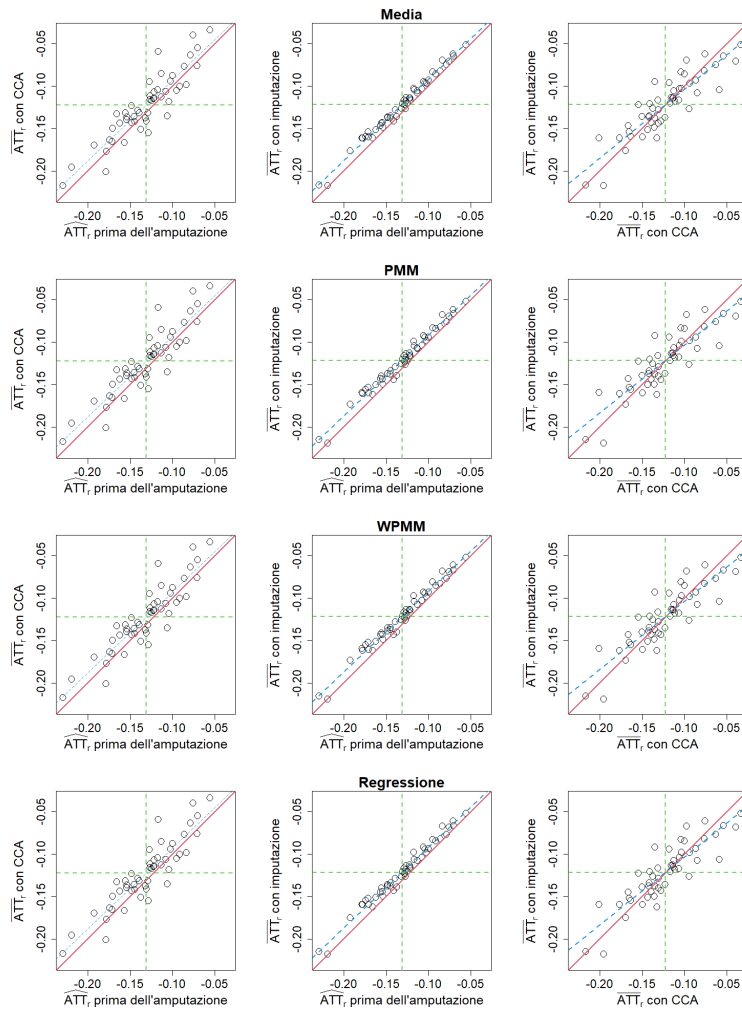


Figura A.11: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 1.

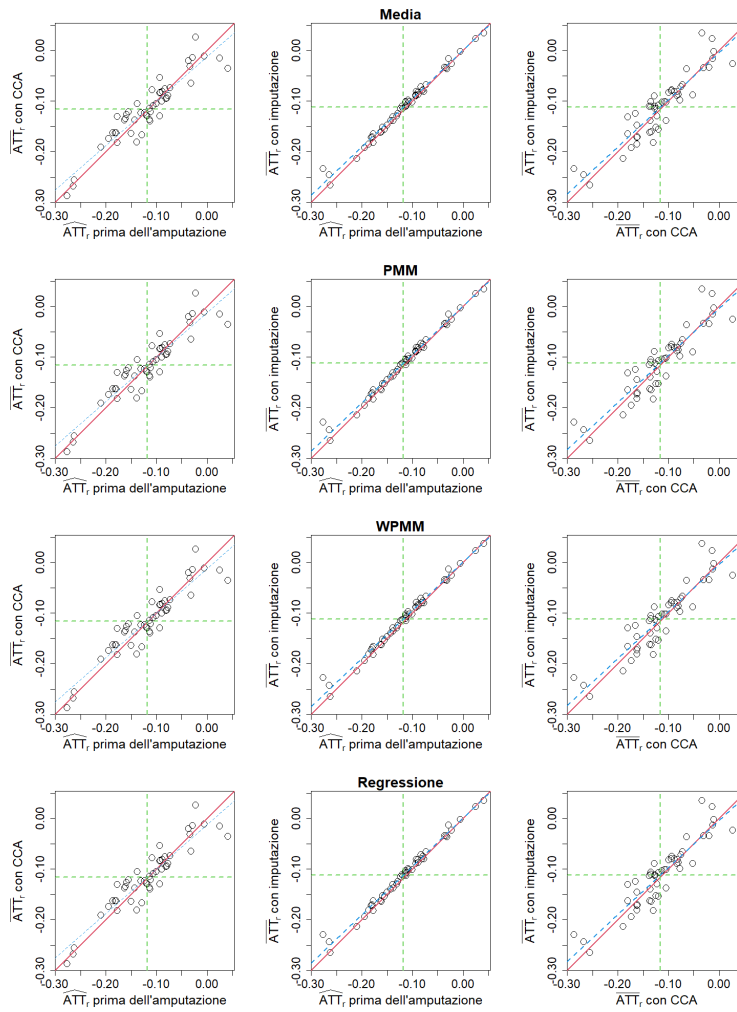


Figura A.12: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 1.

A.4 Simulazione 2 Scenario A

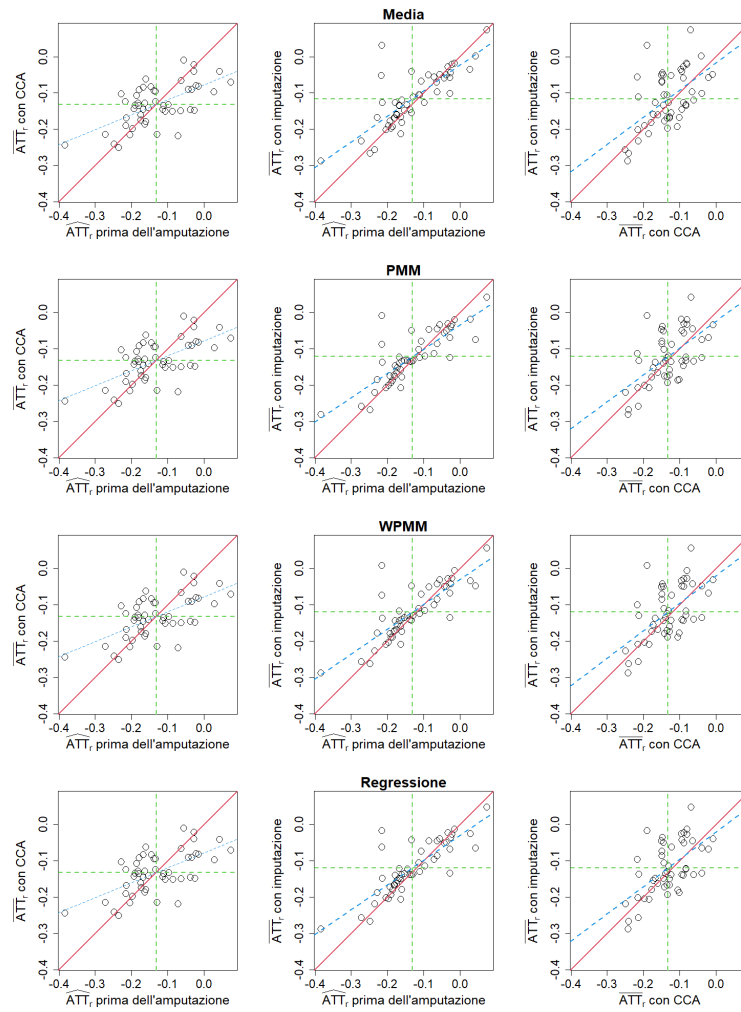


Figura A.13: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 2.

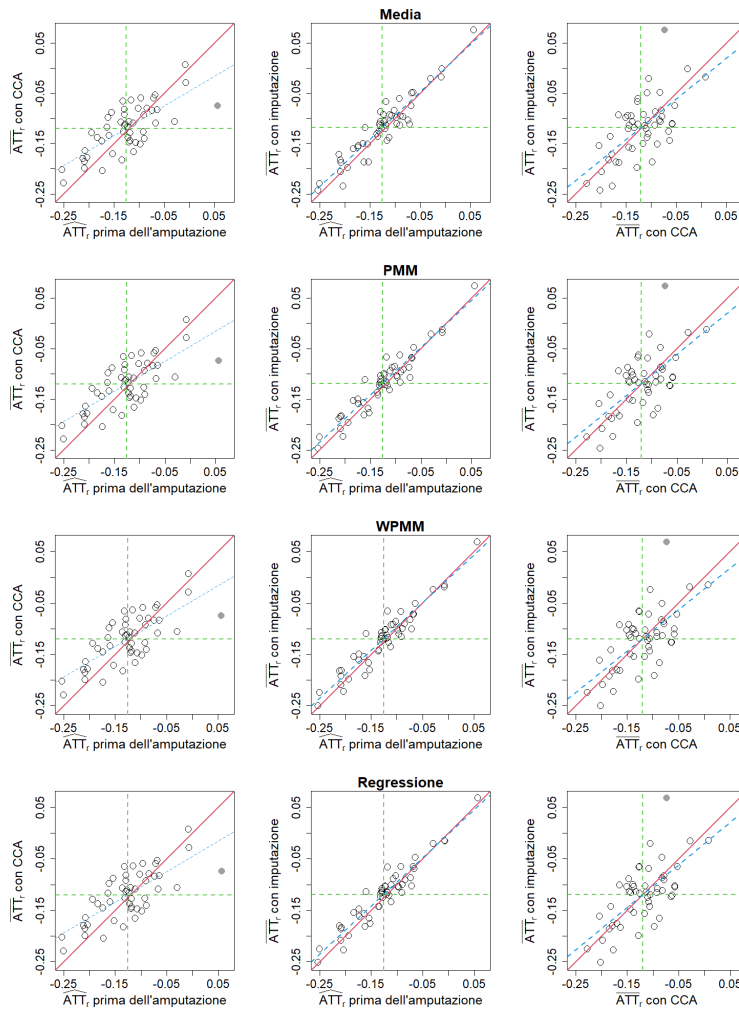


Figura A.14: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 2. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

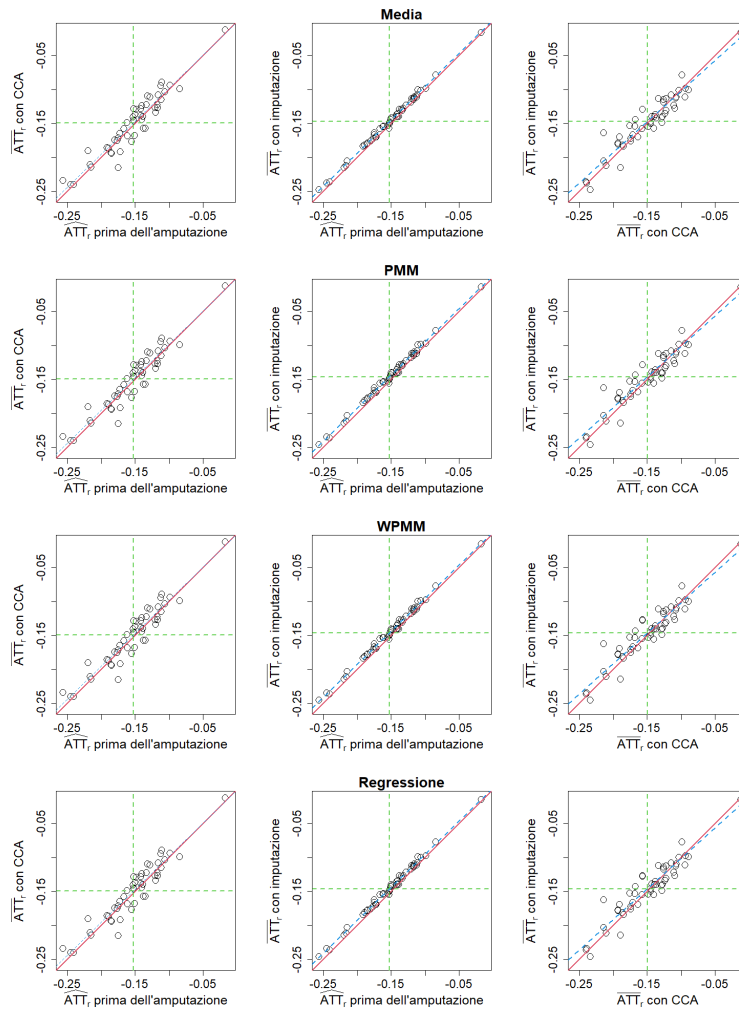


Figura A.15: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 2.

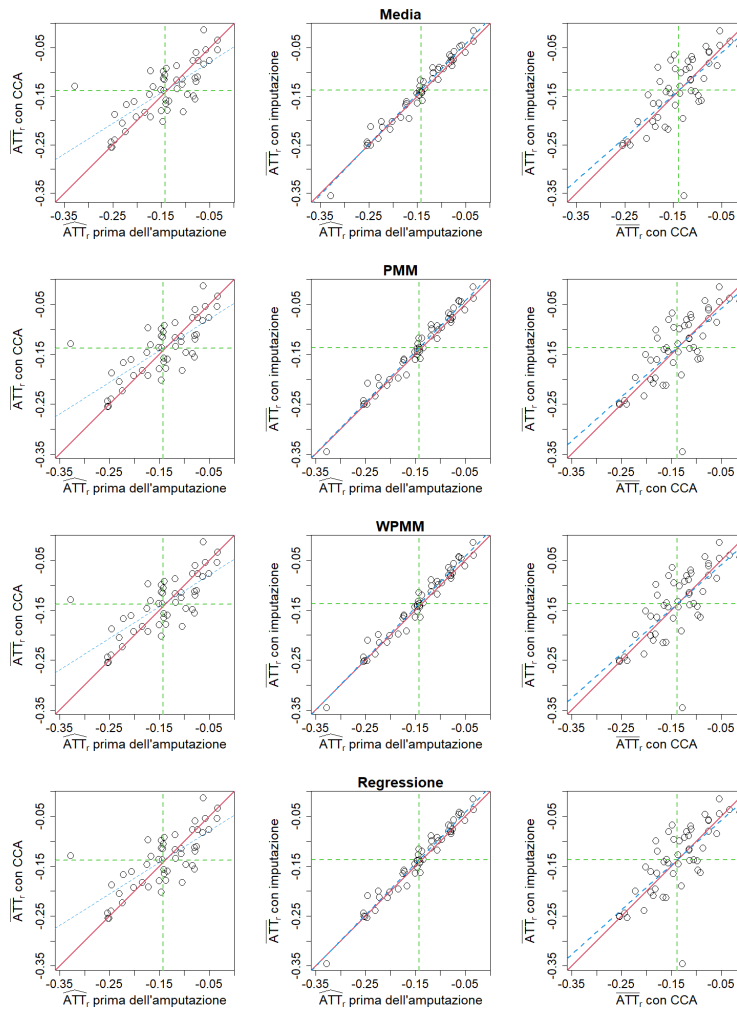


Figura A.16: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 2.

A.5 Simulazione 2 Scenario B

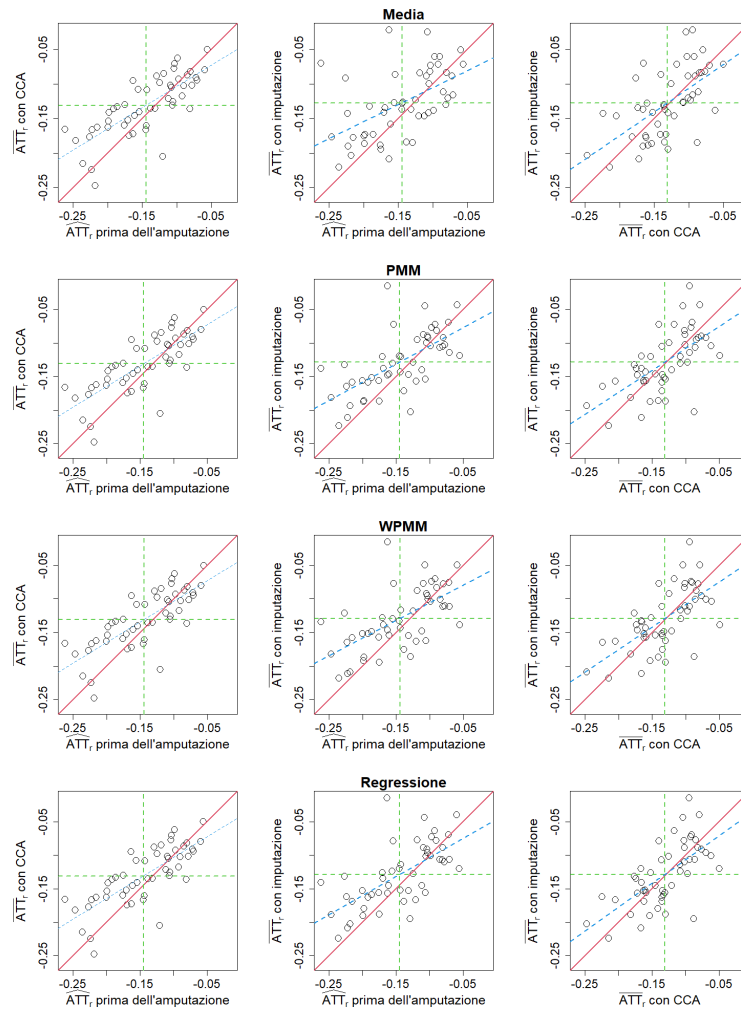


Figura A.17: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 2.

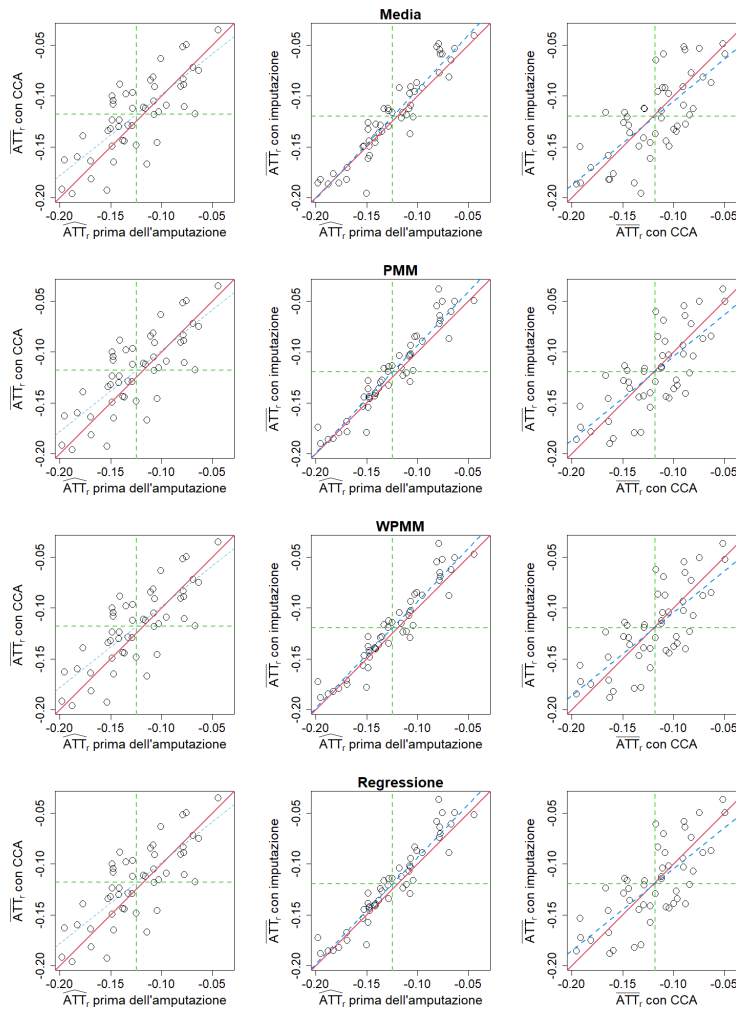


Figura A.18: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 2. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

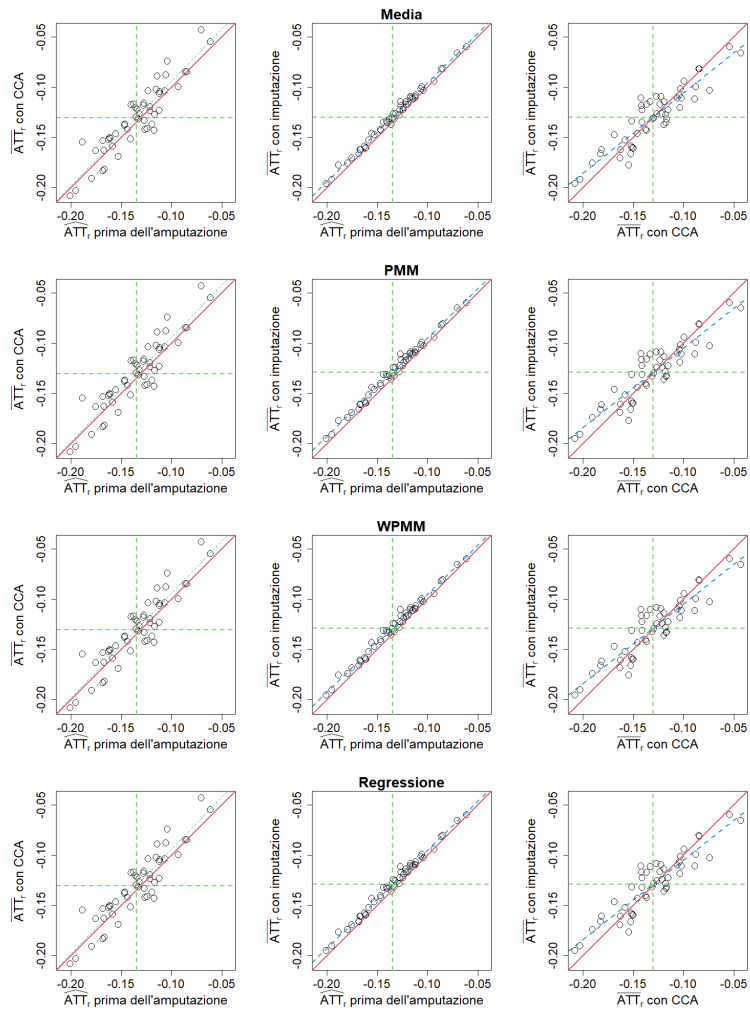


Figura A.19: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 2.

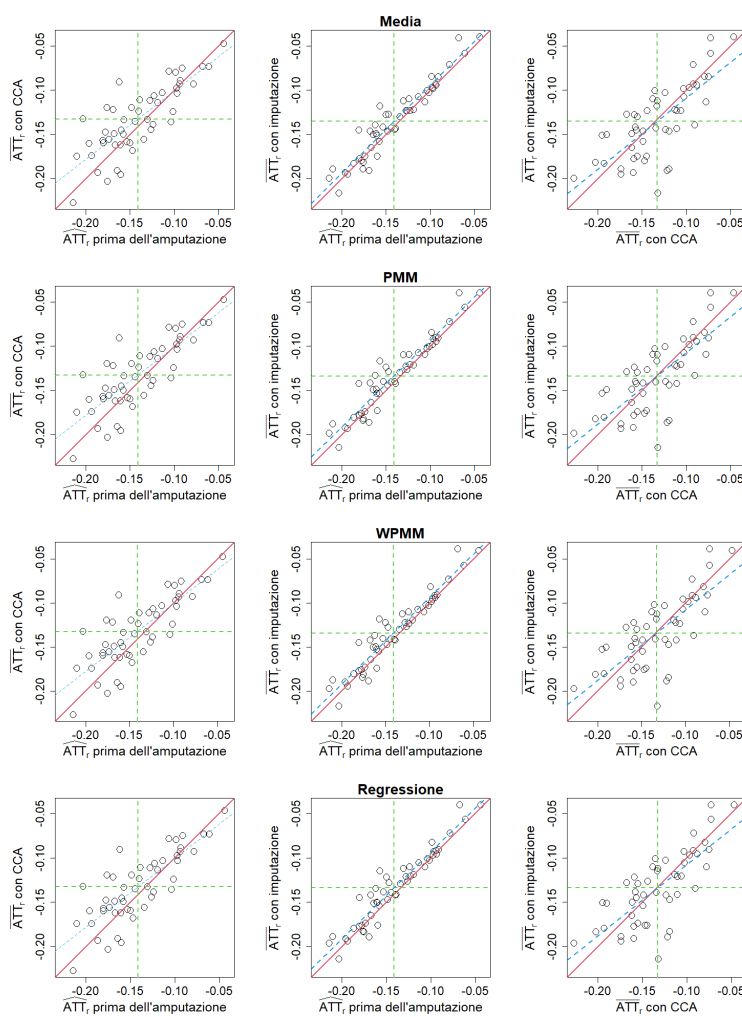


Figura A.20: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 2.

A.6 Simulazione 2 Scenario C

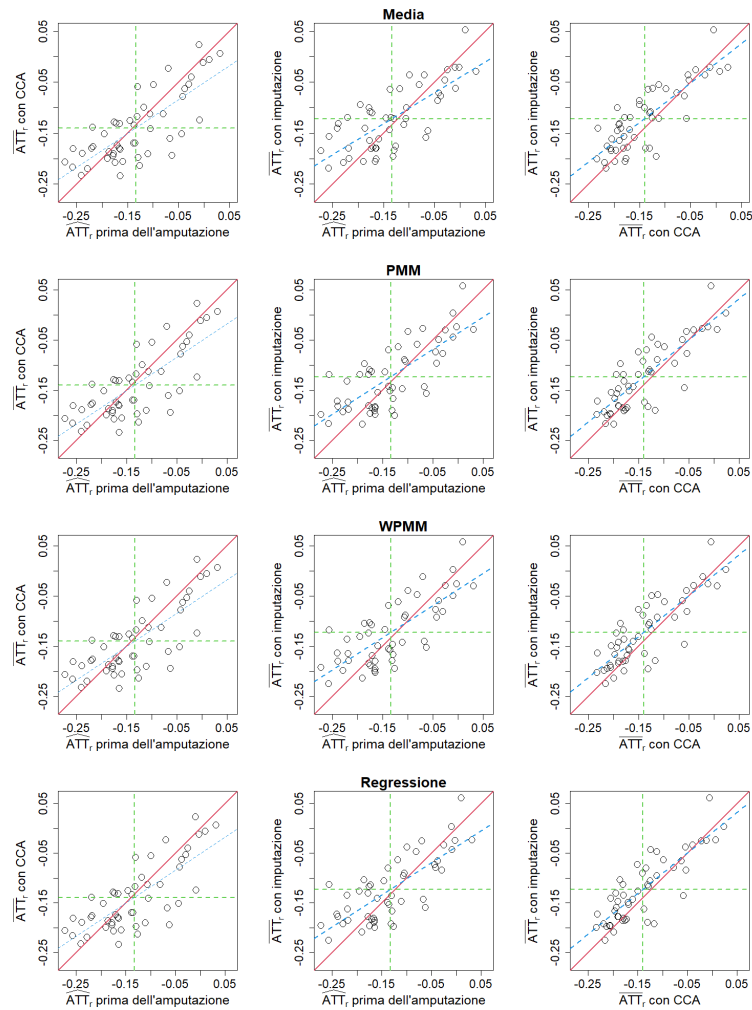


Figura A.21: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 2.

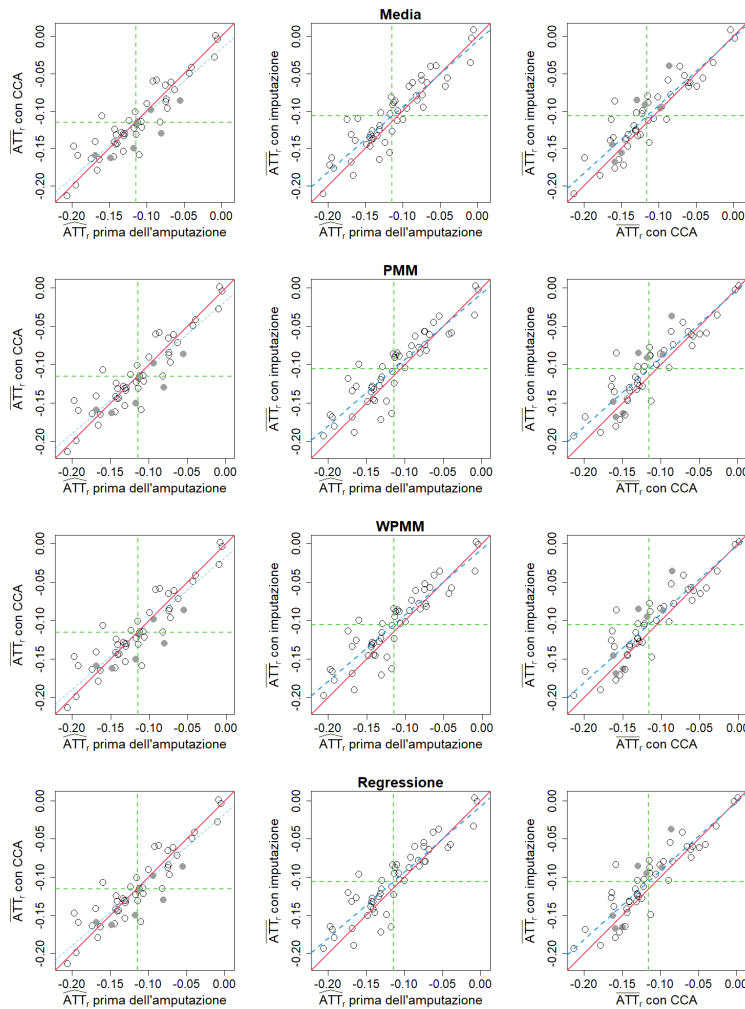


Figura A.22: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 2. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

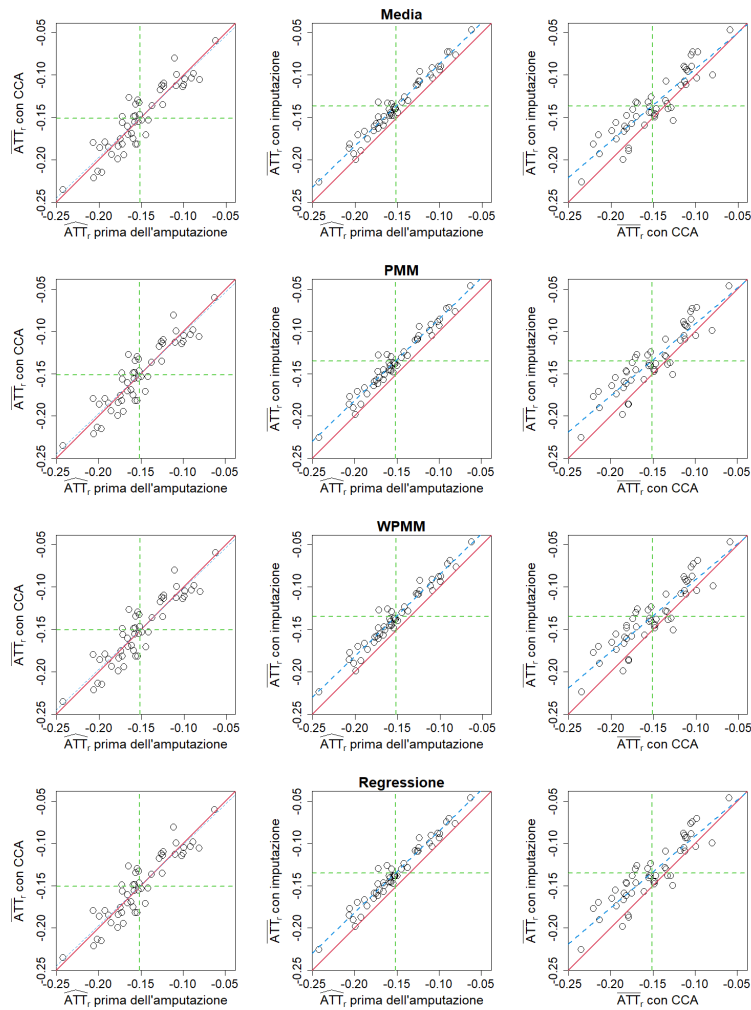


Figura A.23: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 2.

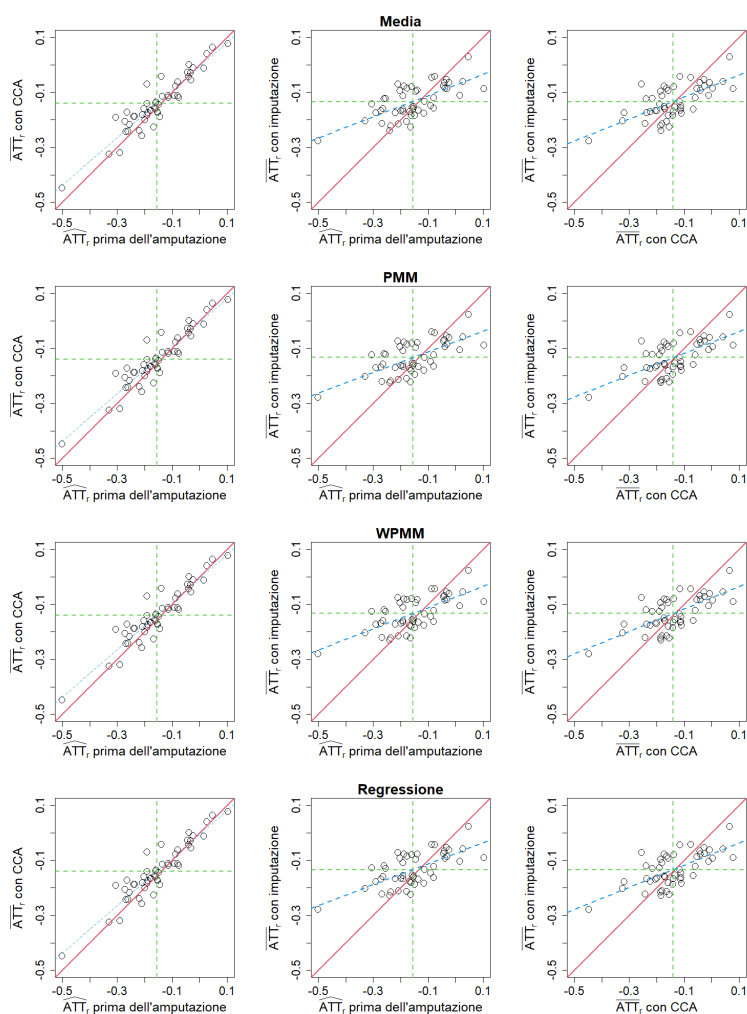


Figura A.24: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 2.

A.7 Simulazione 2 Scenario D

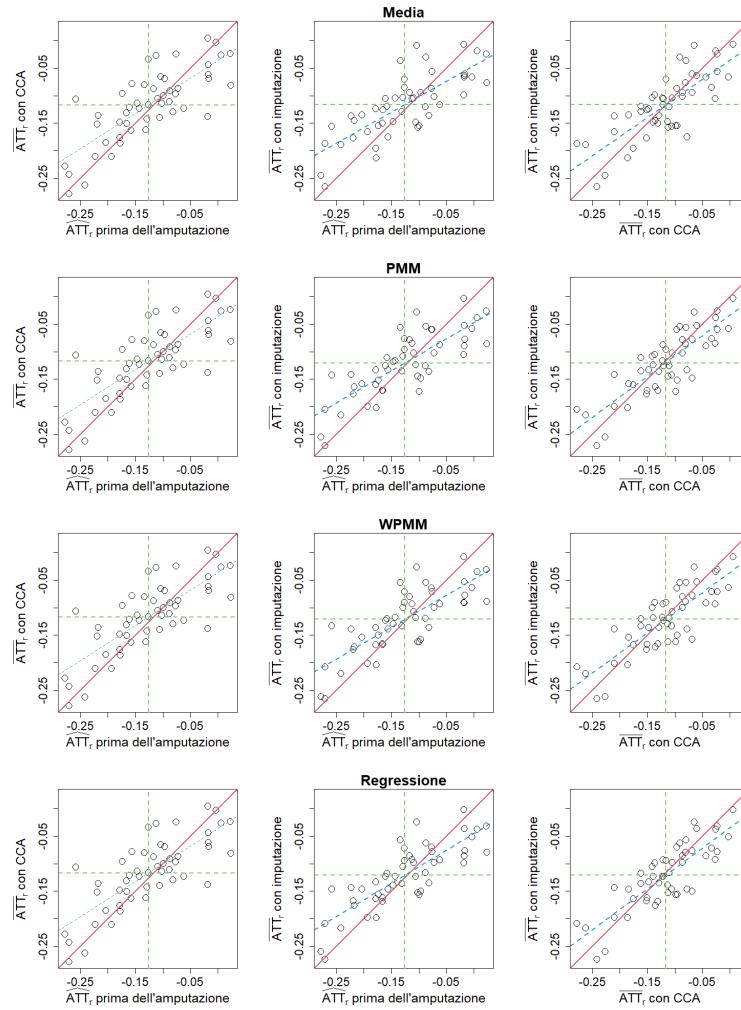


Figura A.25: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 2.

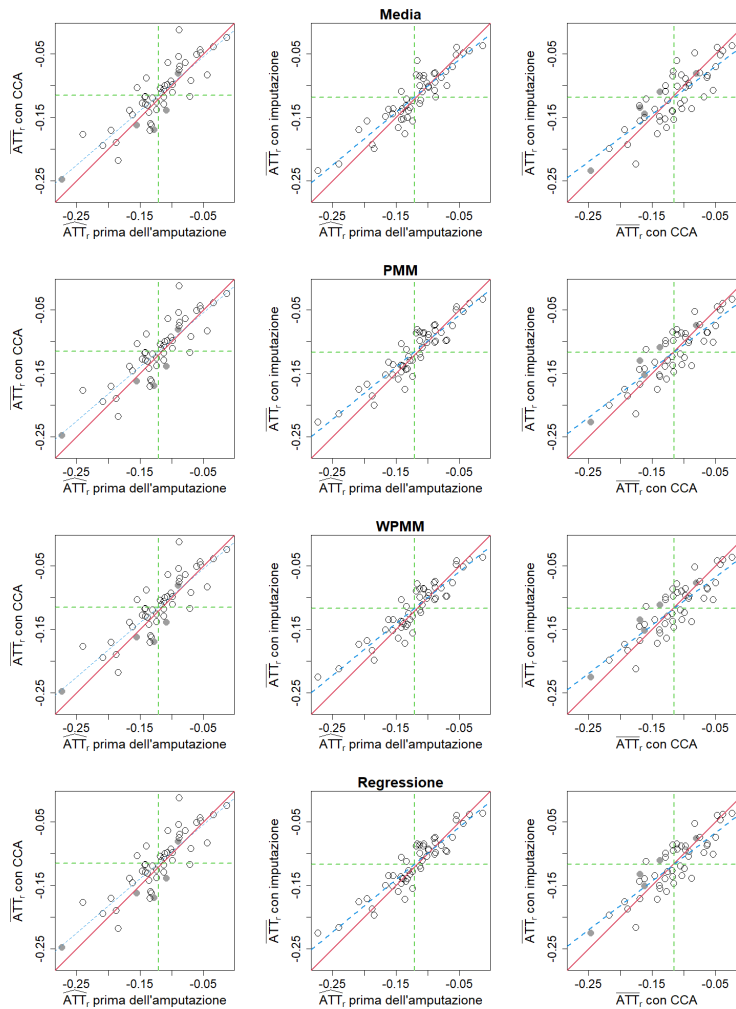


Figura A.26: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 2. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

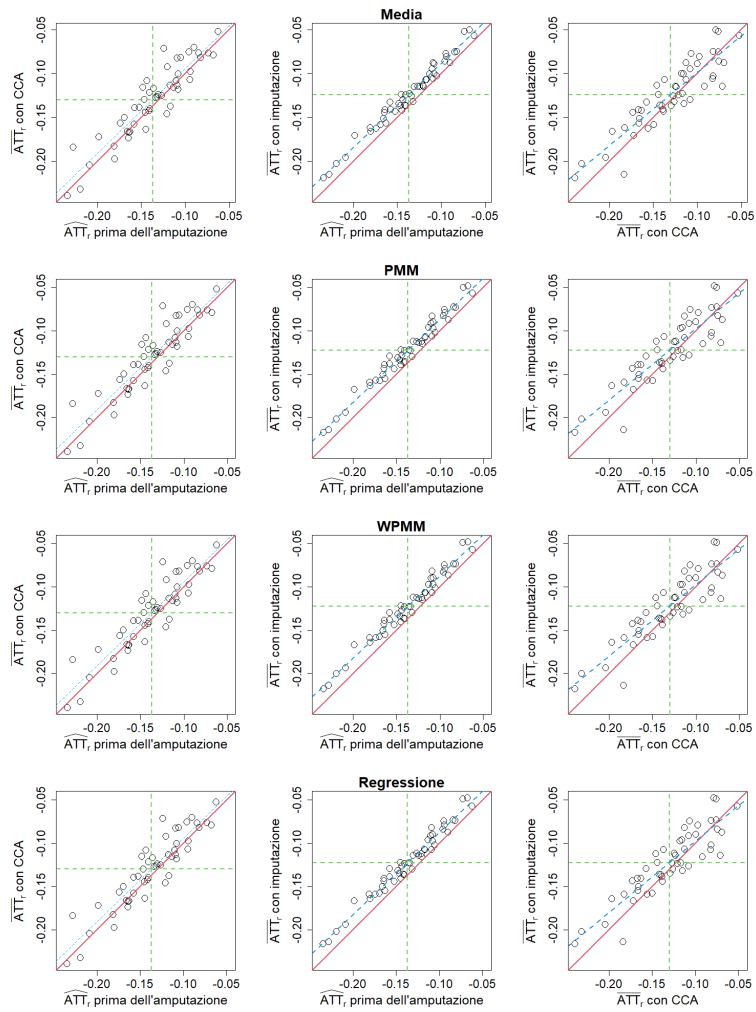


Figura A.27: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 2.

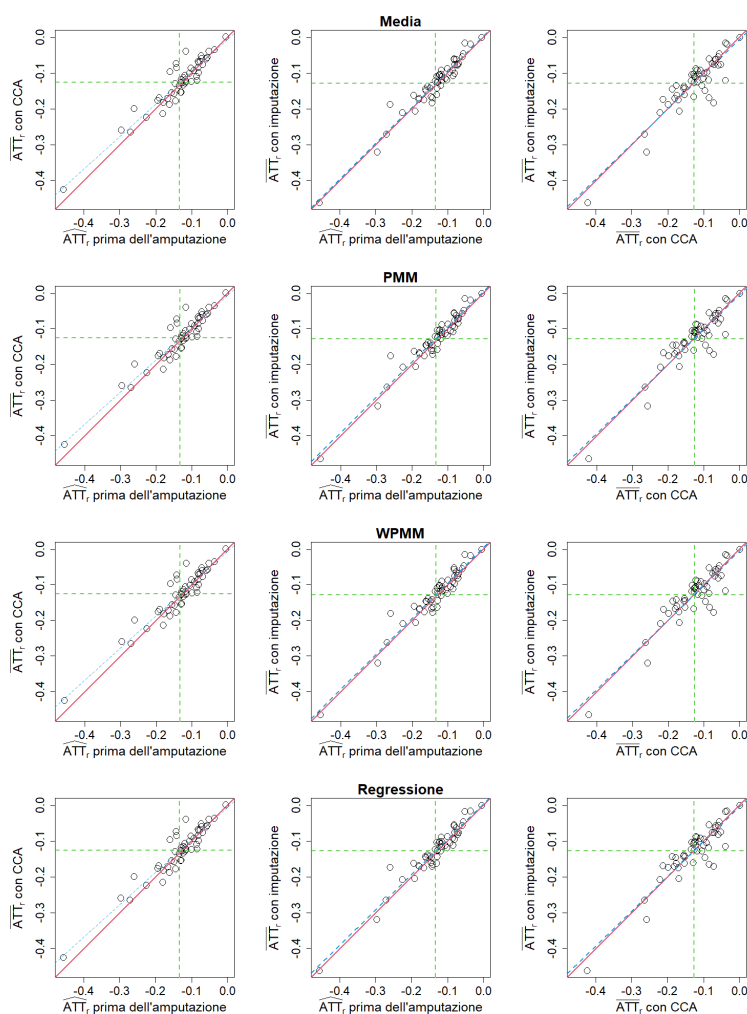


Figura A.28: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 2.

A.8 Simulazione 3 Scenario A

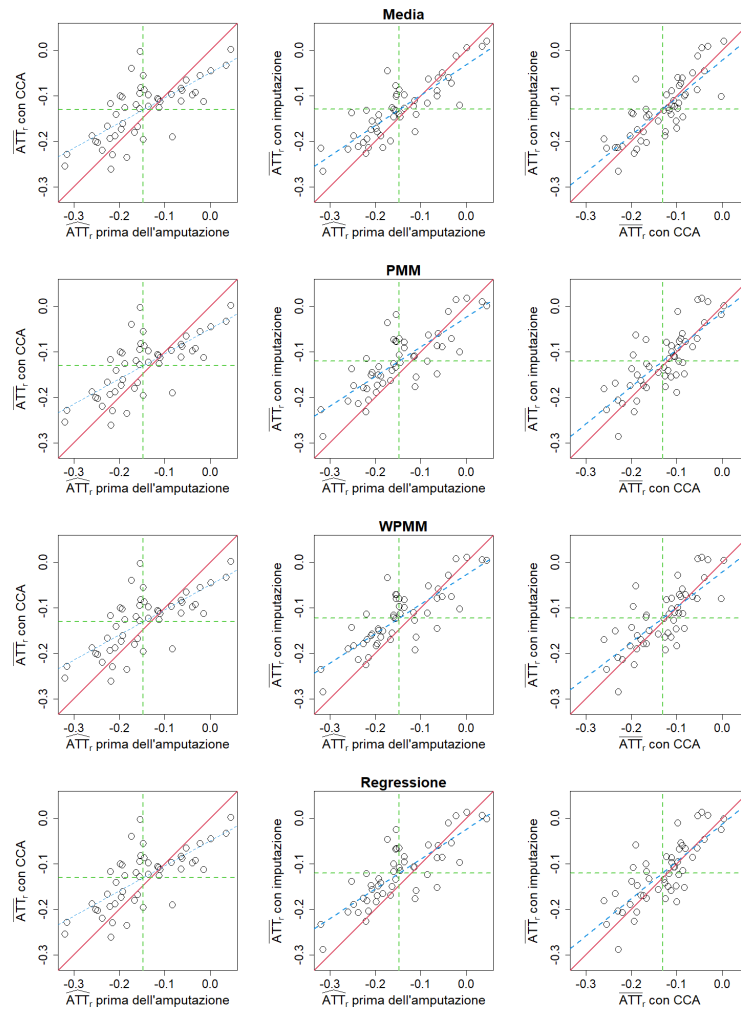


Figura A.29: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 3.

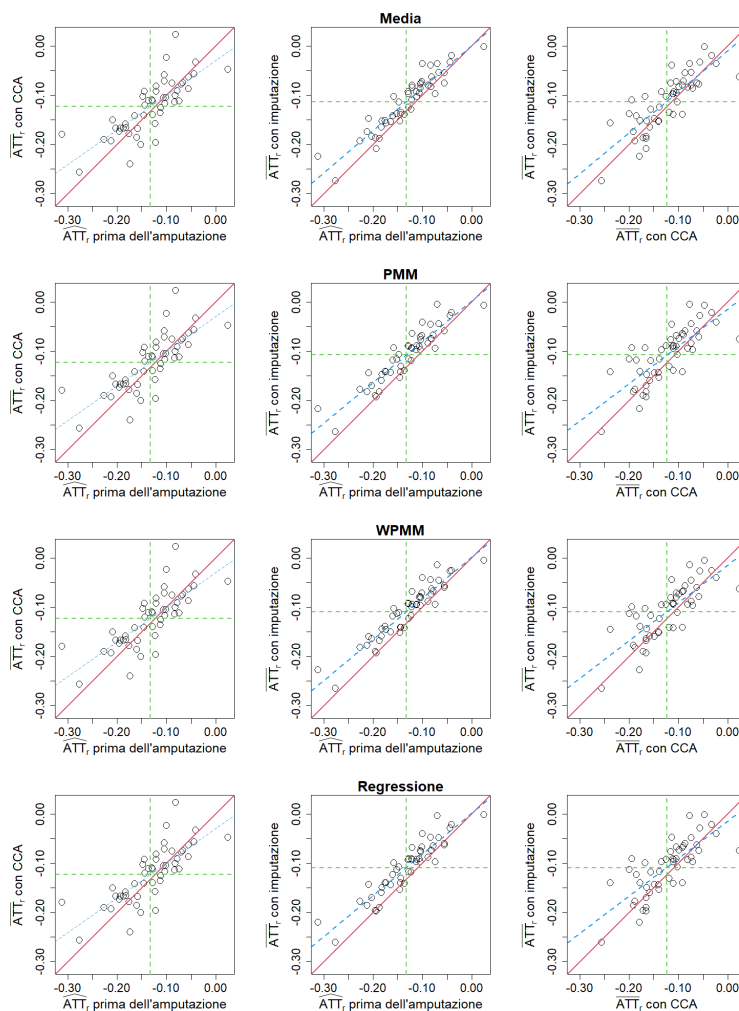


Figura A.30: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 3. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

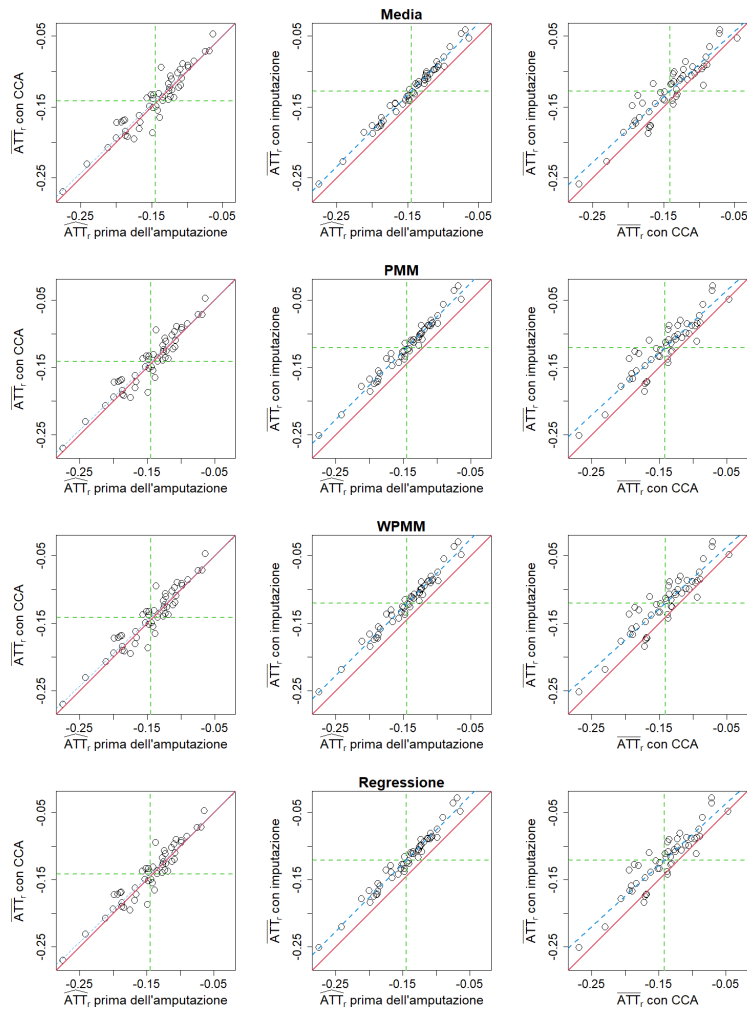


Figura A.31: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 3.

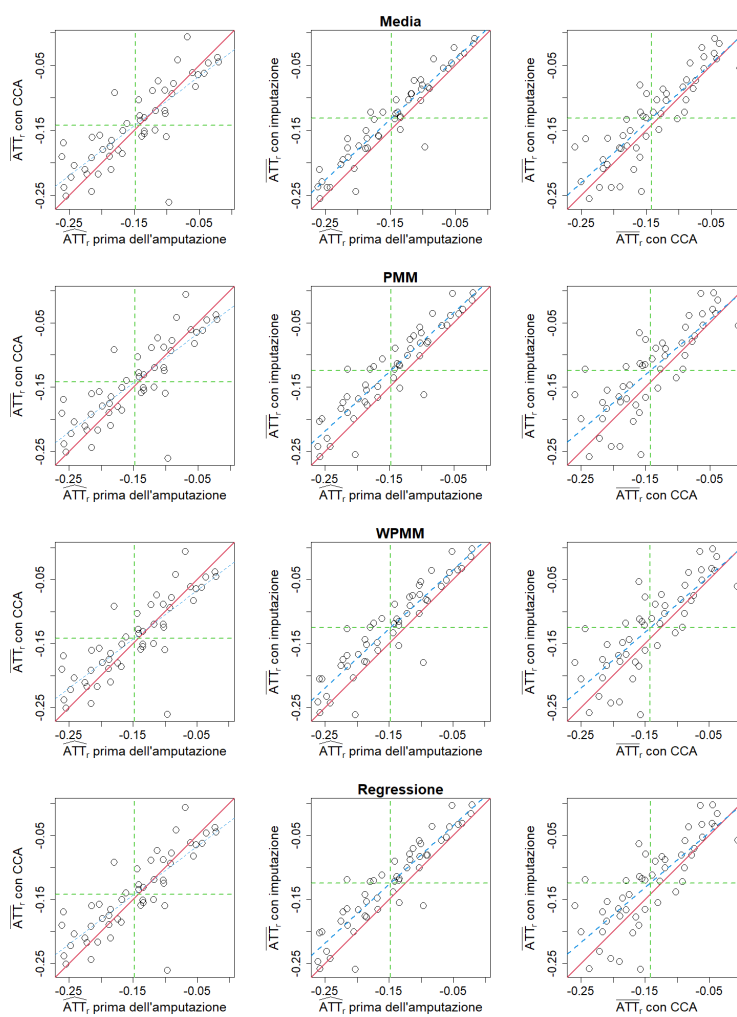


Figura A.32: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 3.

A.9 Simulazione 3 Scenario B

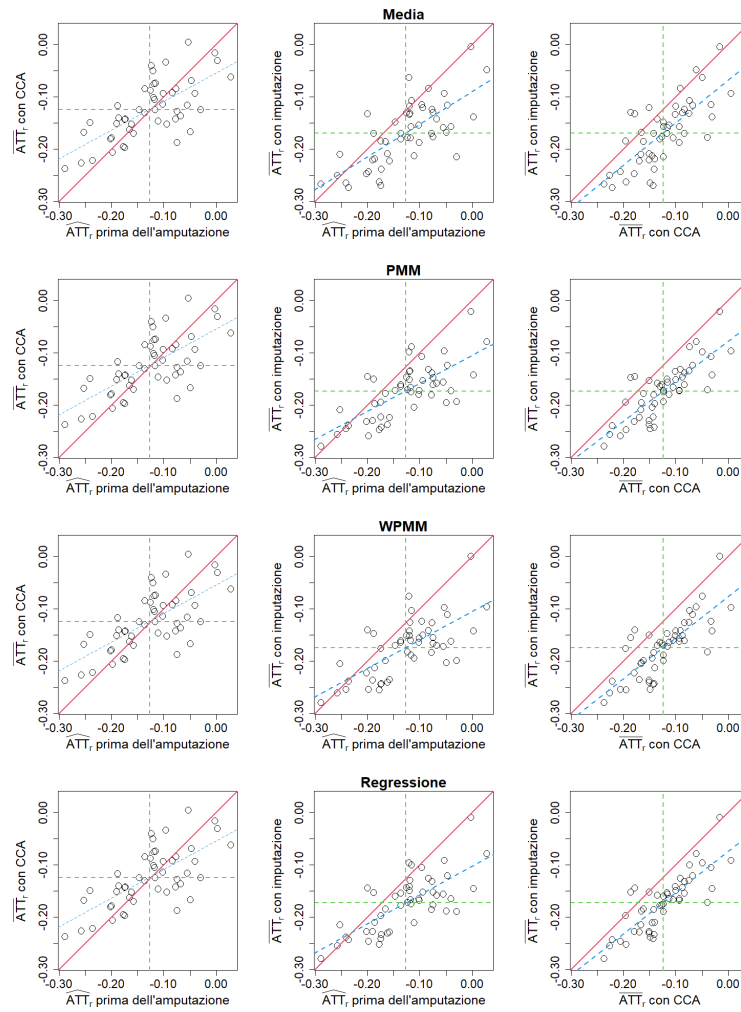


Figura A.33: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 3.

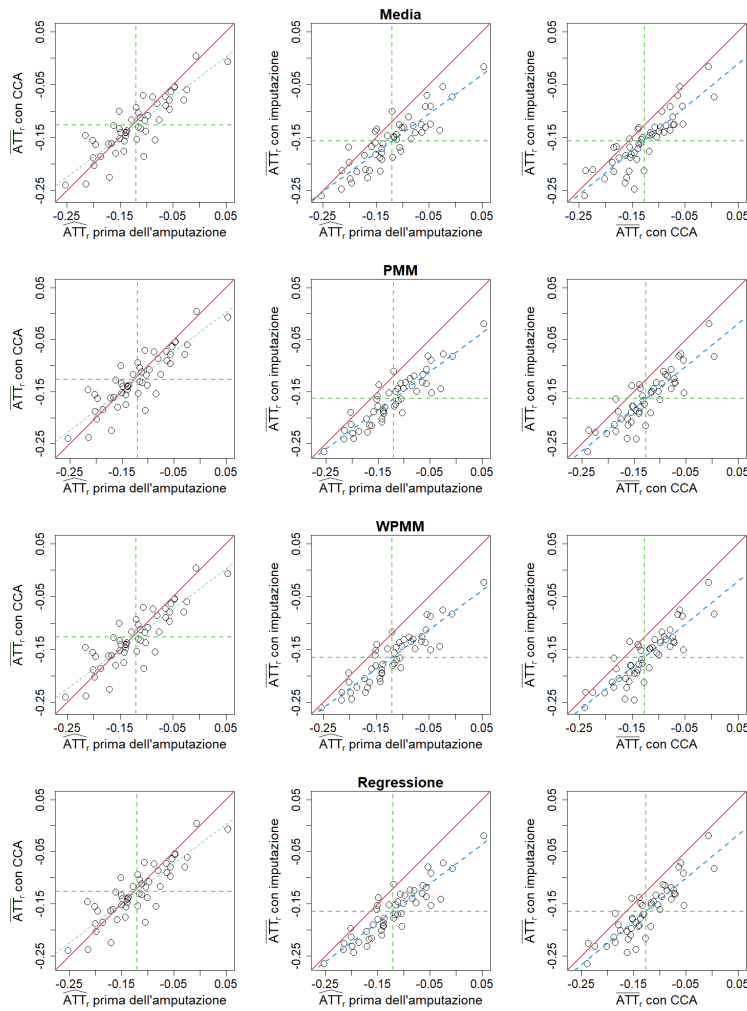


Figura A.34: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 3. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

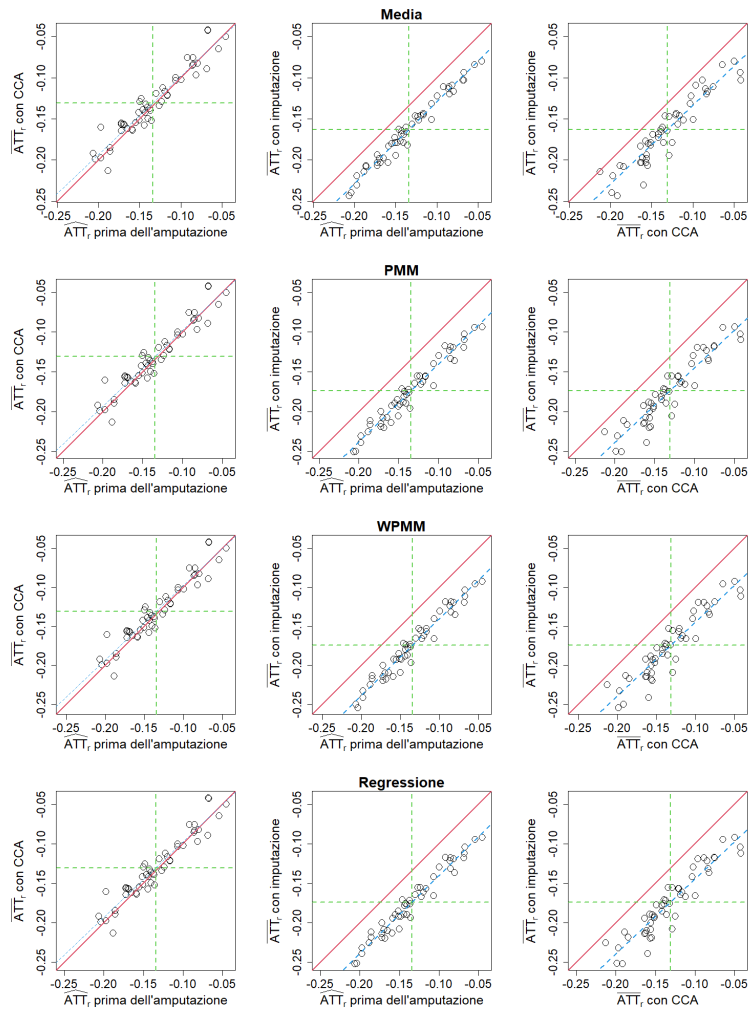


Figura A.35: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 3.

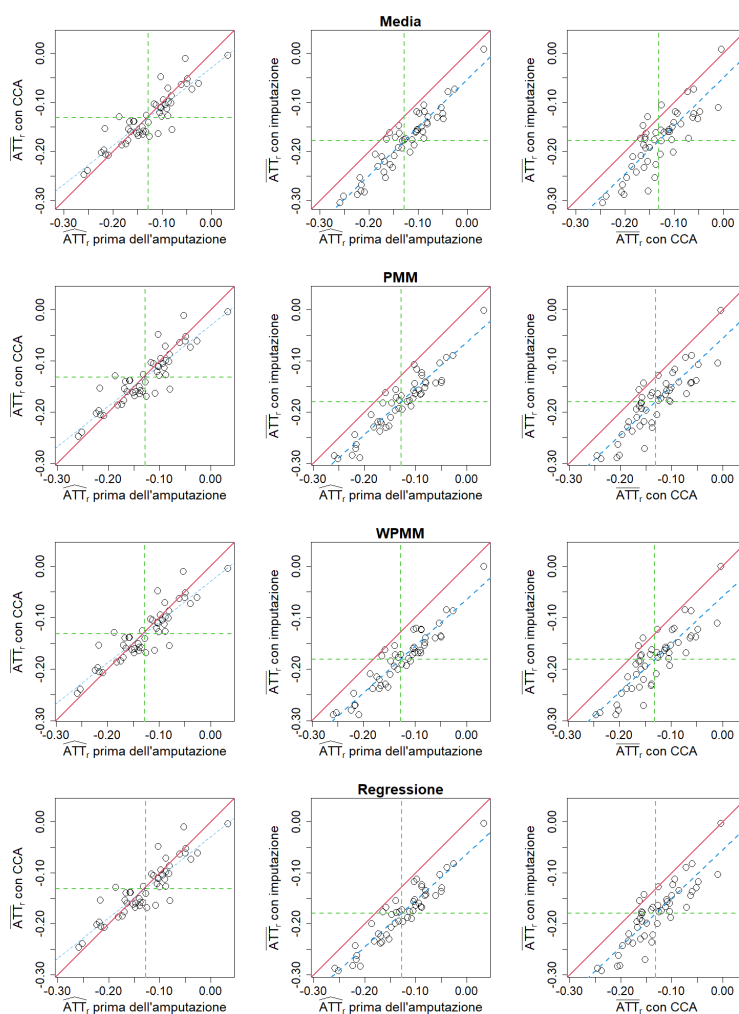


Figura A.36: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 3.

A.10 Simulazione 3 Scenario C

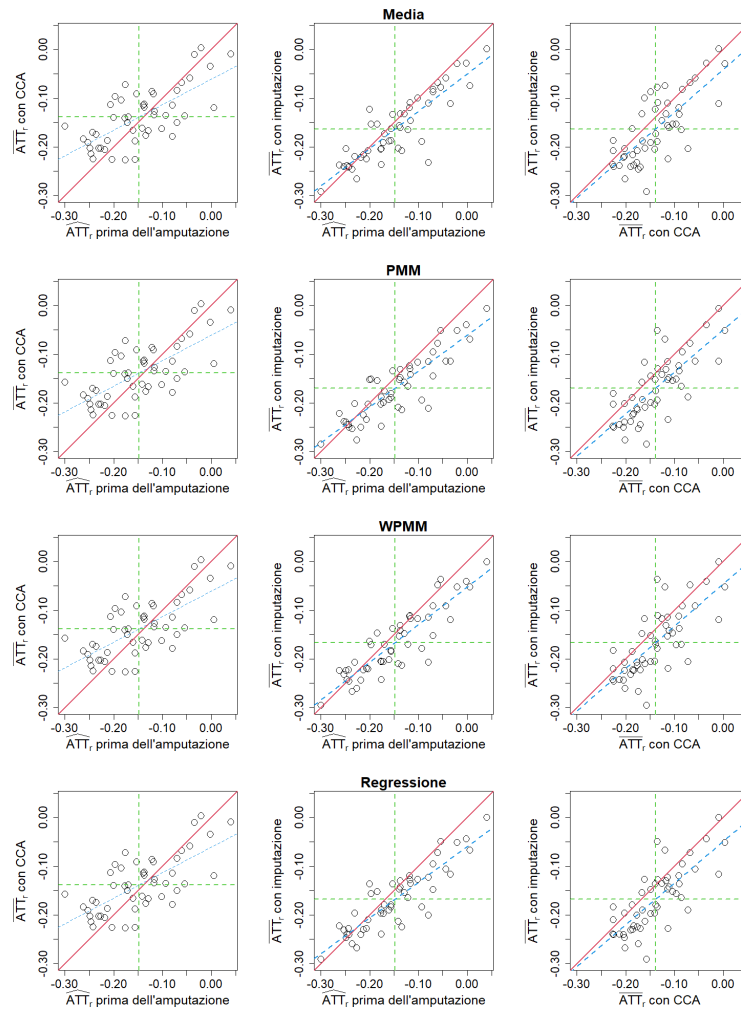


Figura A.37: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 3.

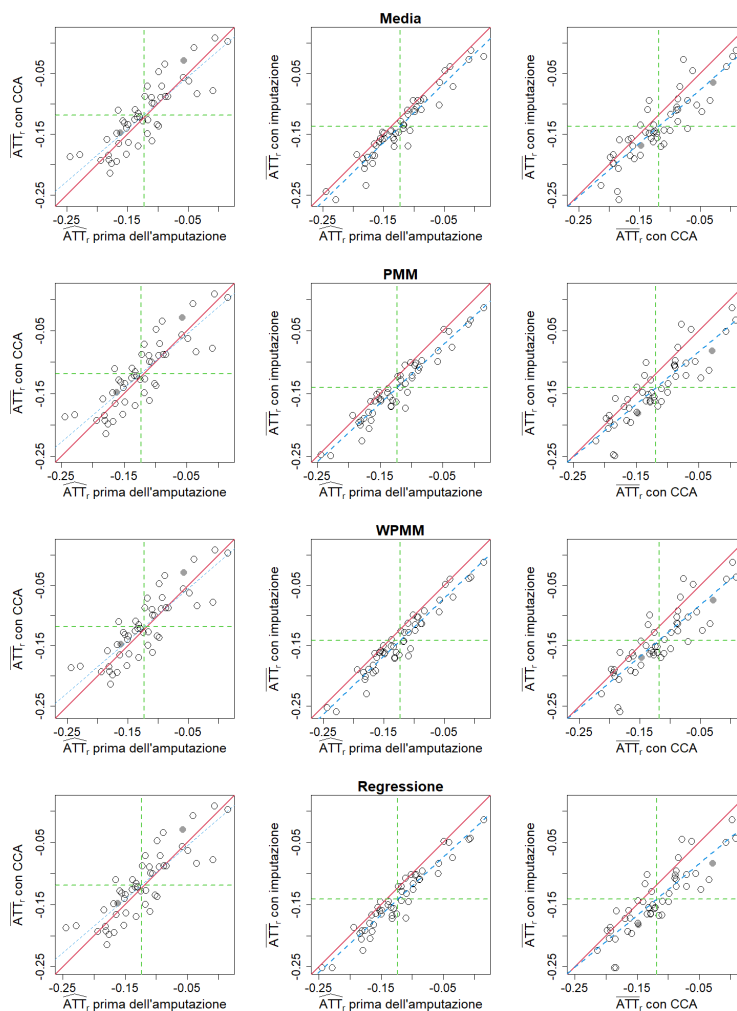


Figura A.38: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 3. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

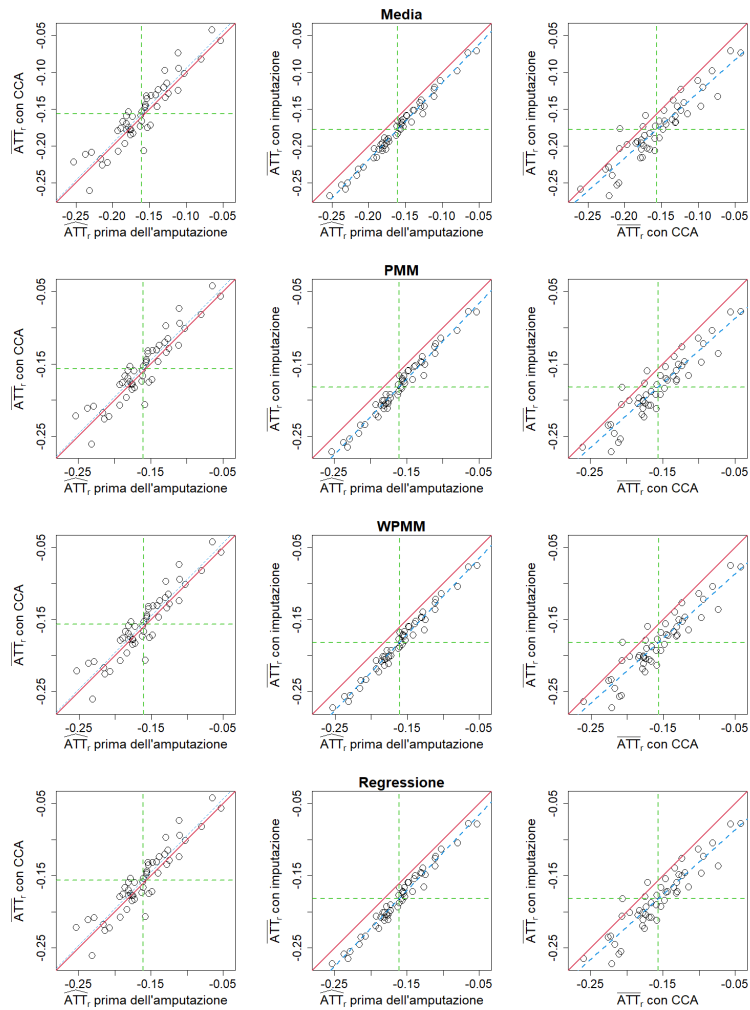


Figura A.39: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 3.

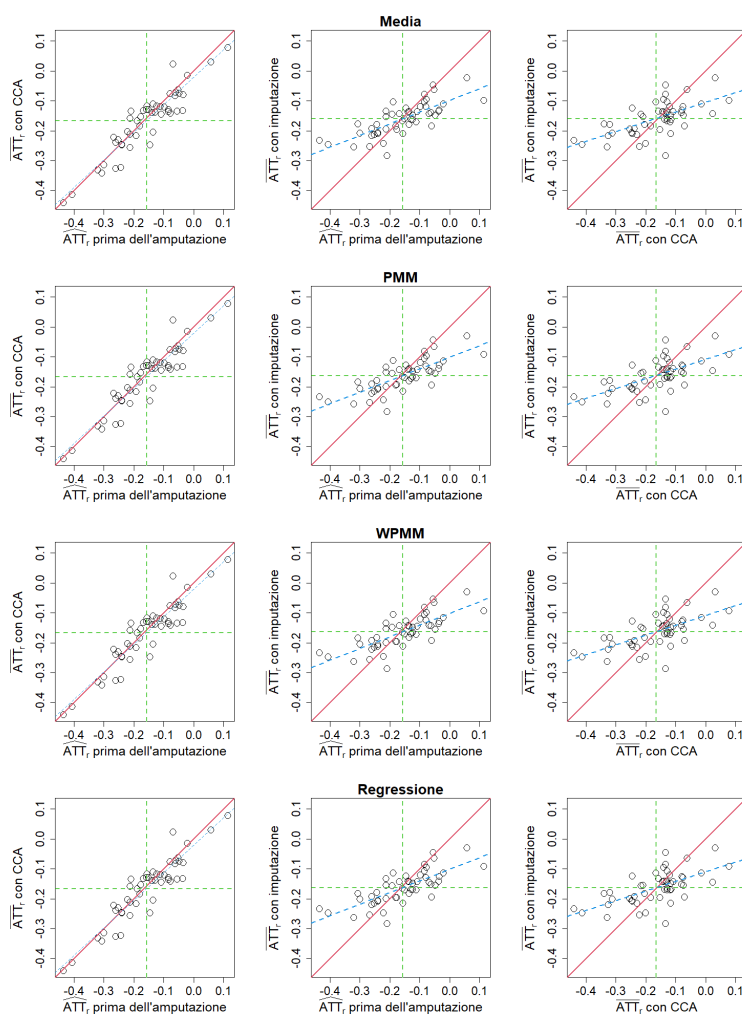


Figura A.40: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 3.

A.11 Simulazione 3 Scenario D

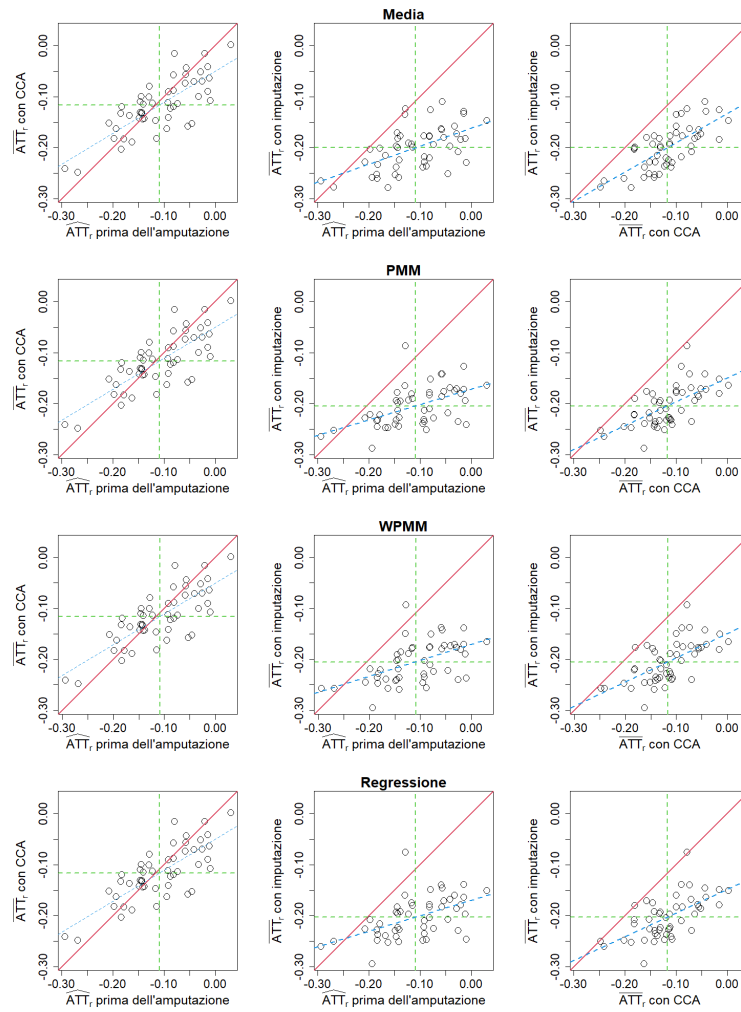


Figura A.41: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 3.

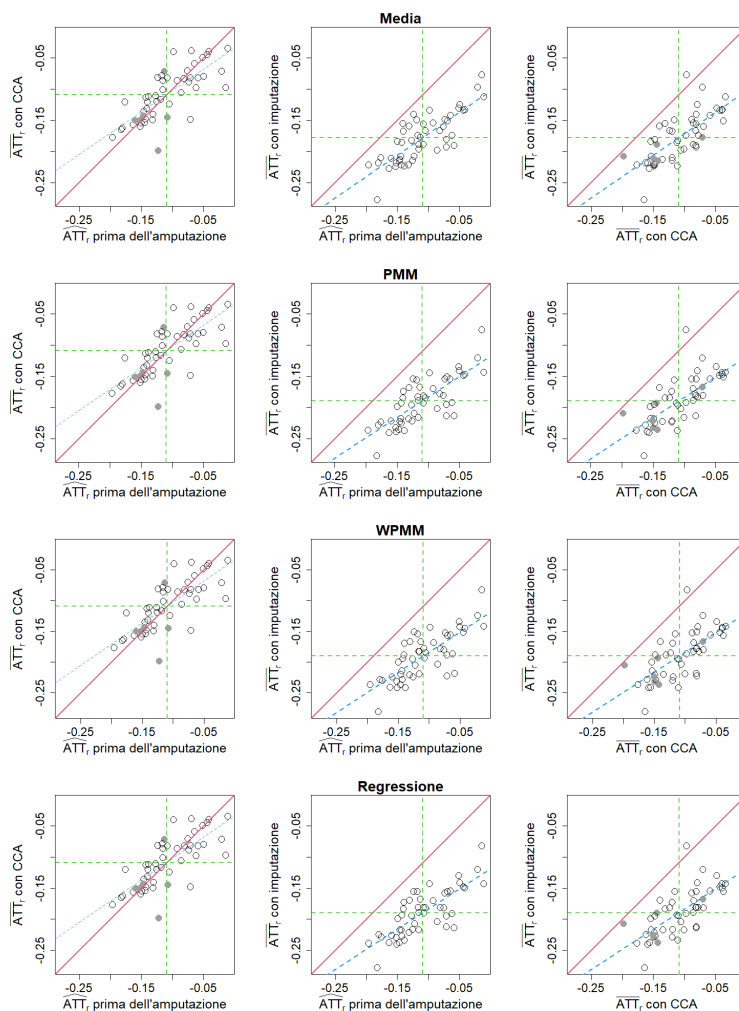


Figura A.42: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 3. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

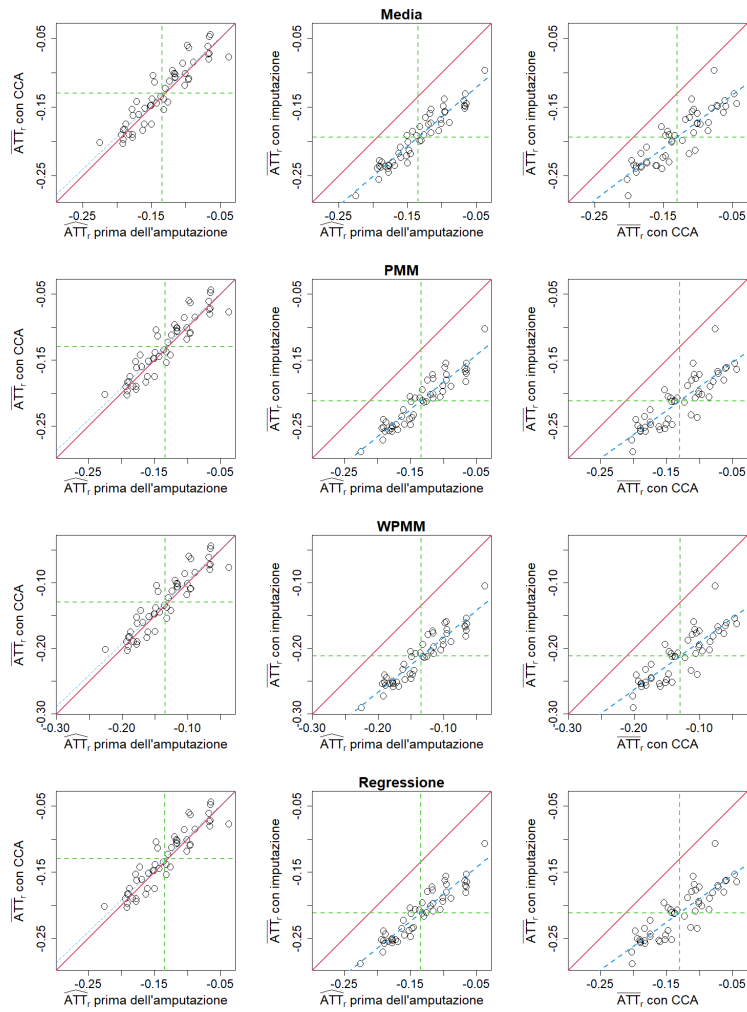


Figura A.43: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 3.

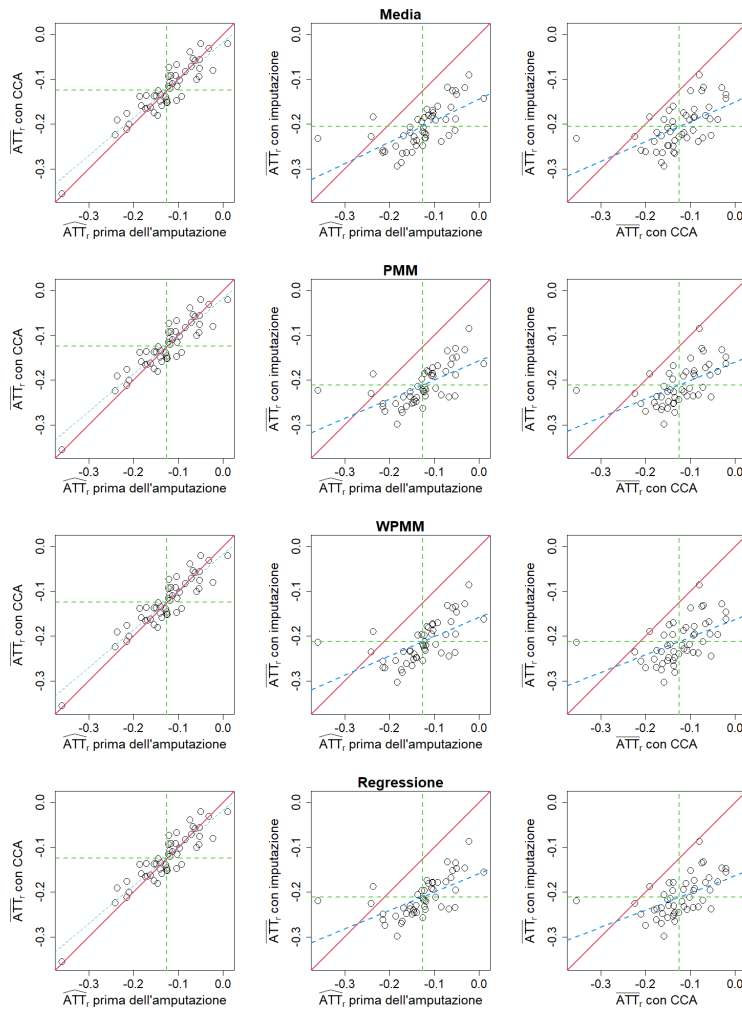


Figura A.44: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 3.

A.12 Simulazione 4 Scenario A

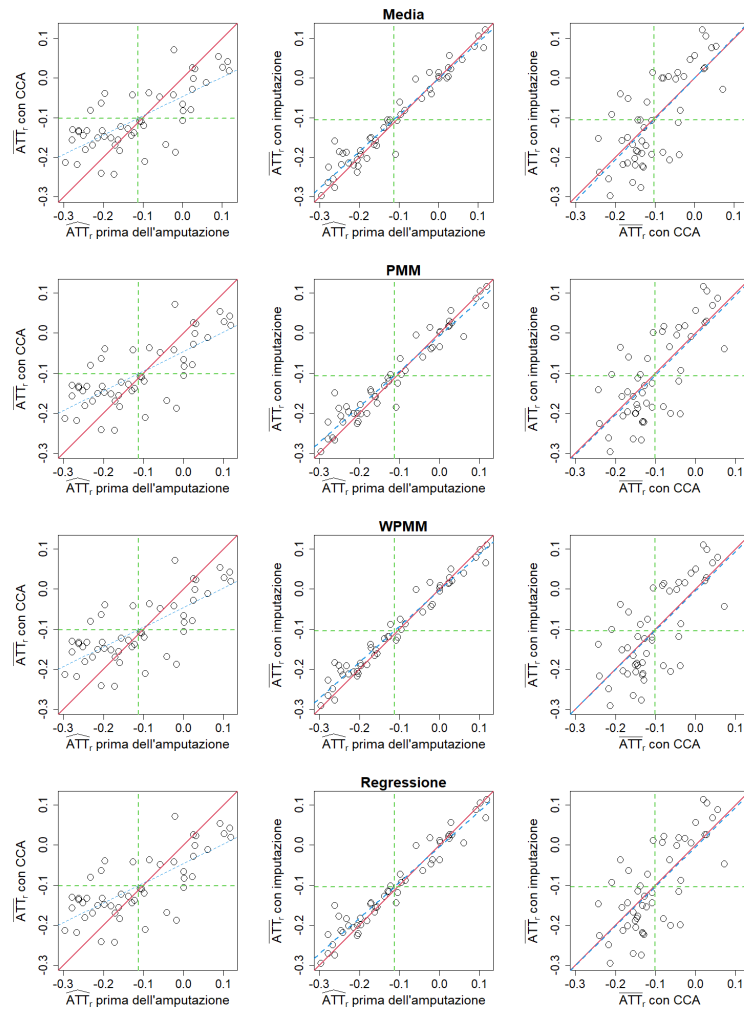


Figura A.45: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 4.

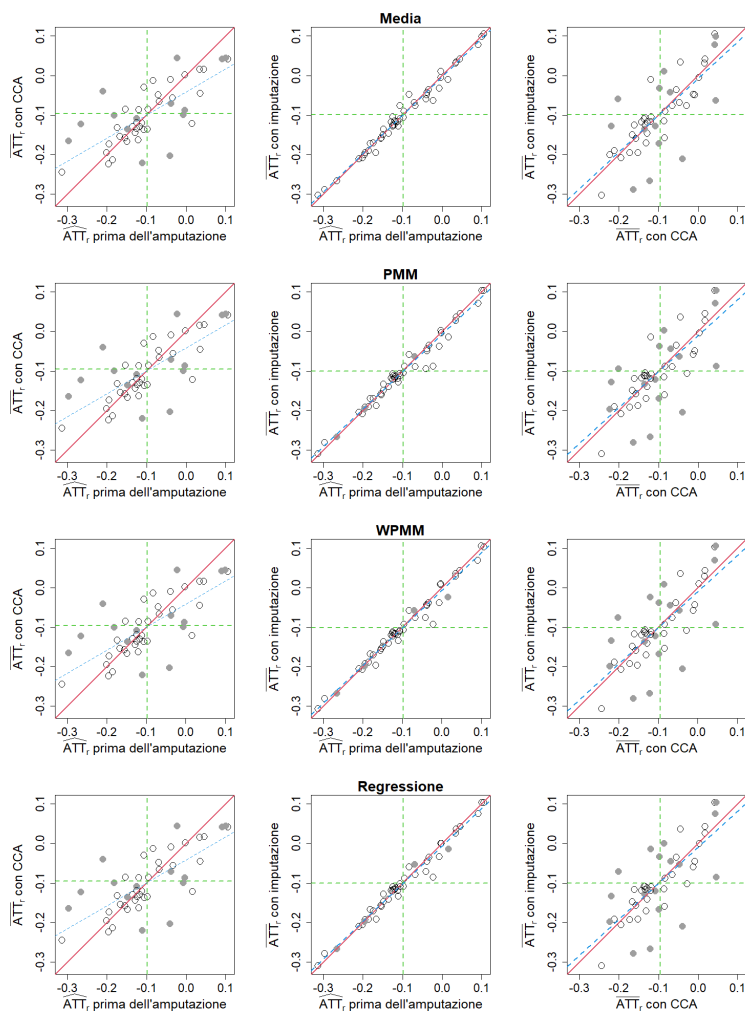


Figura A.46: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 4. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

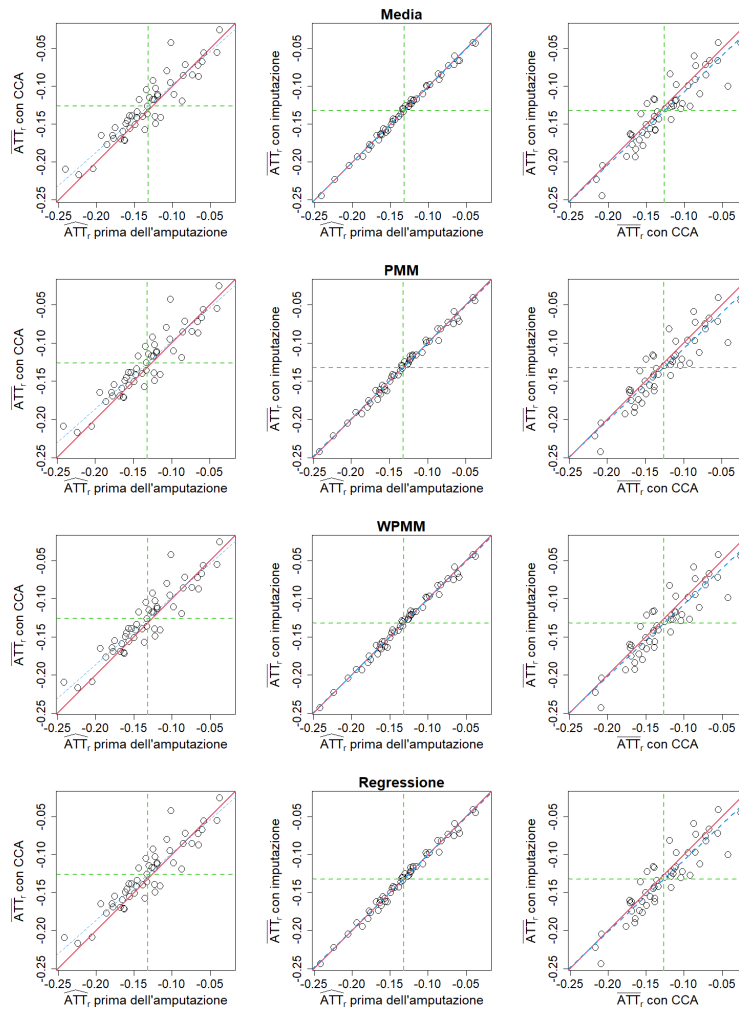


Figura A.47: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 4.

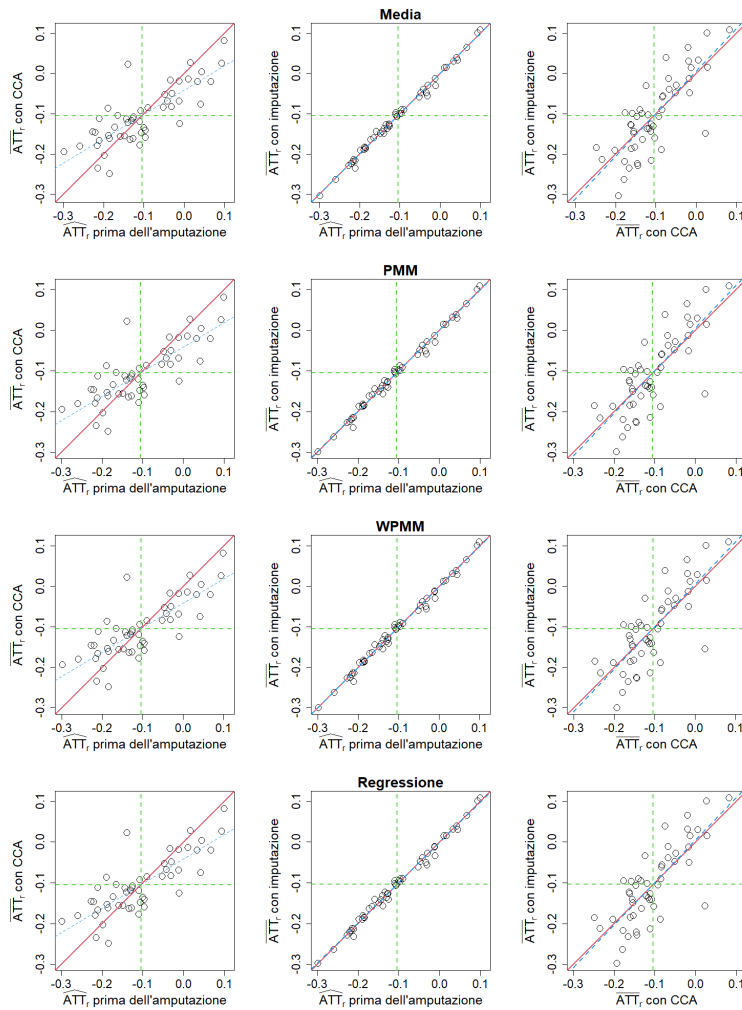


Figura A.48: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 4.

A.13 Simulazione 4 Scenario B

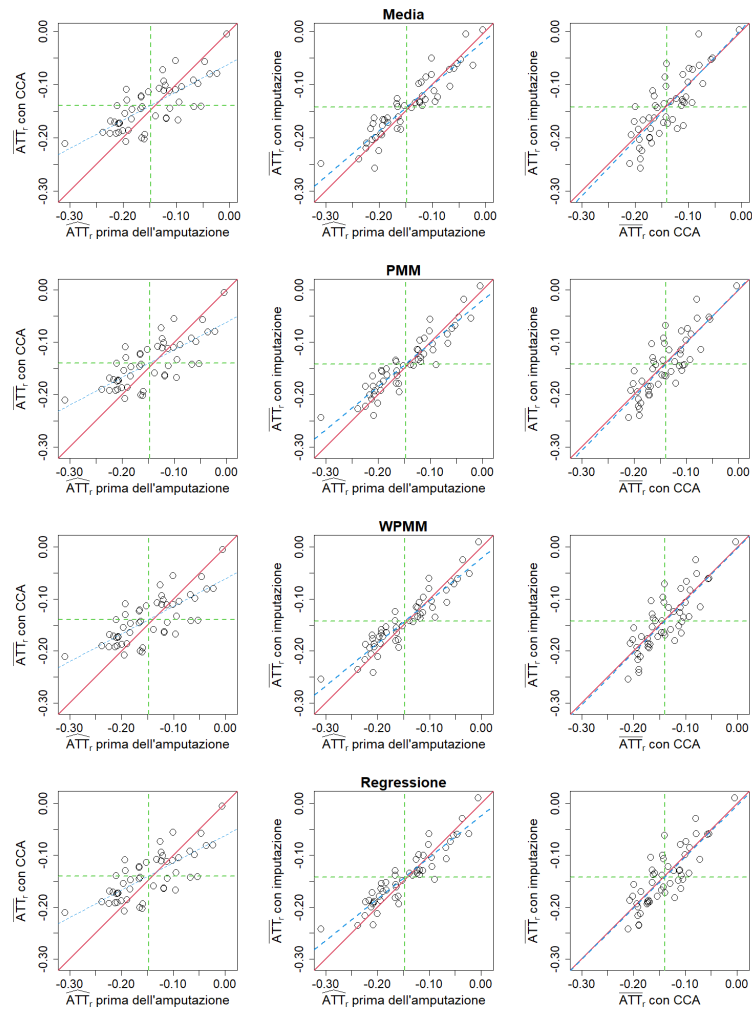


Figura A.49: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 4.

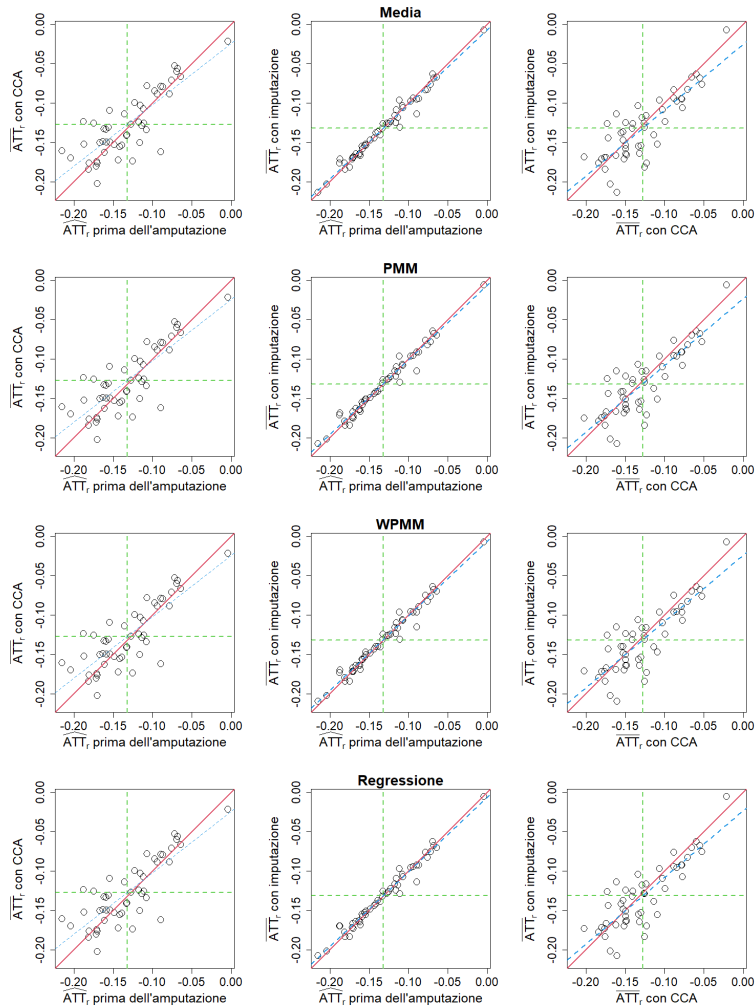


Figura A.50: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 4. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

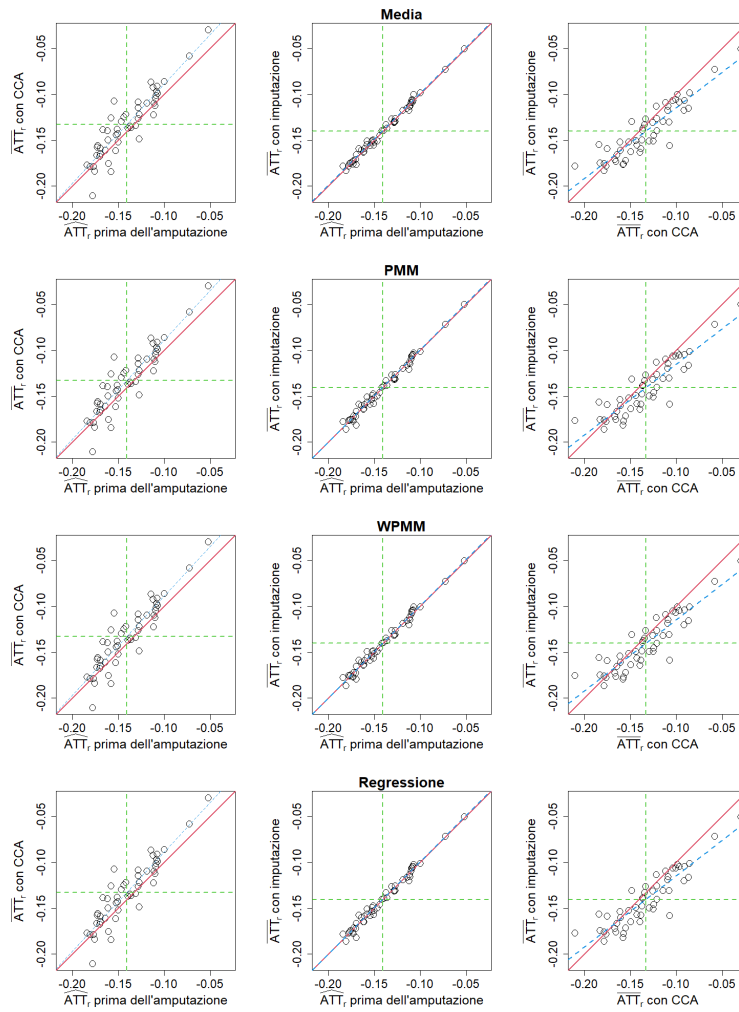


Figura A.51: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 4.

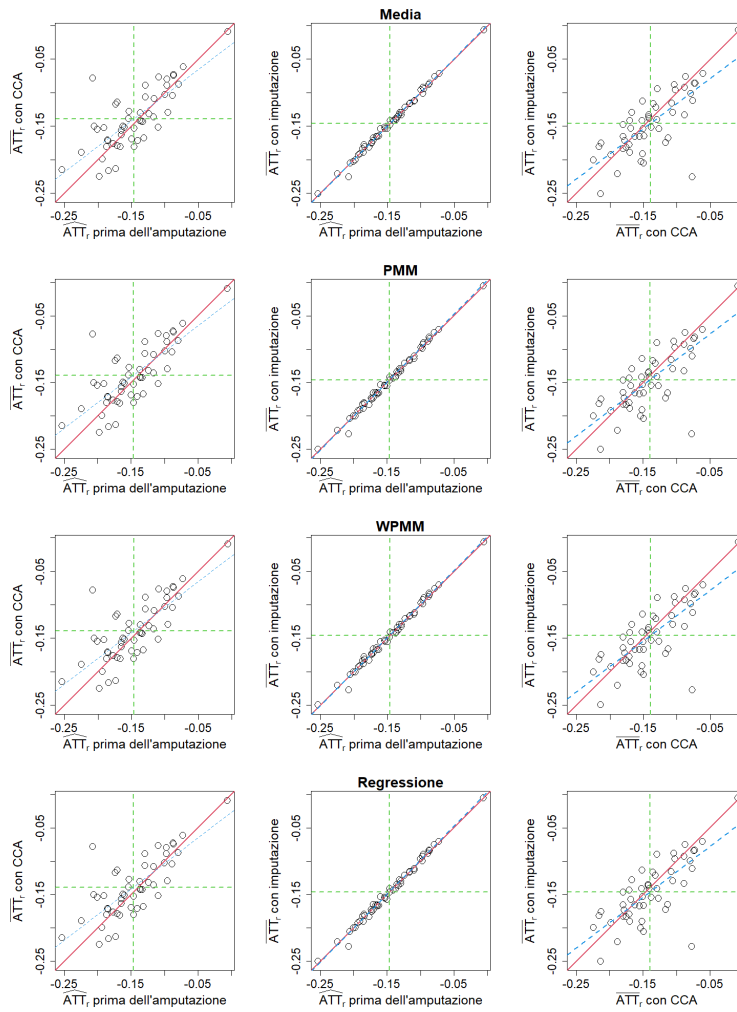


Figura A.52: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 4.

A.14 Simulazione 4 Scenario C

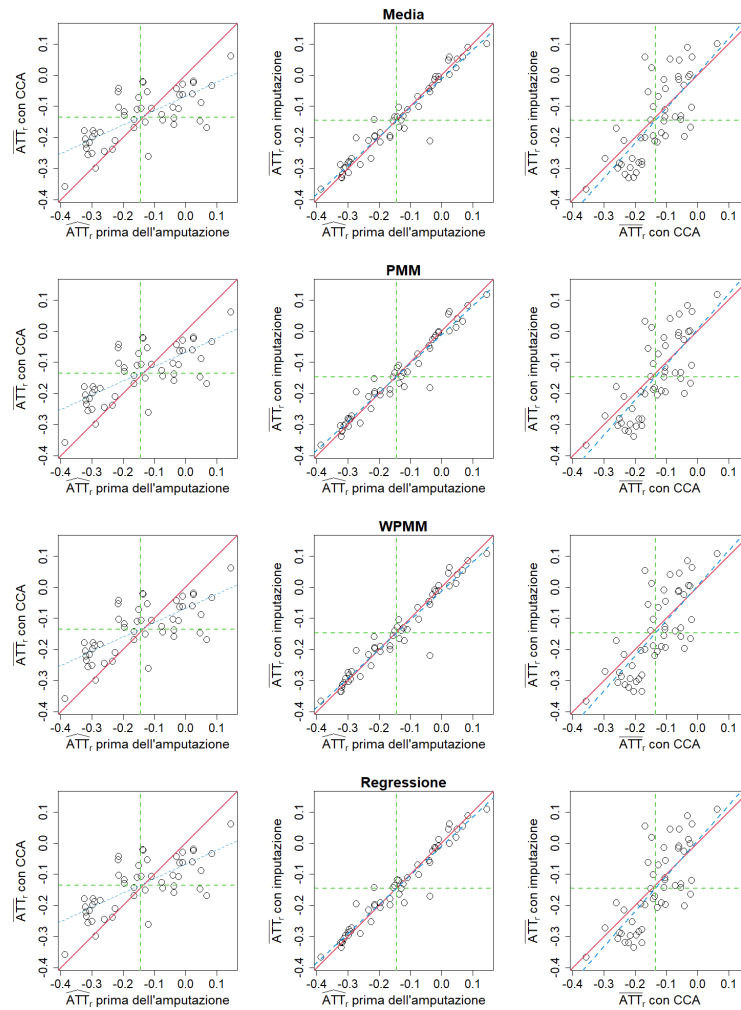


Figura A.53: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 4.

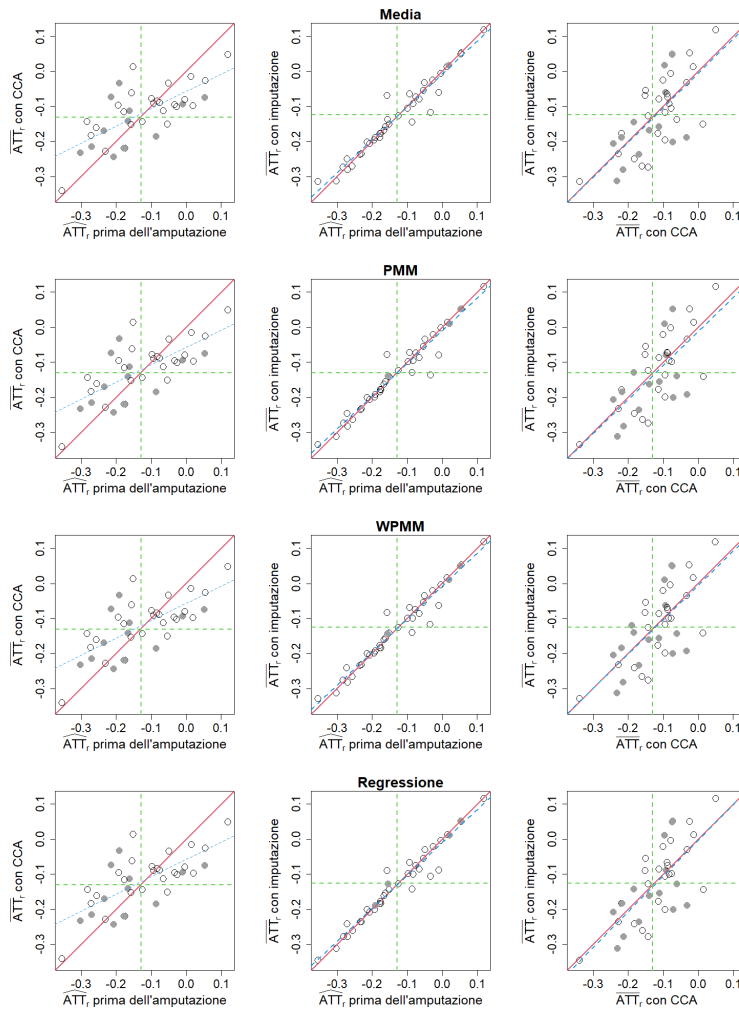


Figura A.54: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 4. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

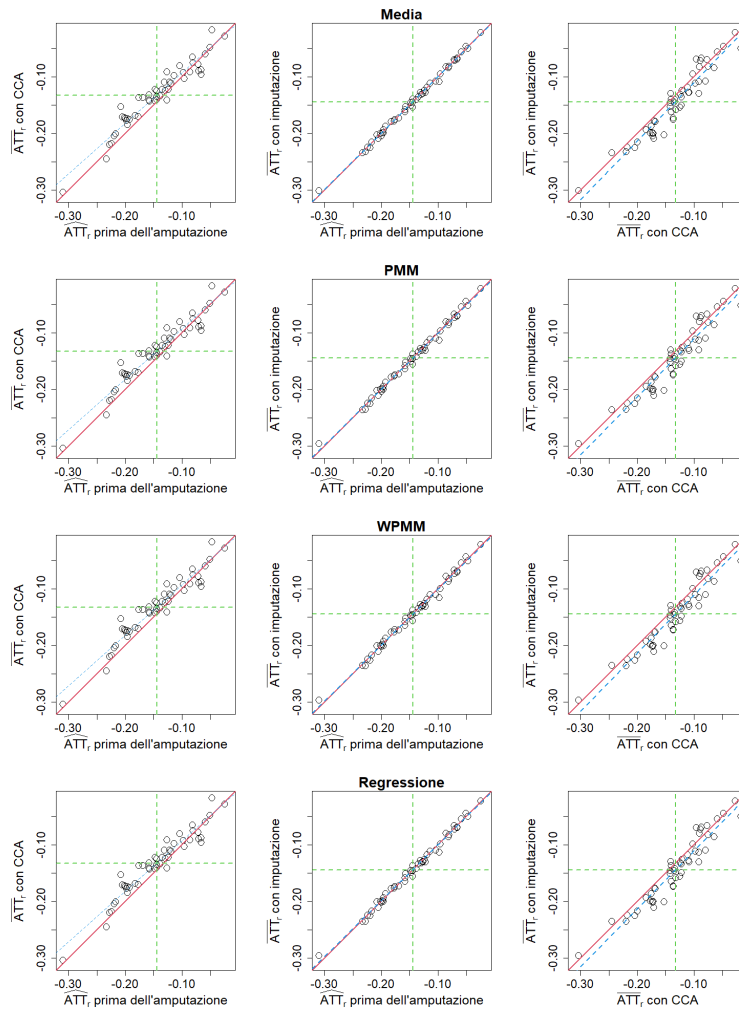


Figura A.55: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 4.

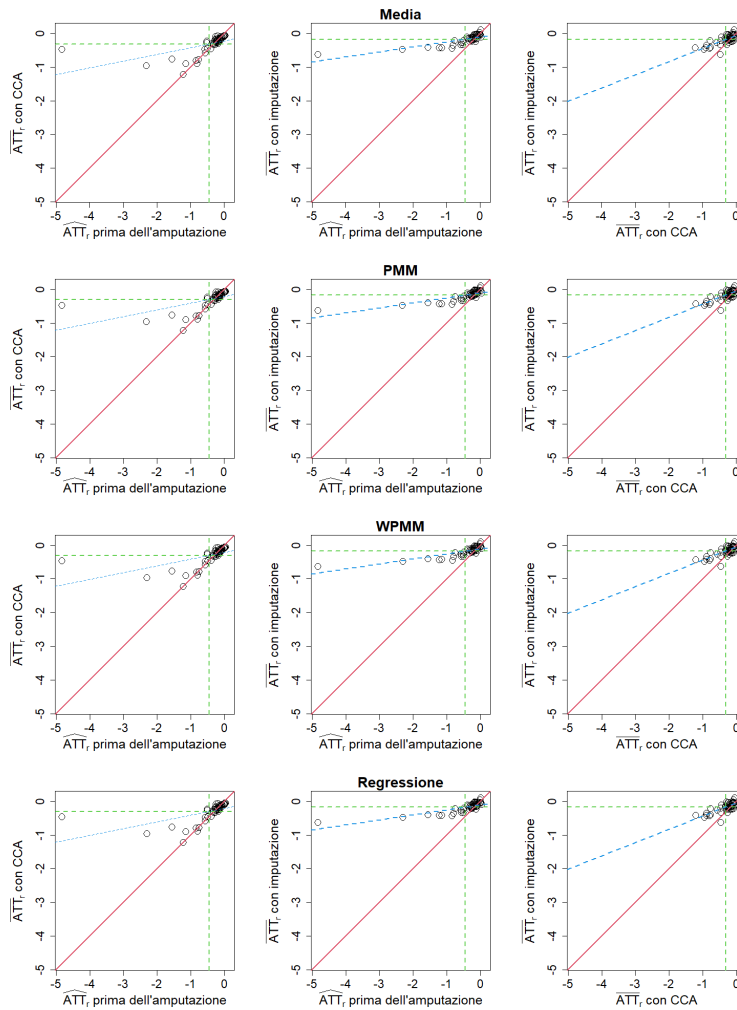


Figura A.56: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 4.

A.15 Simulazione 4 Scenario D

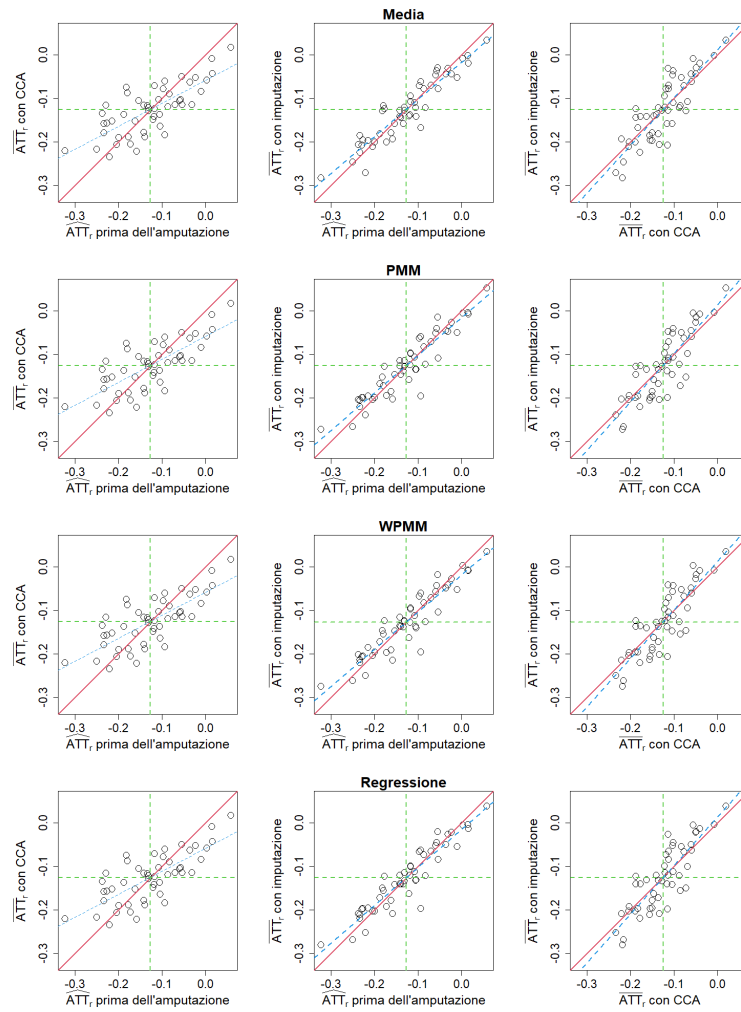


Figura A.57: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 4.

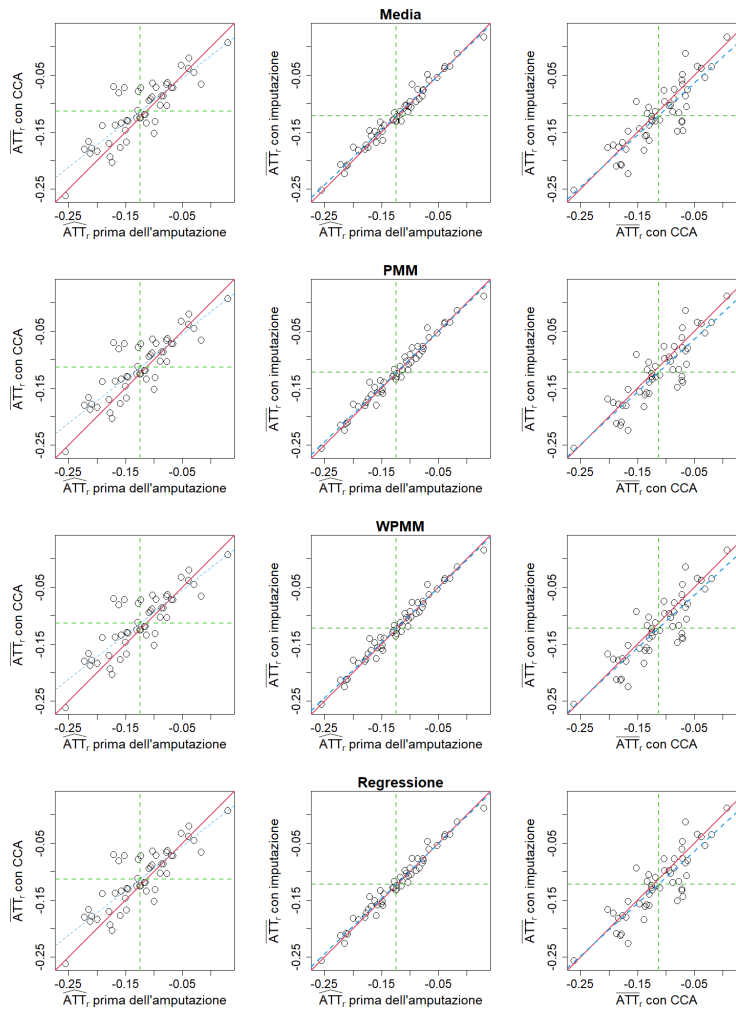


Figura A.58: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 4. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

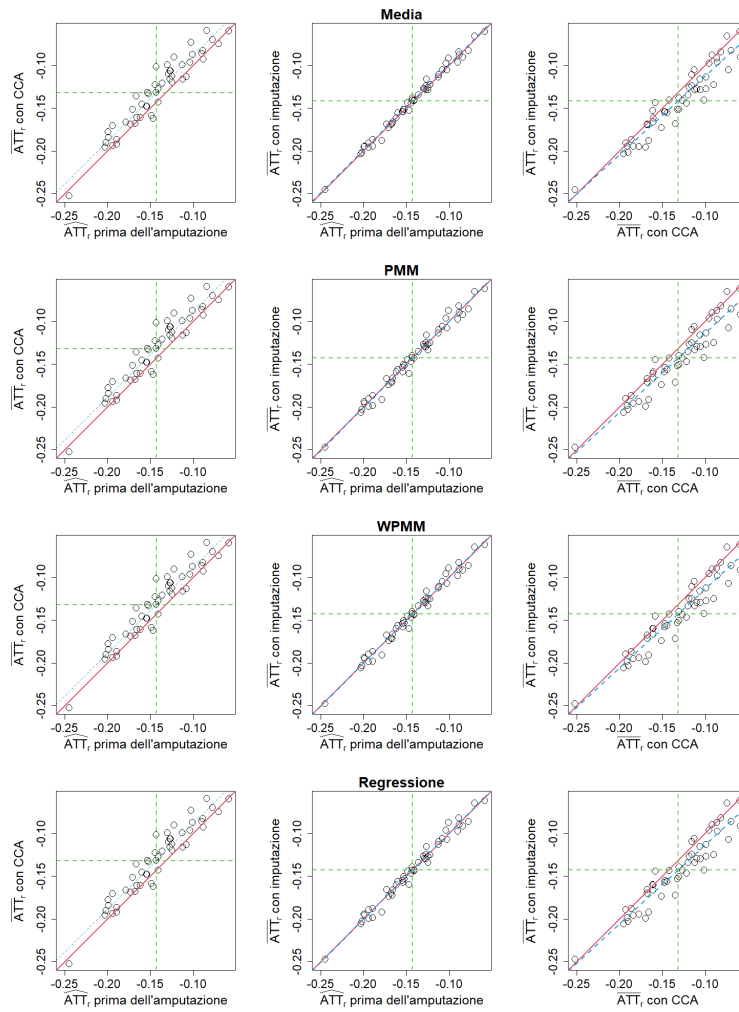


Figura A.59: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 4.

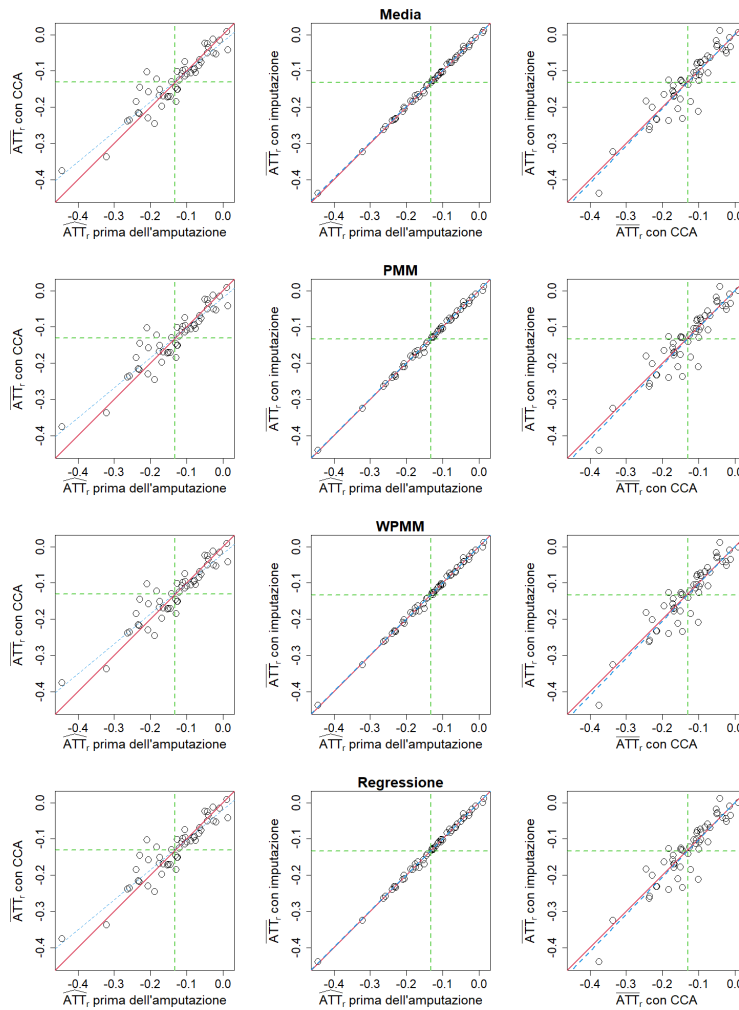


Figura A.60: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 4.

A.16 Simulazione 5 Scenario A

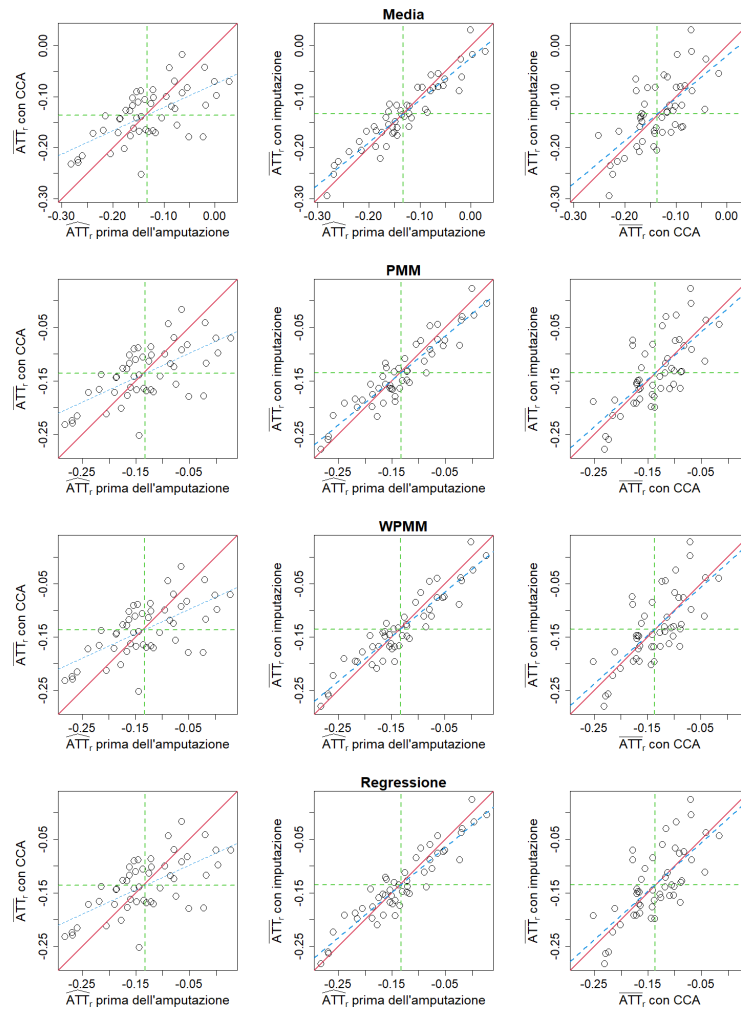


Figura A.61: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 5.

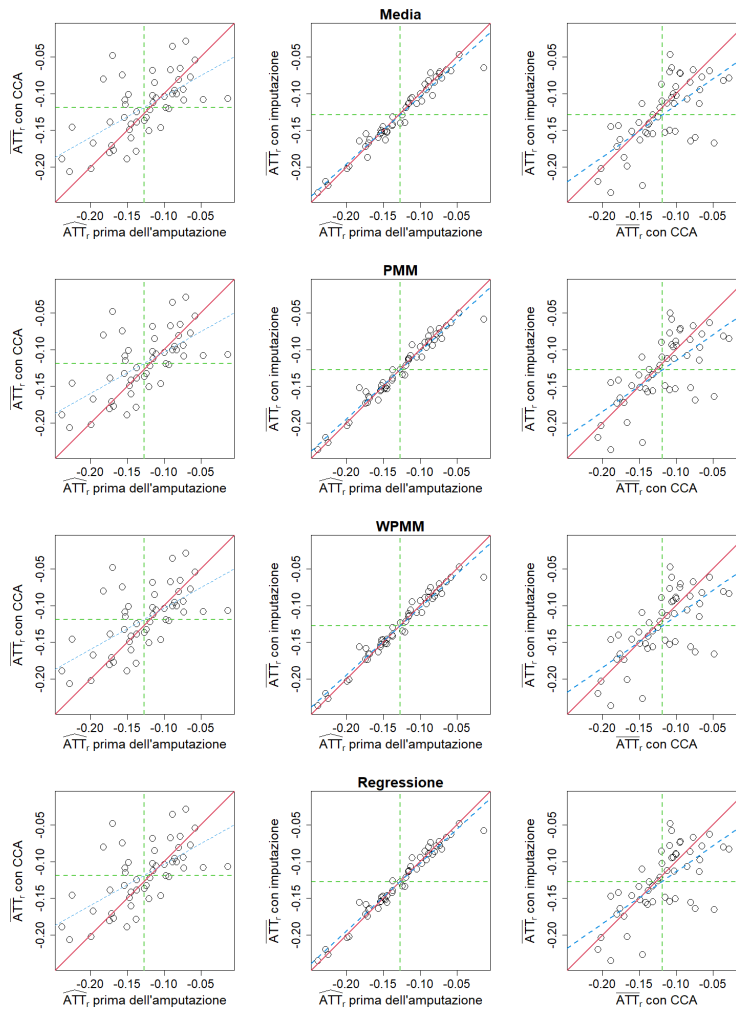


Figura A.62: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 5. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

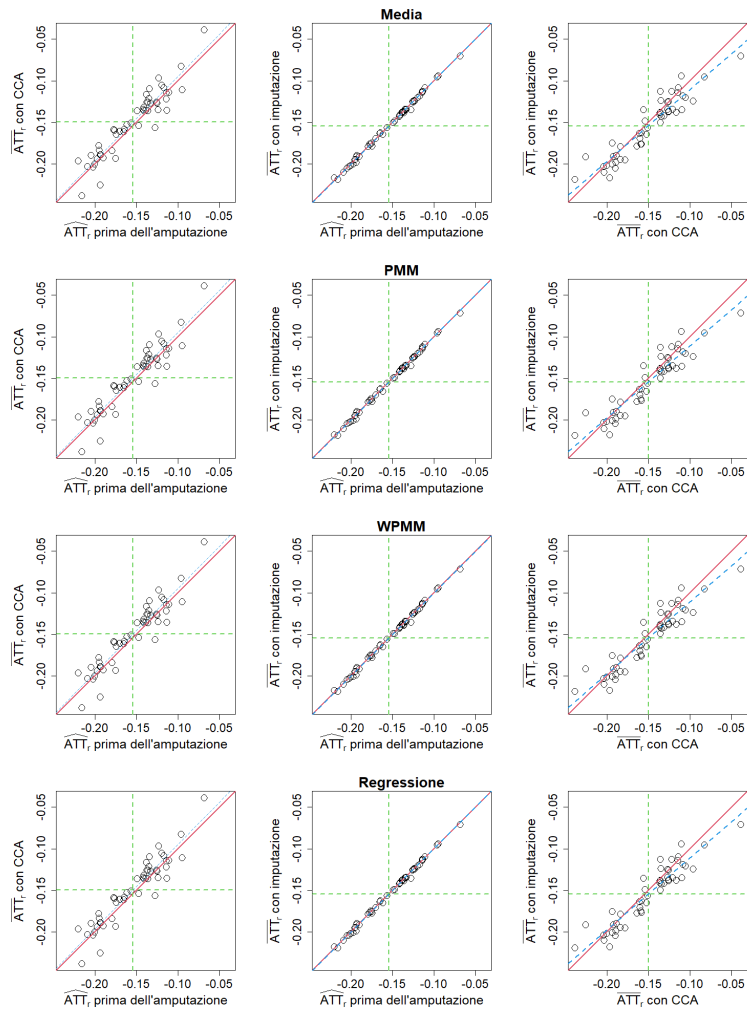


Figura A.63: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 5.

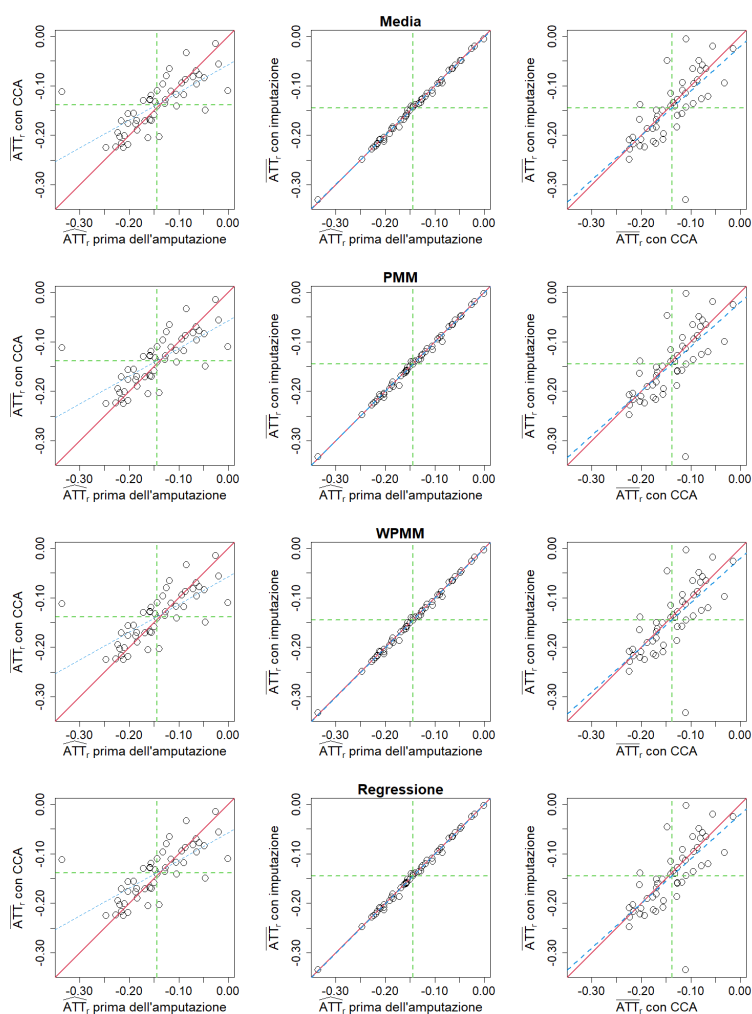


Figura A.64: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 5.

A.17 Simulazione 5 Scenario B

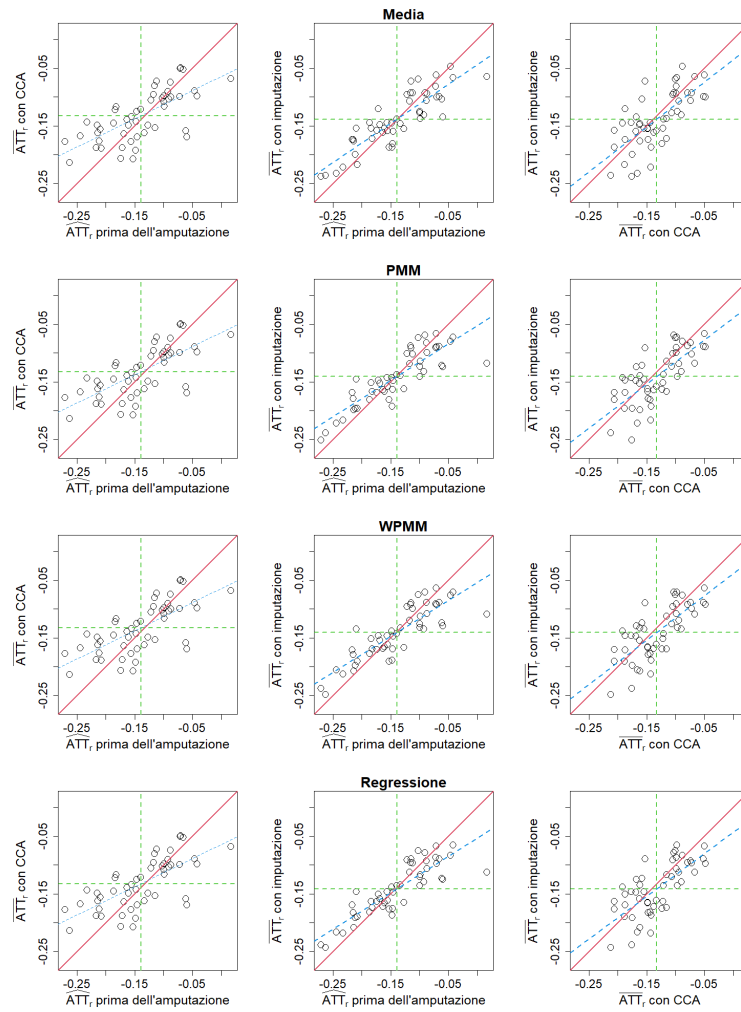


Figura A.65: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 5.

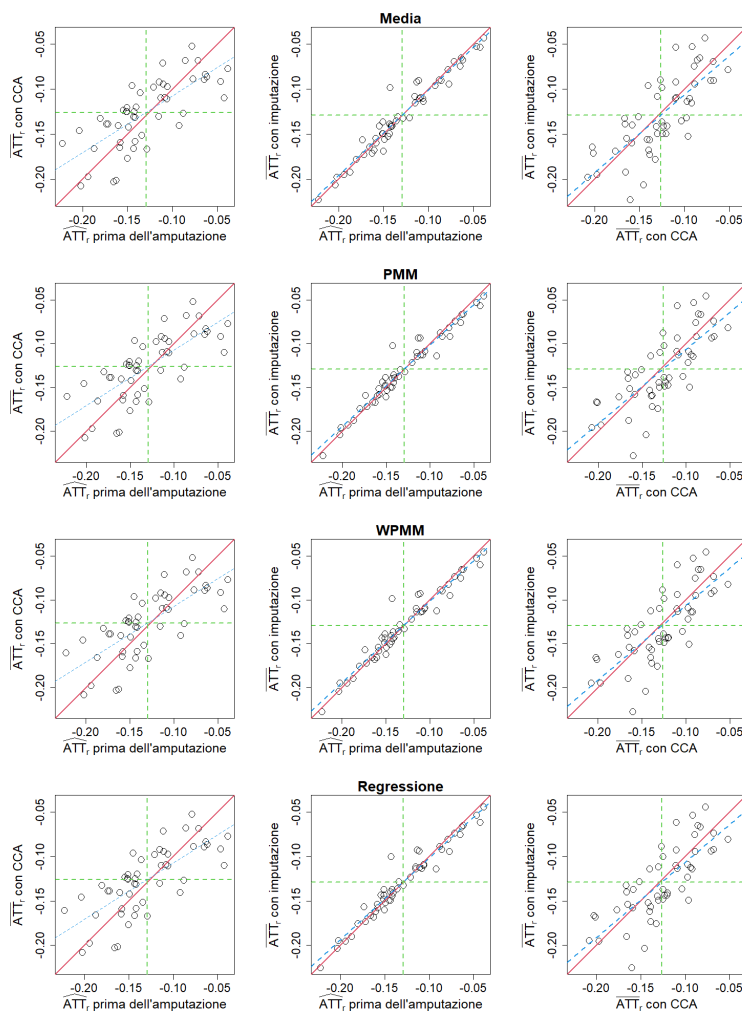


Figura A.66: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 5. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

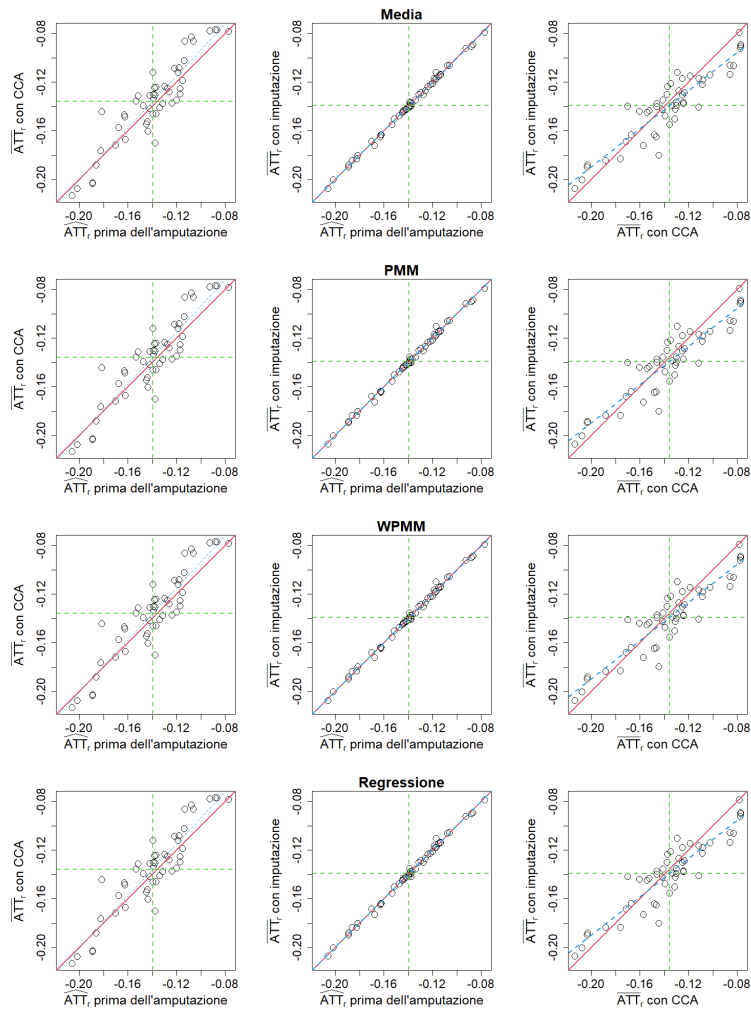


Figura A.67: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 5.

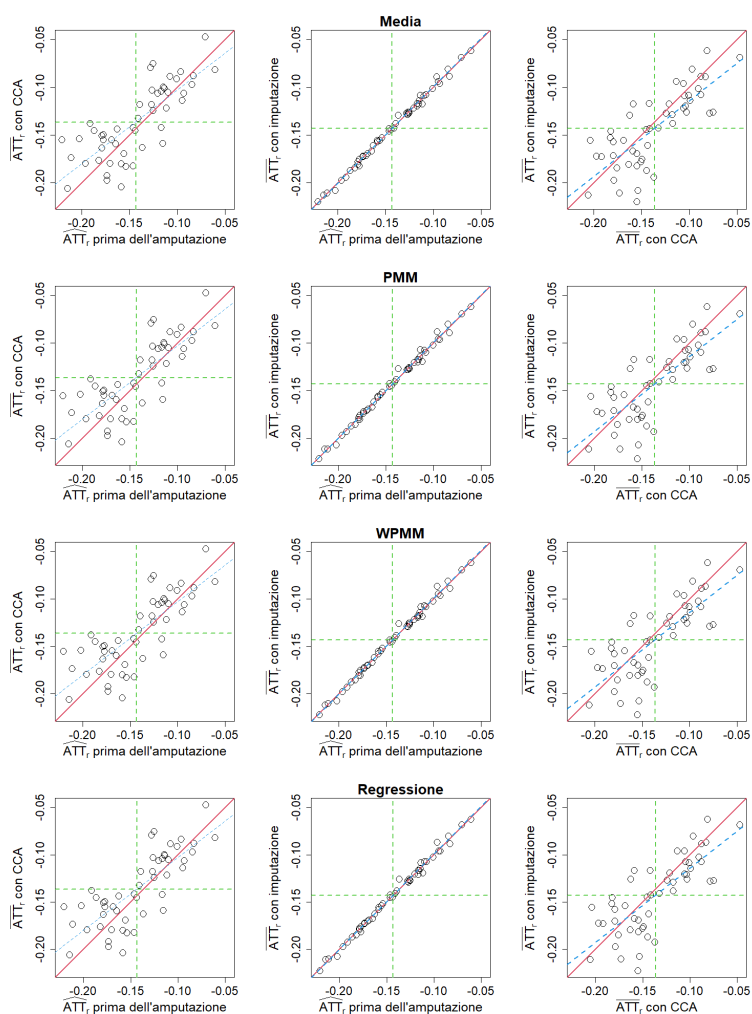


Figura A.68: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 5.

A.18 Simulazione 5 Scenario C

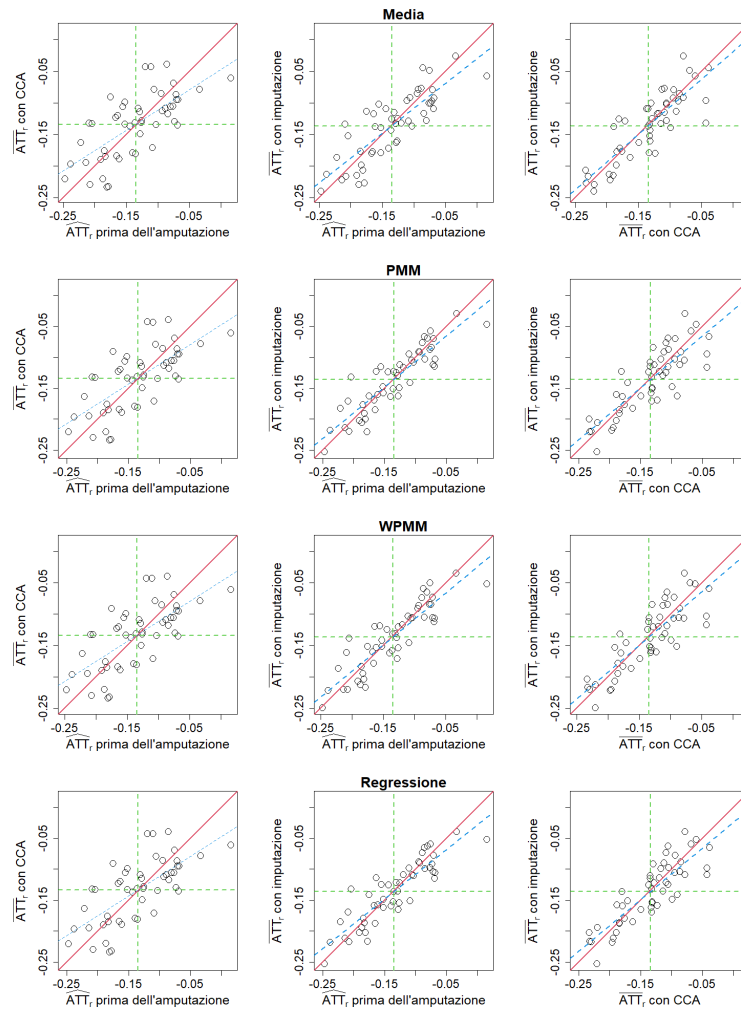


Figura A.69: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 5.

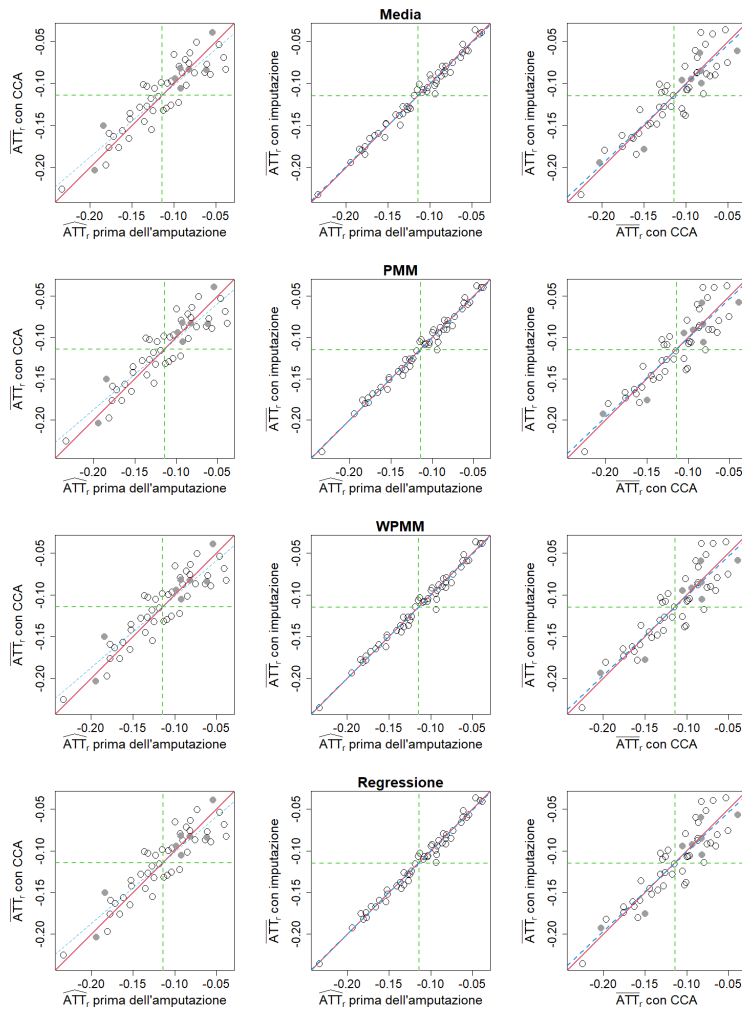


Figura A.70: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 5. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

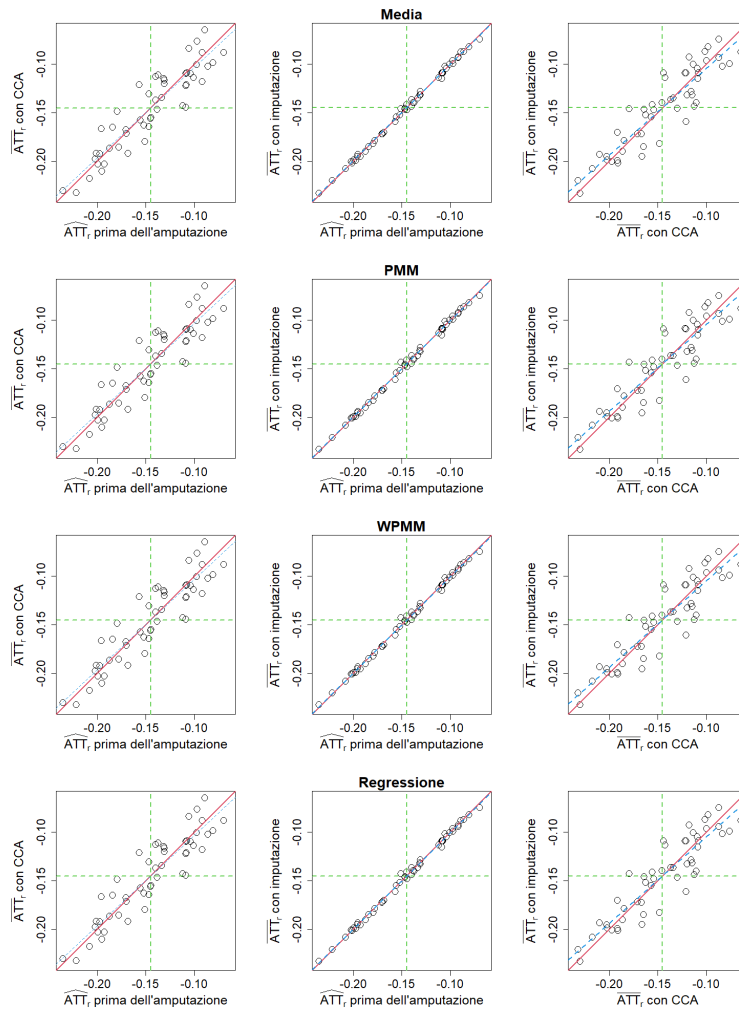


Figura A.71: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 5.

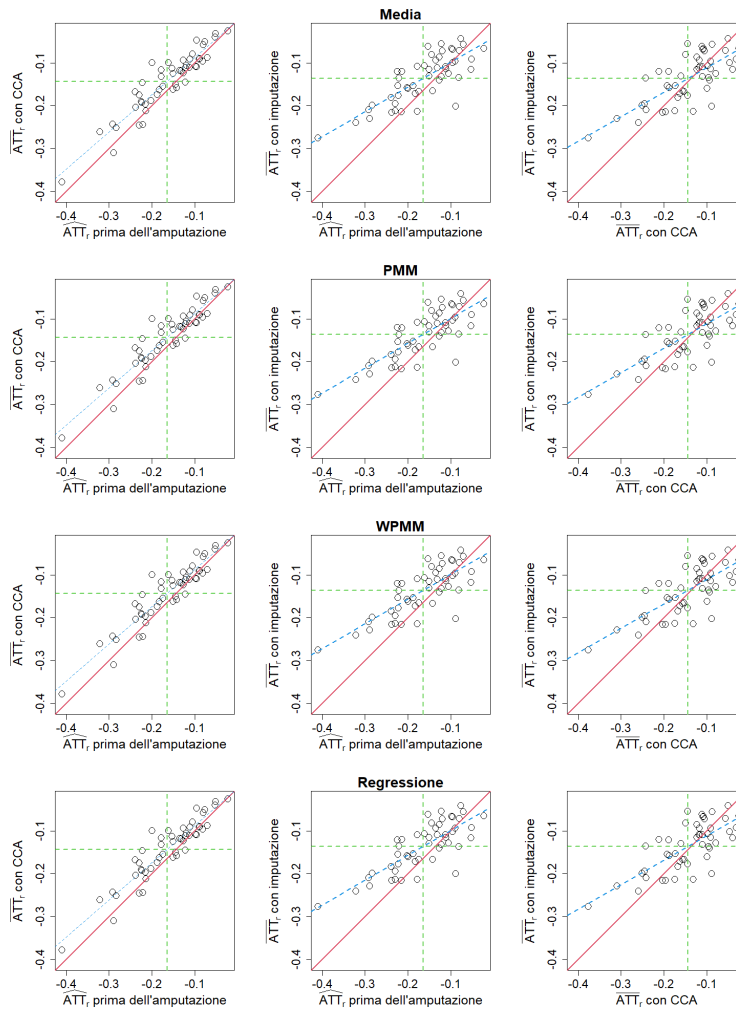


Figura A.72: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 5.

A.19 Simulazione 5 Scenario D

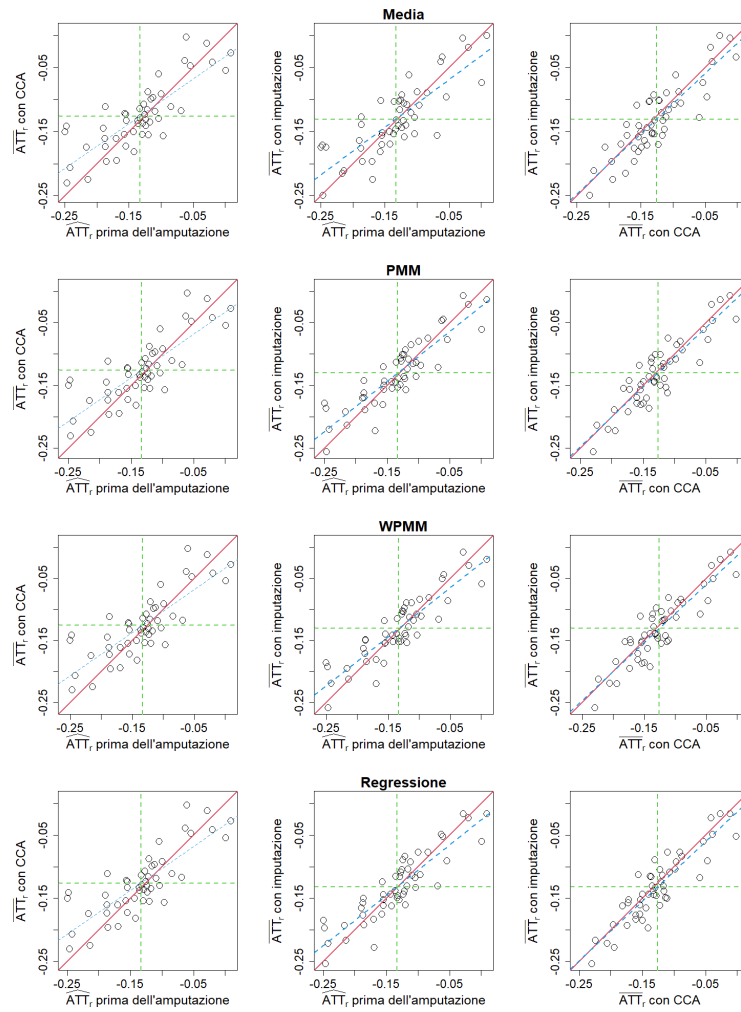


Figura A.73: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 5.

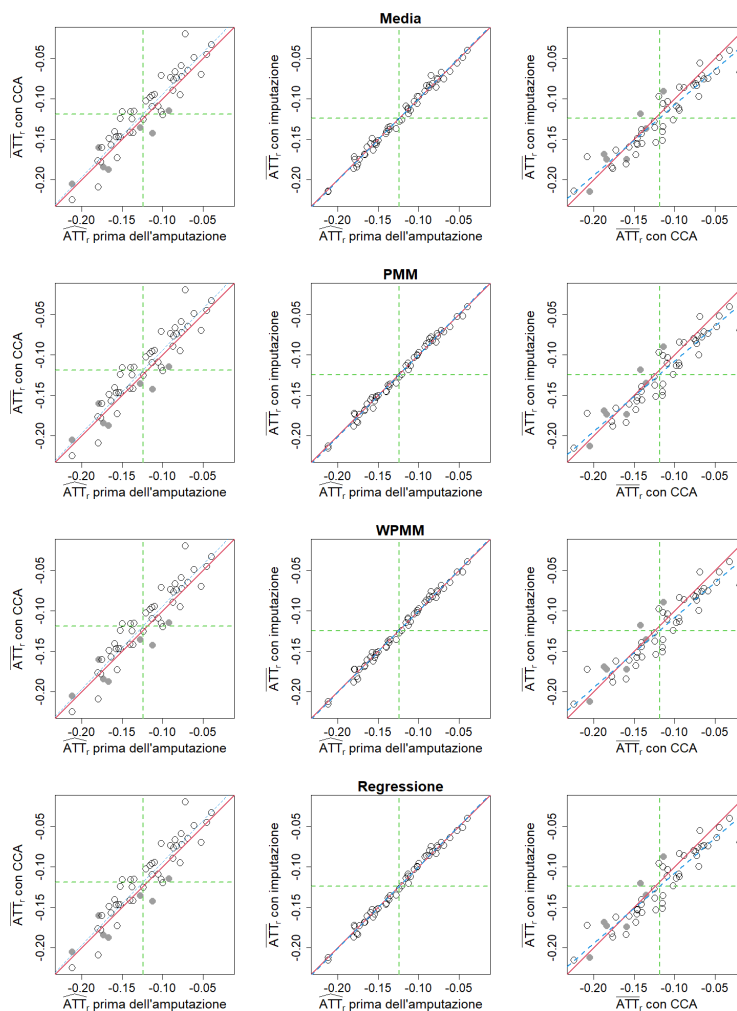


Figura A.74: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 5. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

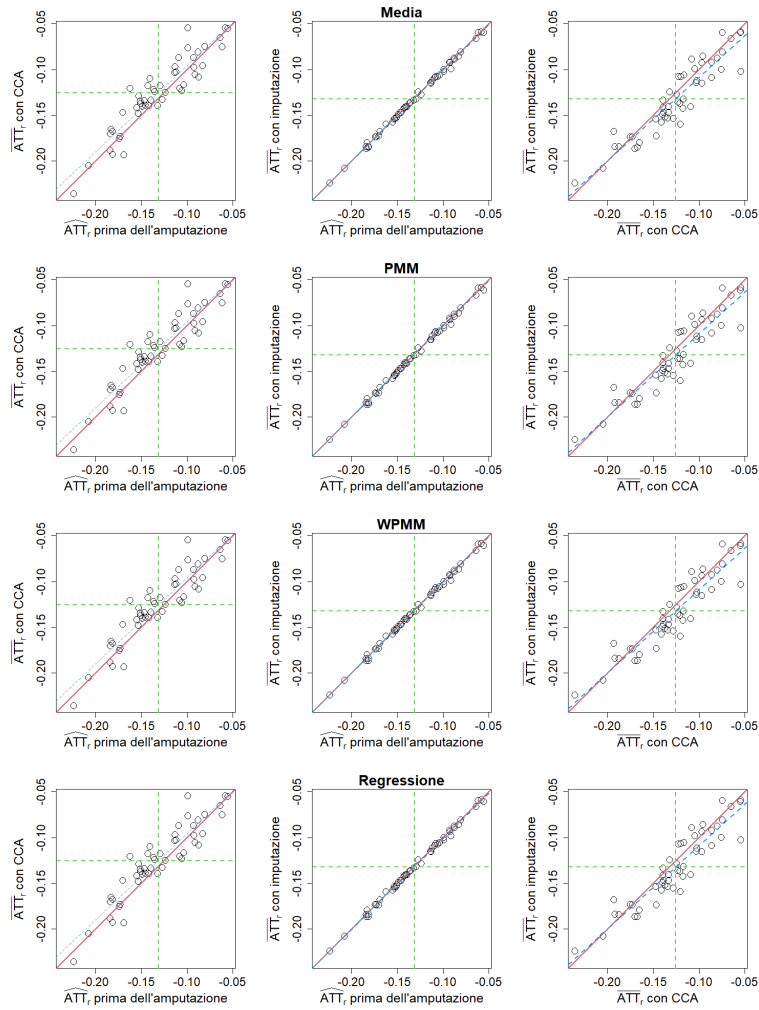


Figura A.75: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 5.

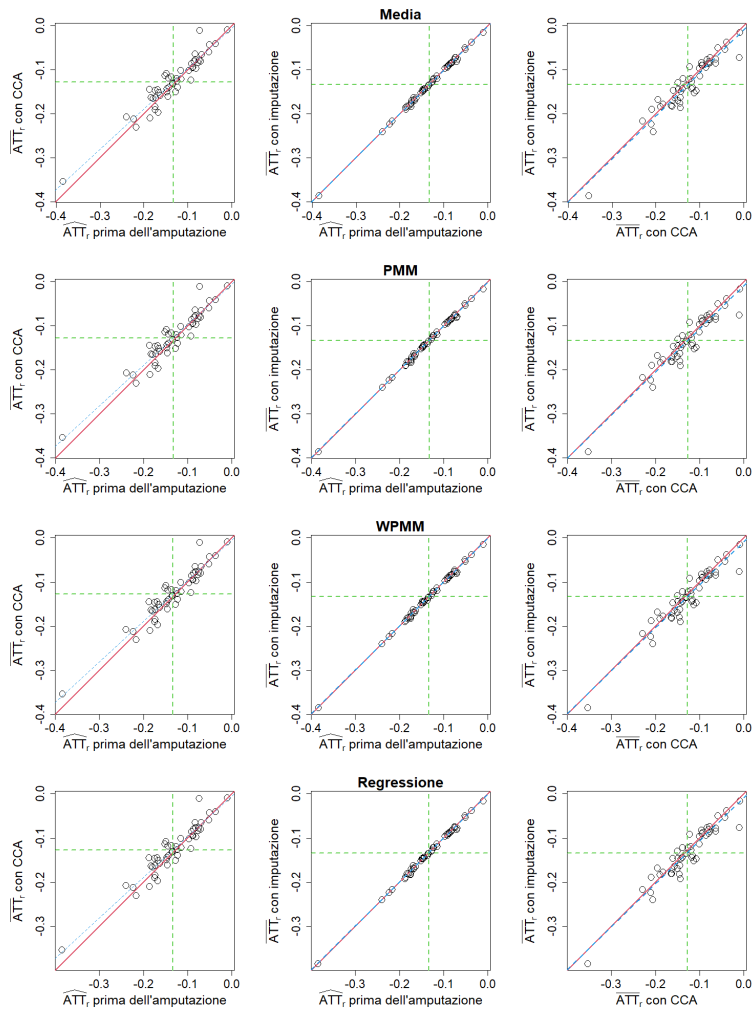


Figura A.76: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 5.

A.20 Simulazione 6 Scenario A

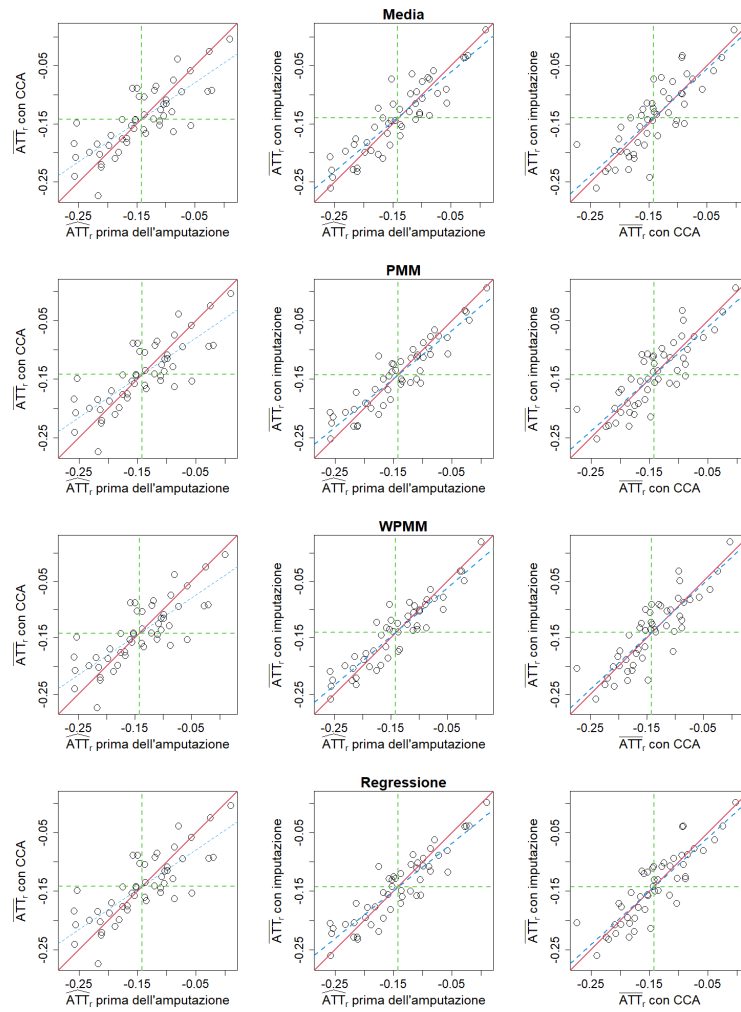


Figura A.77: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 6.

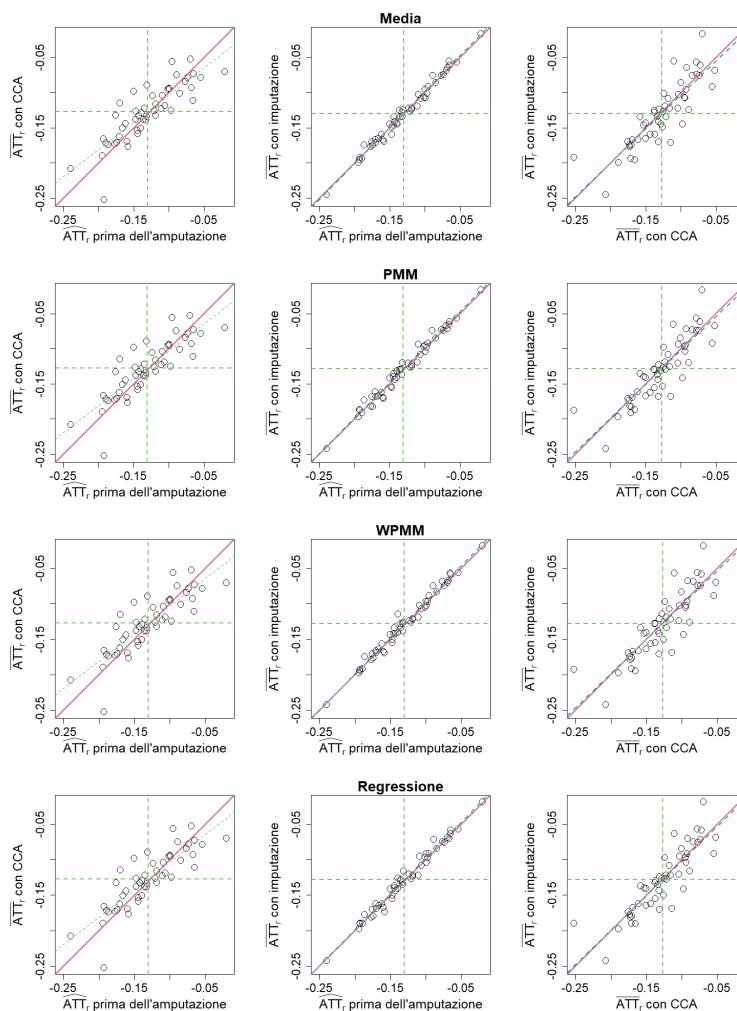


Figura A.78: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 6. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

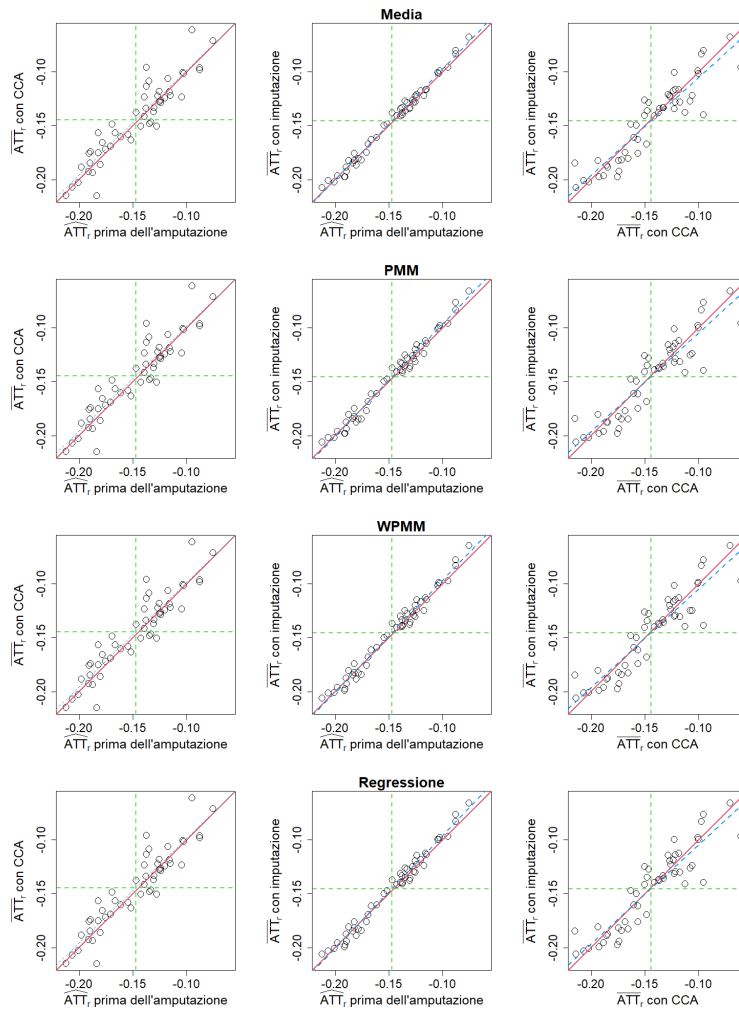


Figura A.79: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 6.

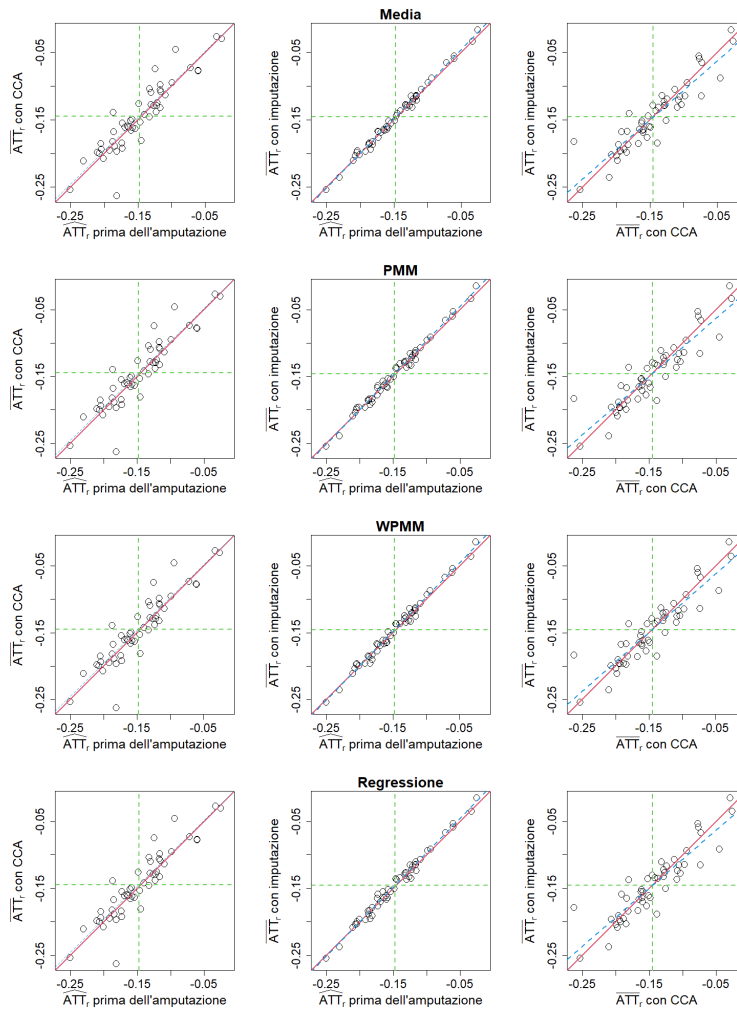


Figura A.80: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 6.

A.21 Simulazione 6 Scenario B

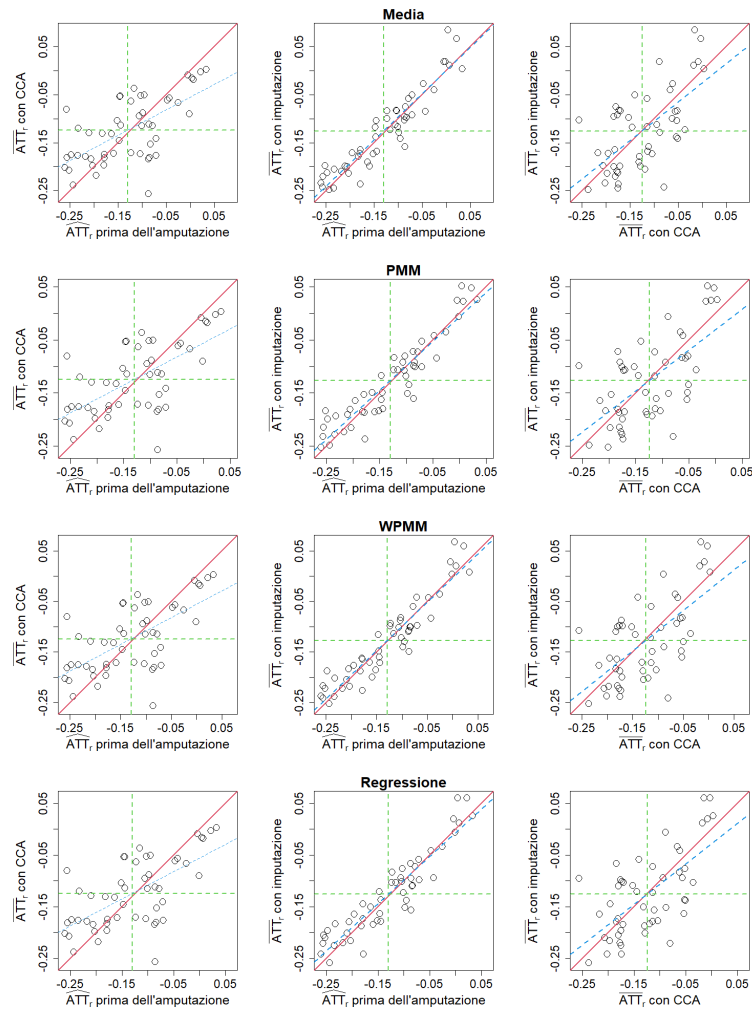


Figura A.81: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 6.

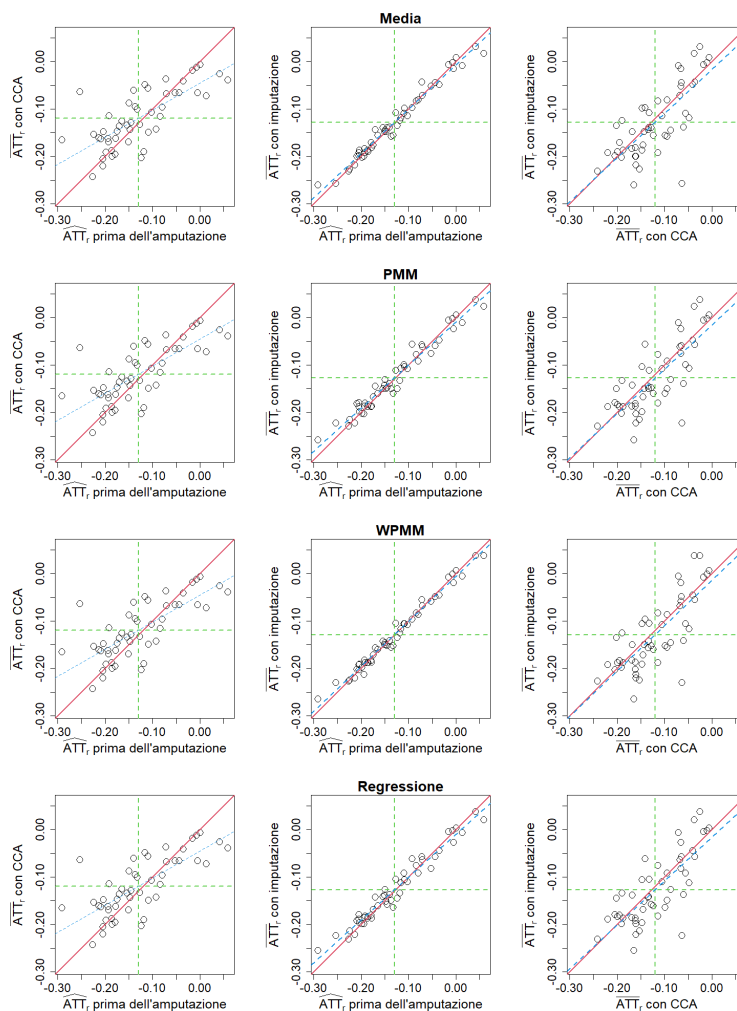


Figura A.82: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 6. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

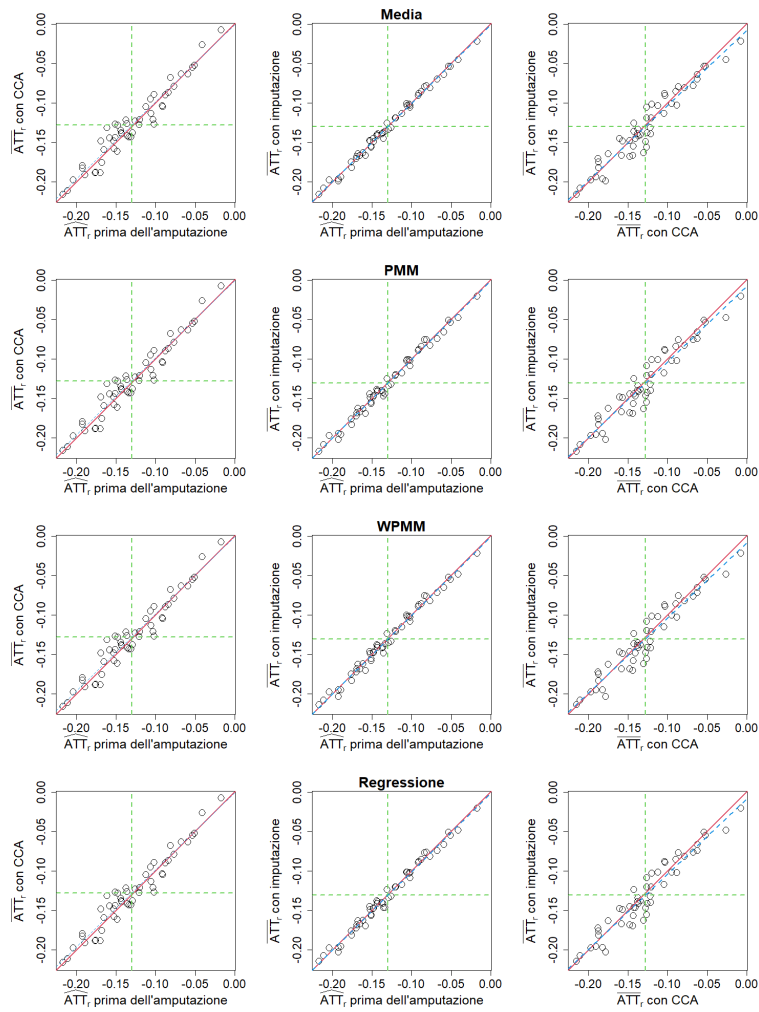


Figura A.83: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 6.

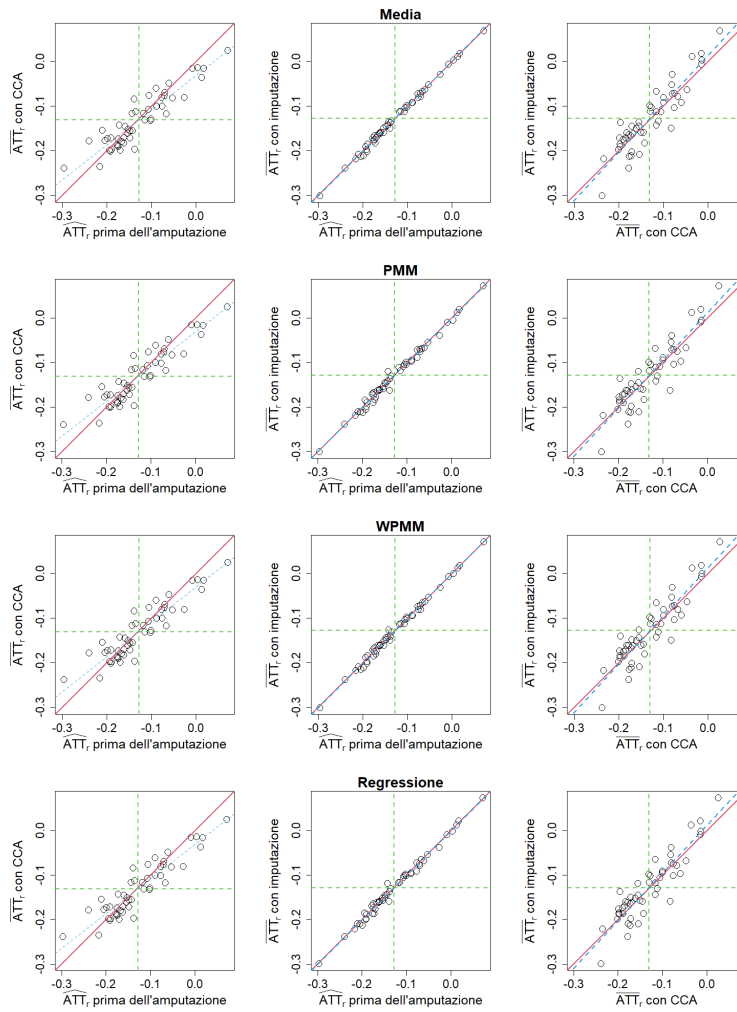


Figura A.84: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 6.

A.22 Simulazione 6 Scenario C

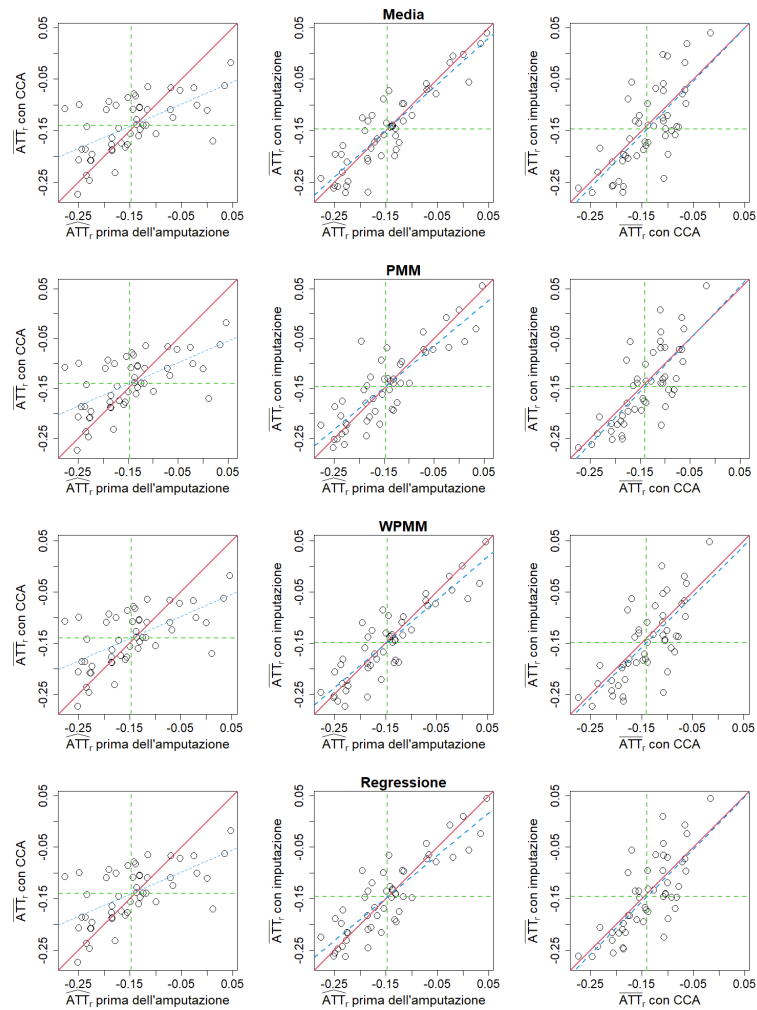


Figura A.85: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 6.

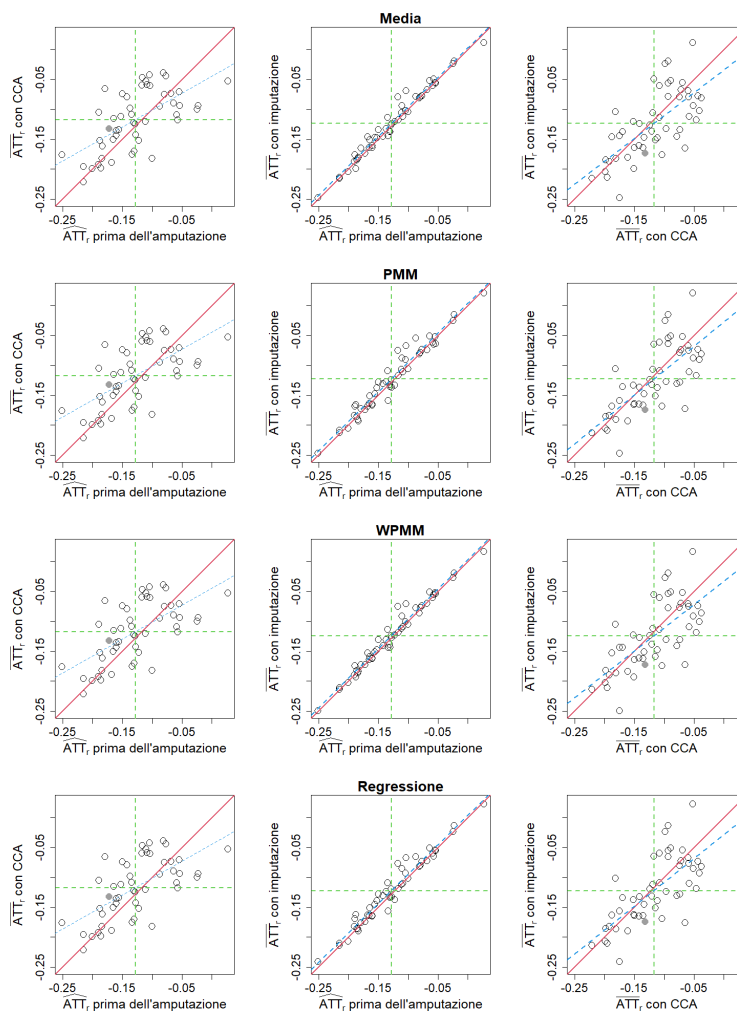


Figura A.86: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 6. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

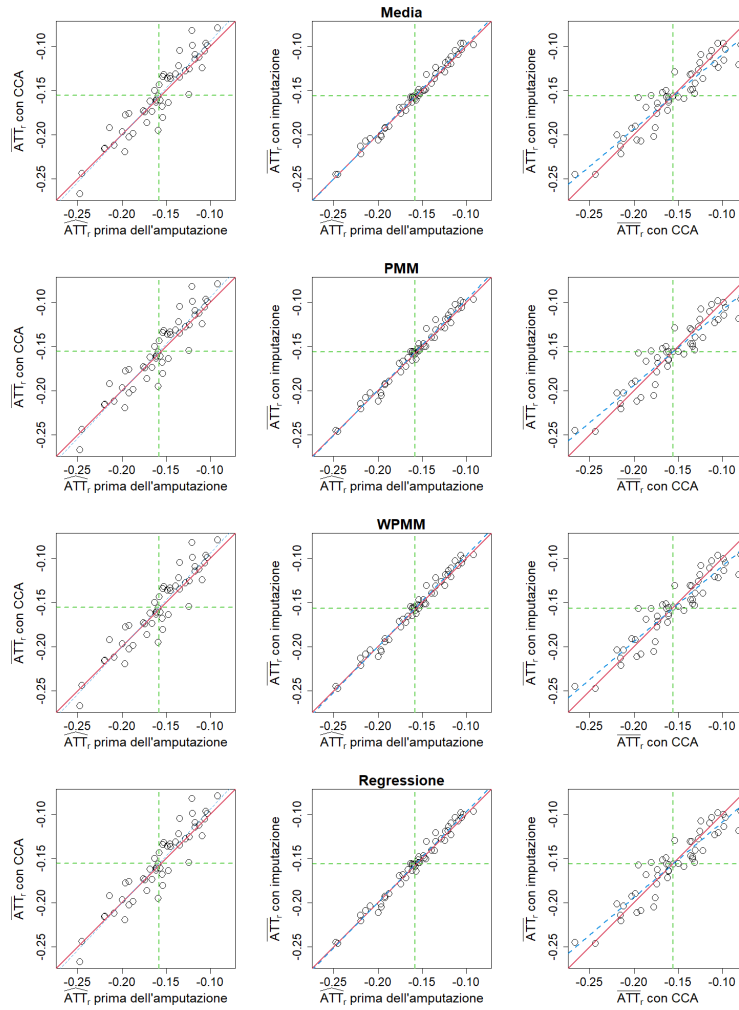


Figura A.87: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 6.

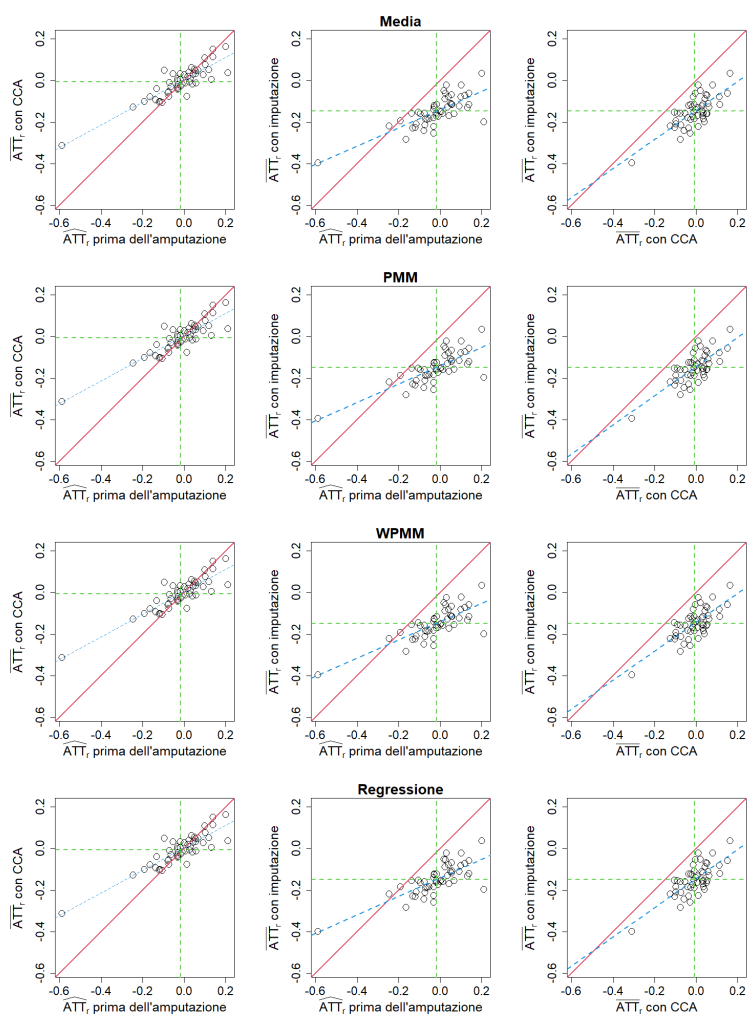


Figura A.88: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 6.

A.23 Simulazione 6 Scenario D

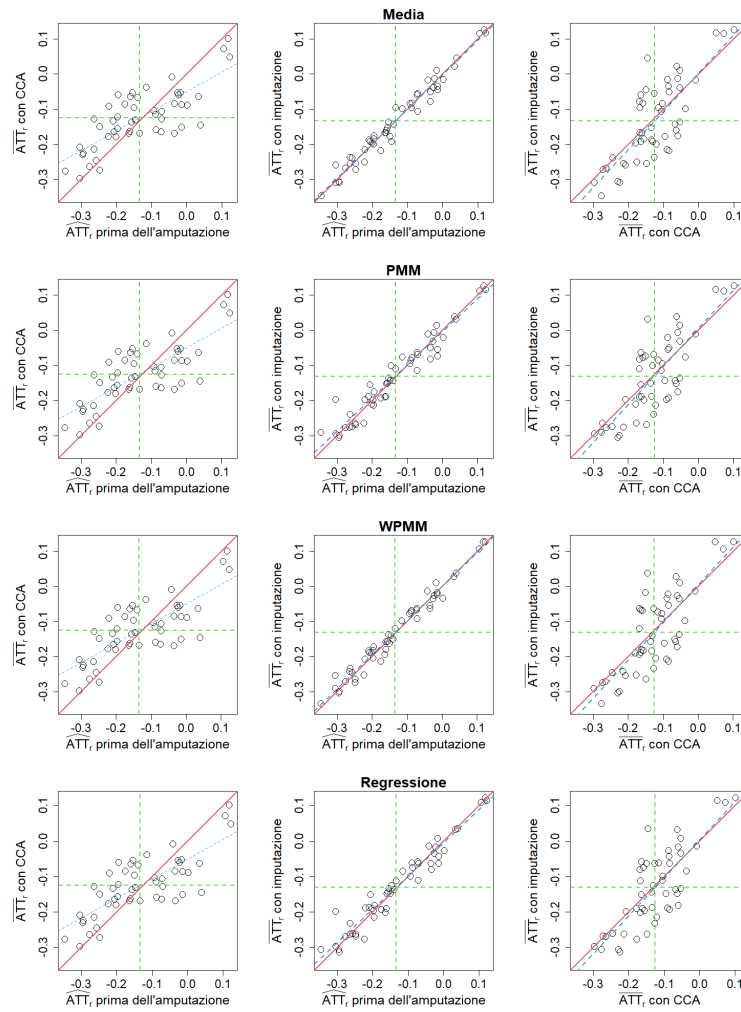


Figura A.89: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 6.

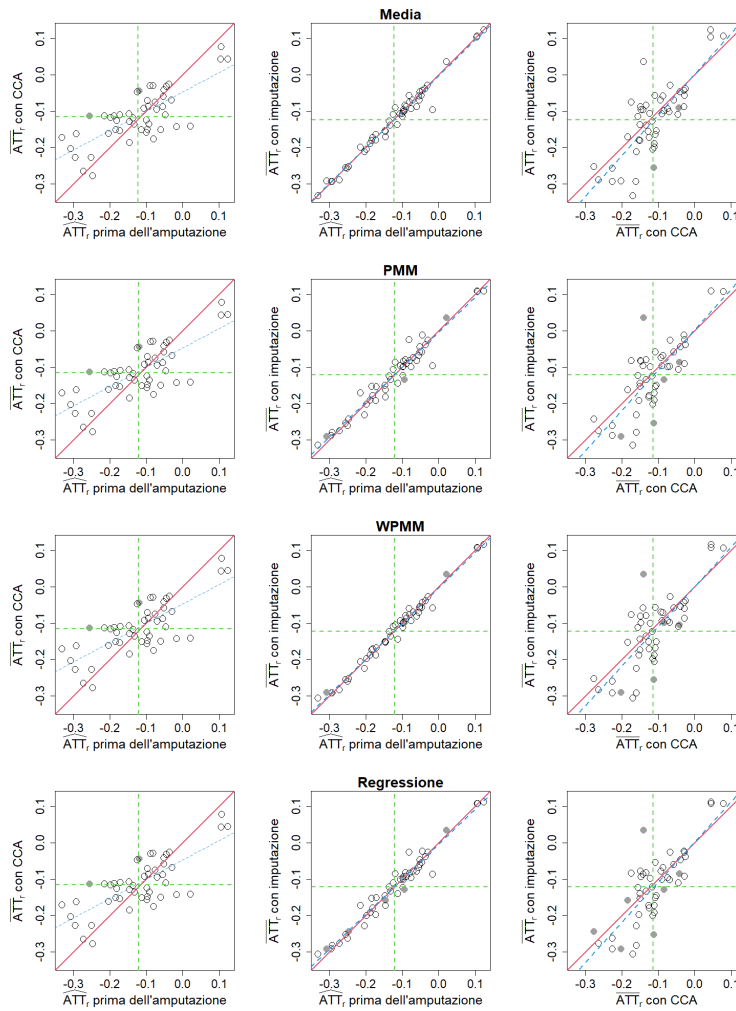


Figura A.90: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 6. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

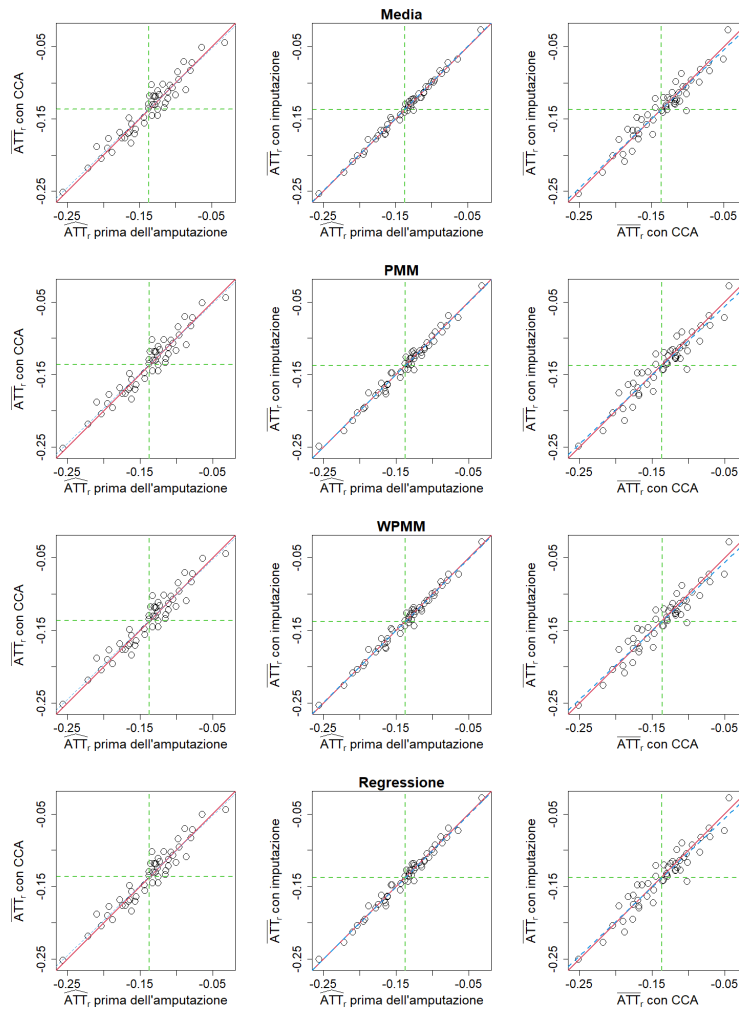


Figura A.91: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 6.

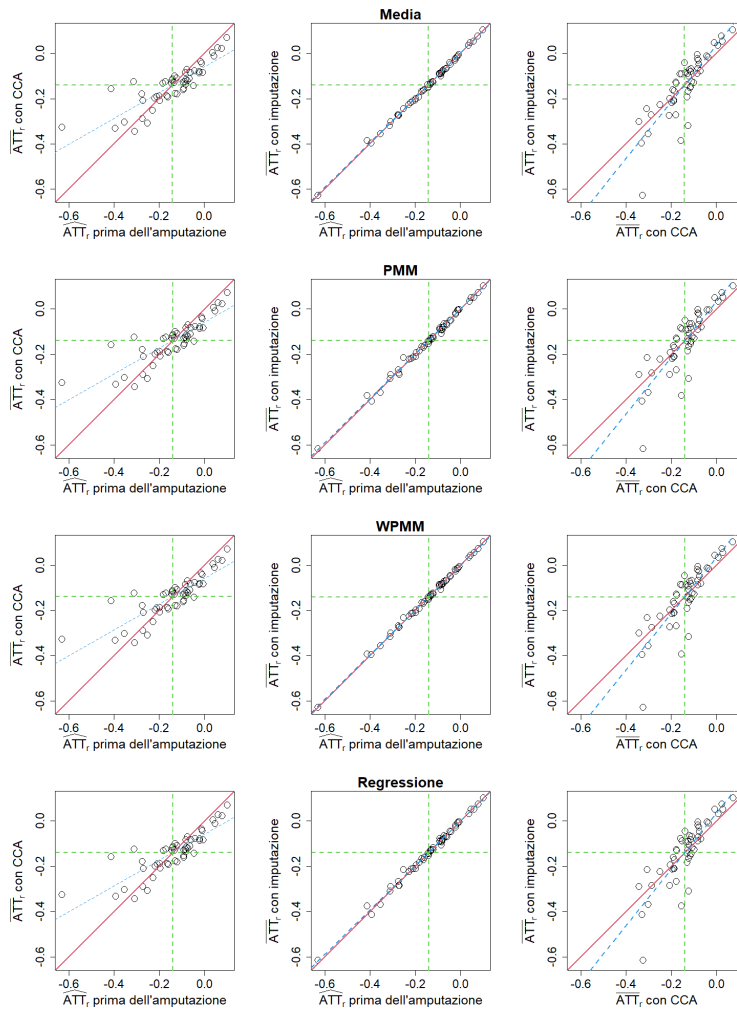


Figura A.92: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 6.

A.24 Simulazione 7 Scenario A

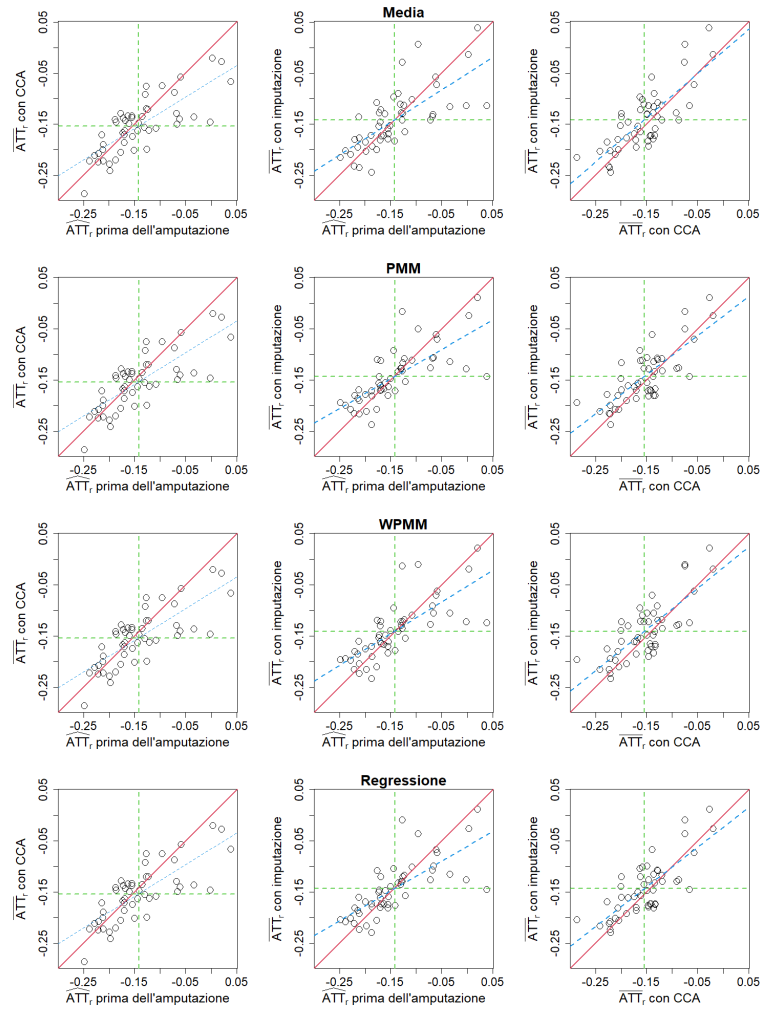


Figura A.93: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 7.

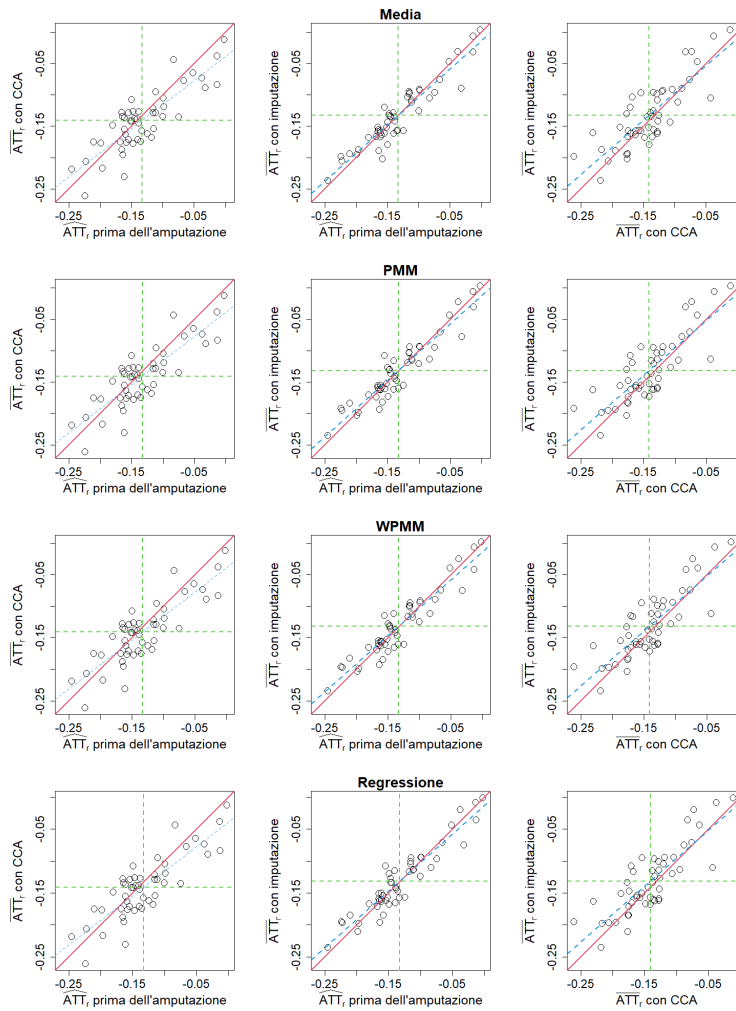


Figura A.94: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 7. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

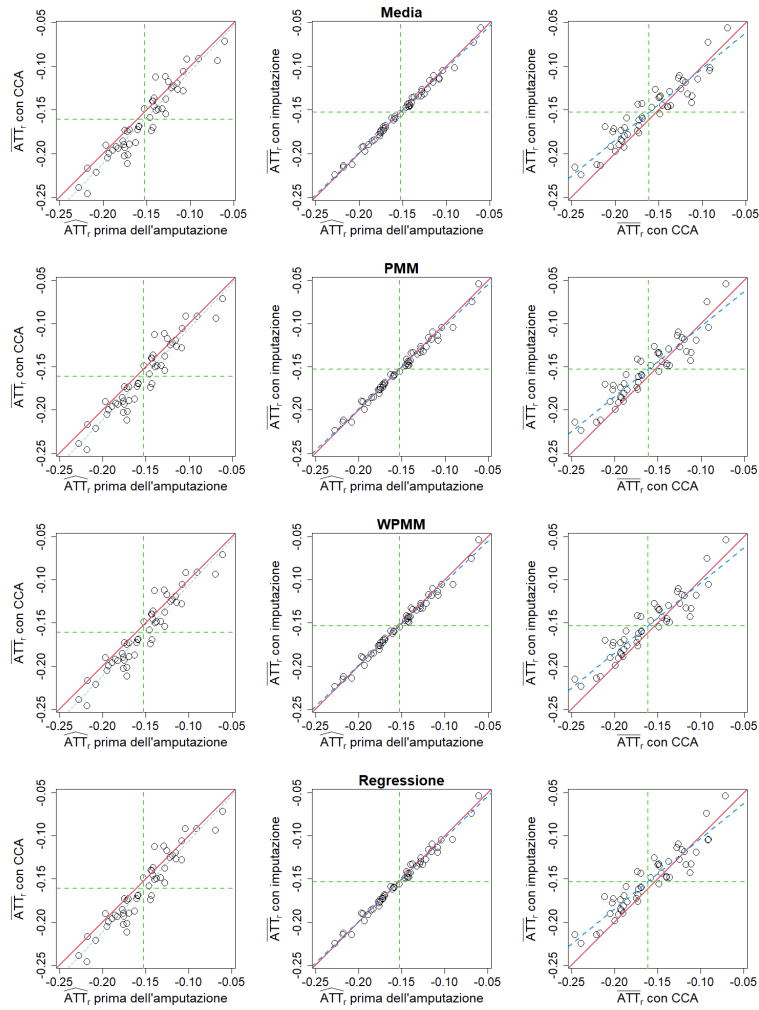


Figura A.95: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 7.

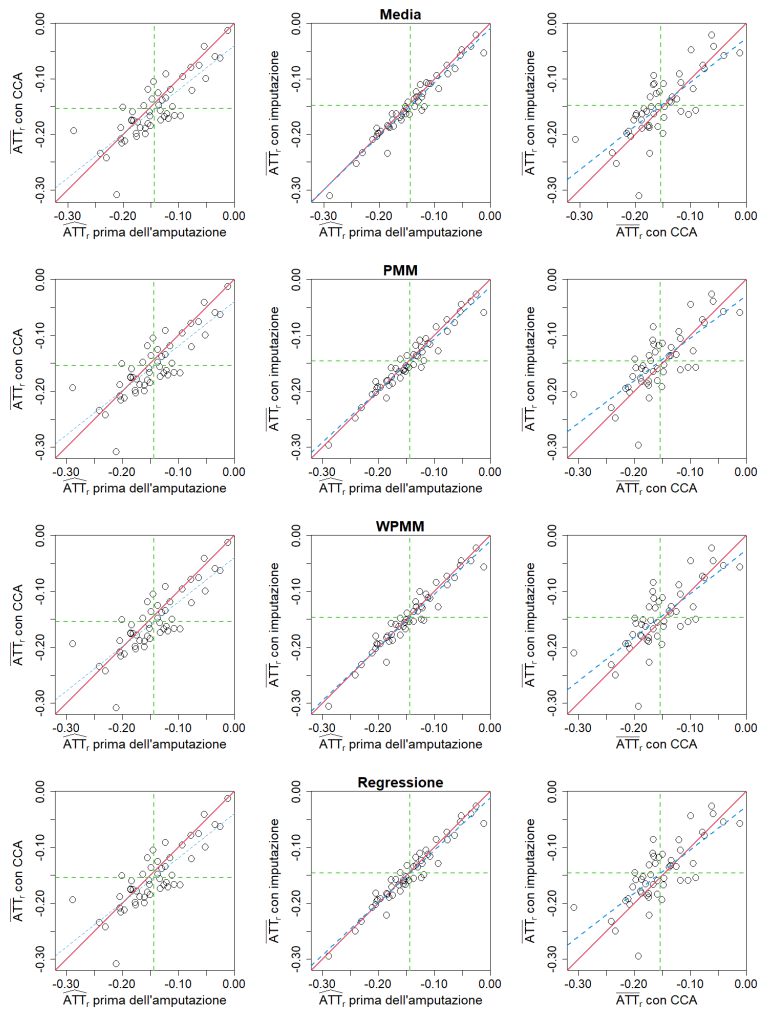


Figura A.96: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 7.

A.25 Simulazione 7 Scenario B

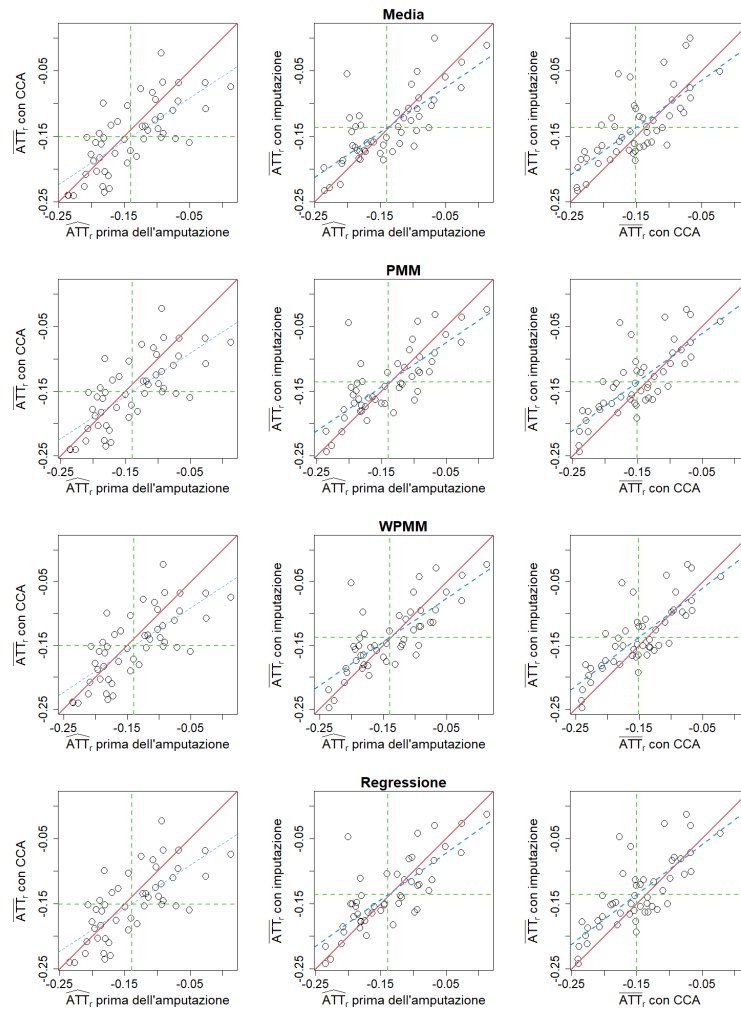


Figura A.97: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 7.

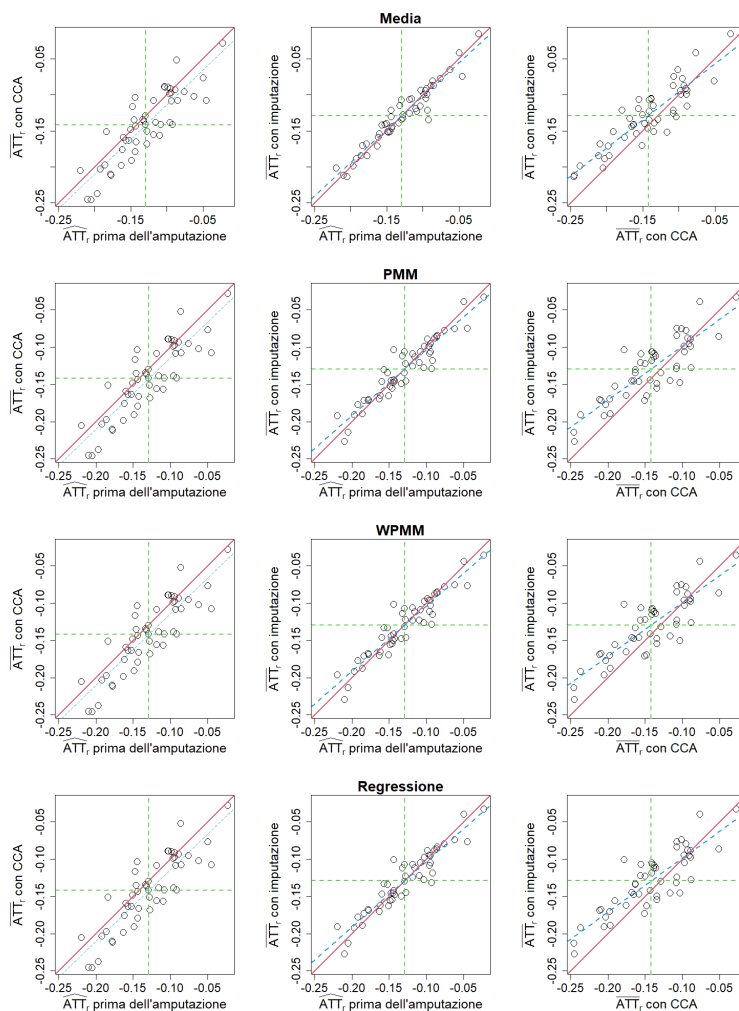


Figura A.98: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 7. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

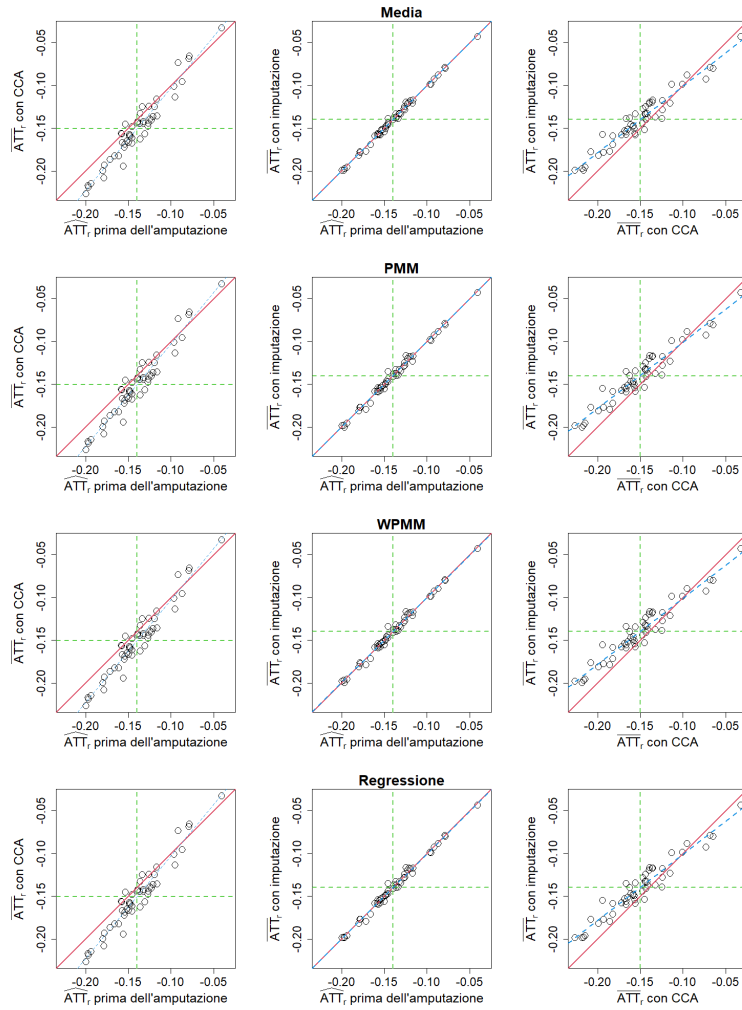


Figura A.99: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 7.

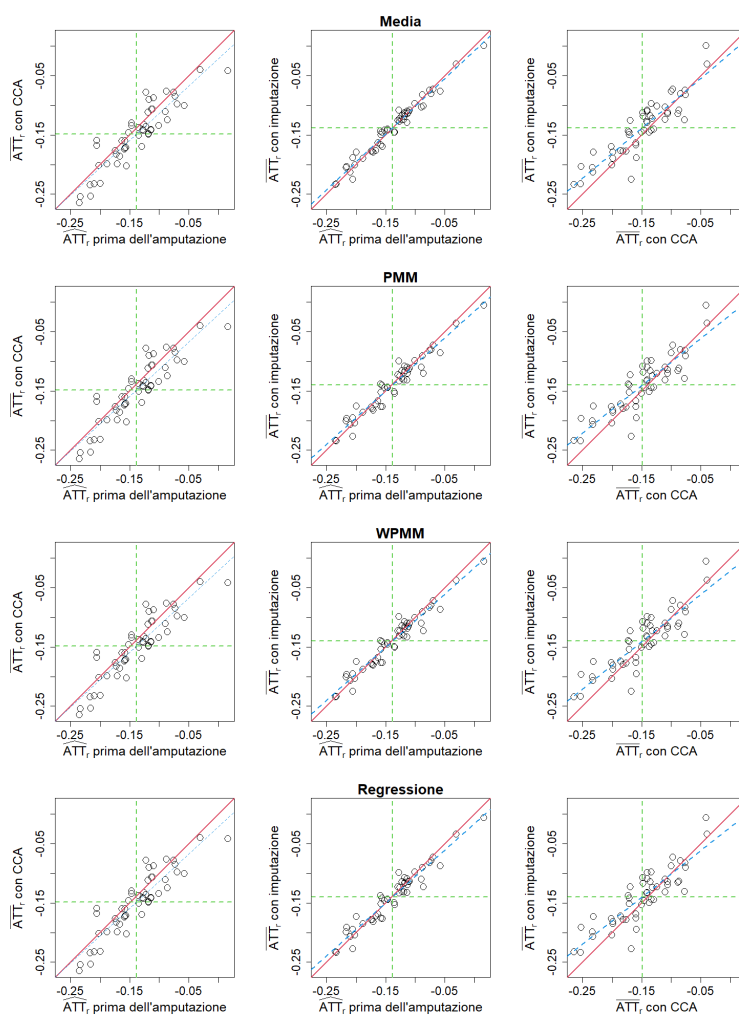


Figura A.100: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 7.

A.26 Simulazione 7 Scenario C

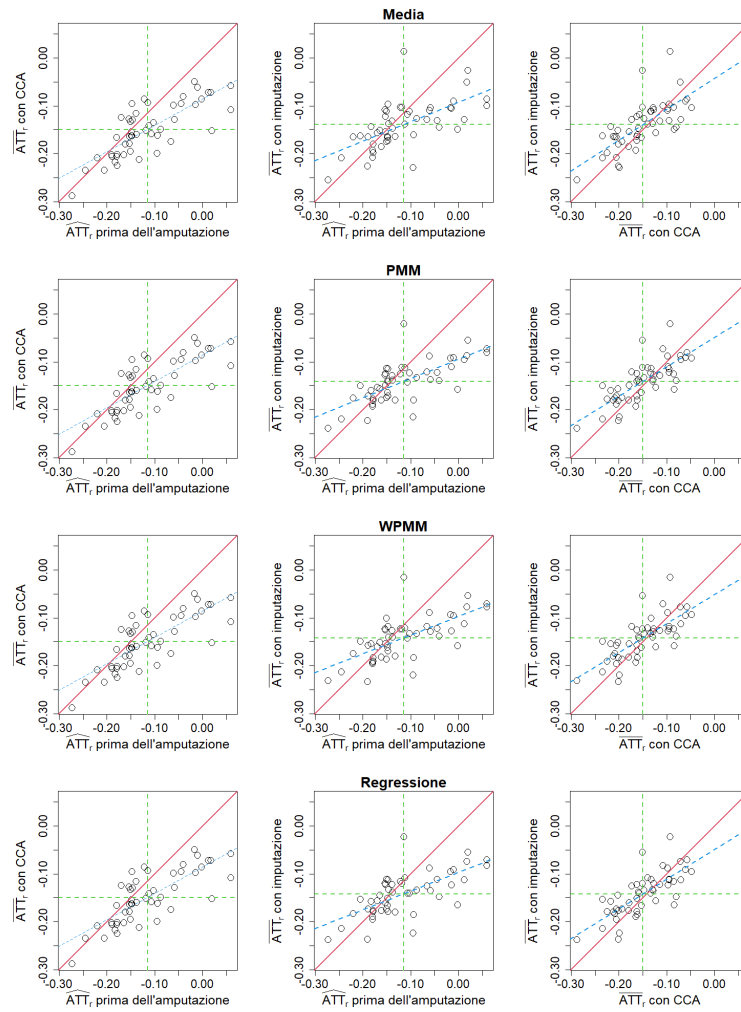


Figura A.101: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 7.

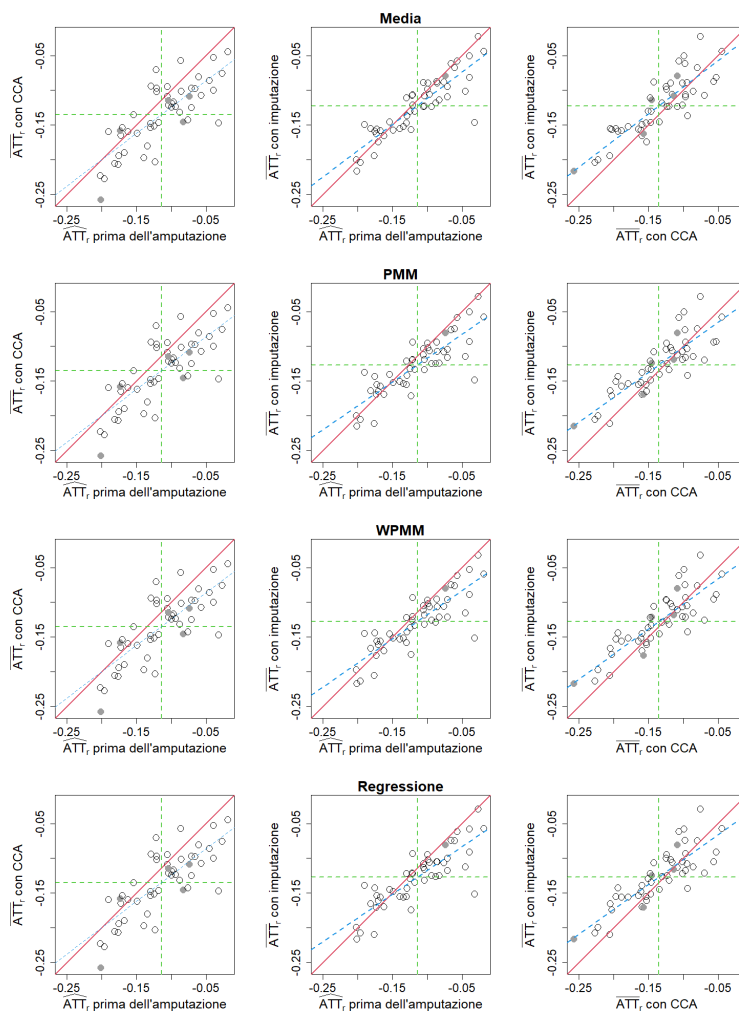


Figura A.102: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 7. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

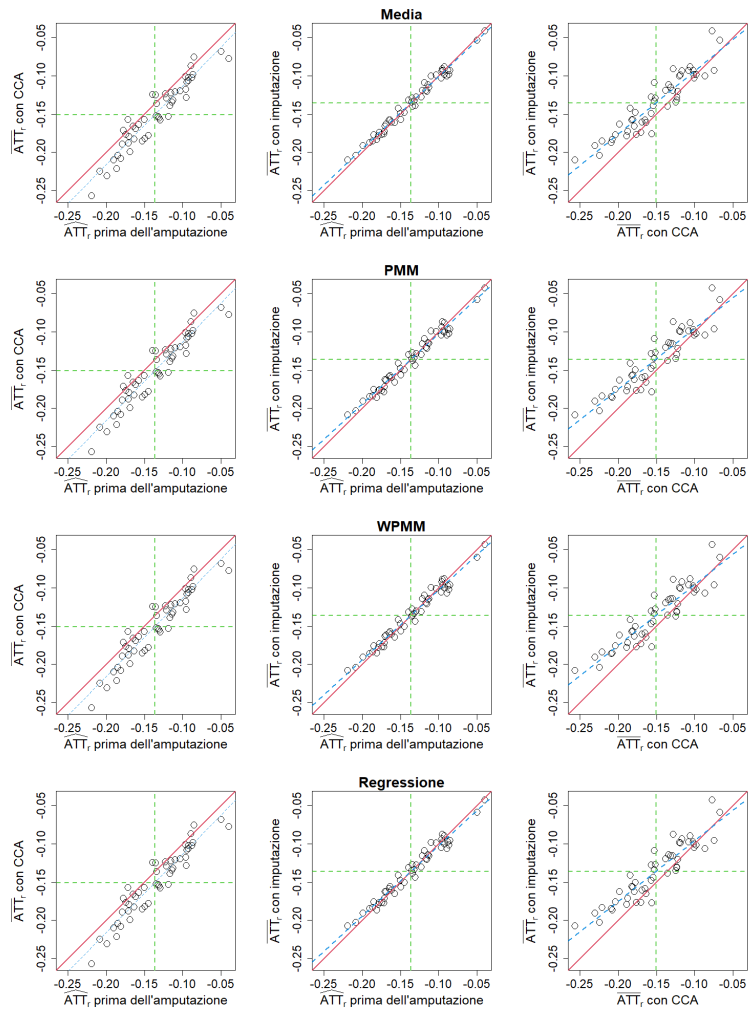


Figura A.103: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 7.

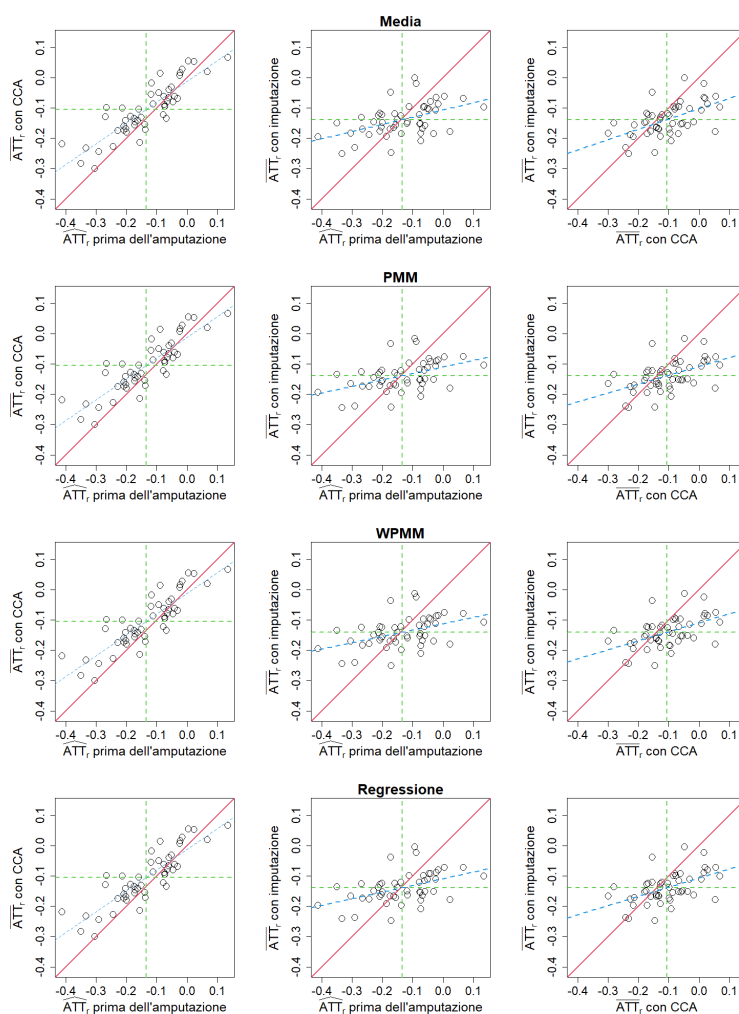


Figura A.104: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 7.

A.27 Simulazione 7 Scenario D

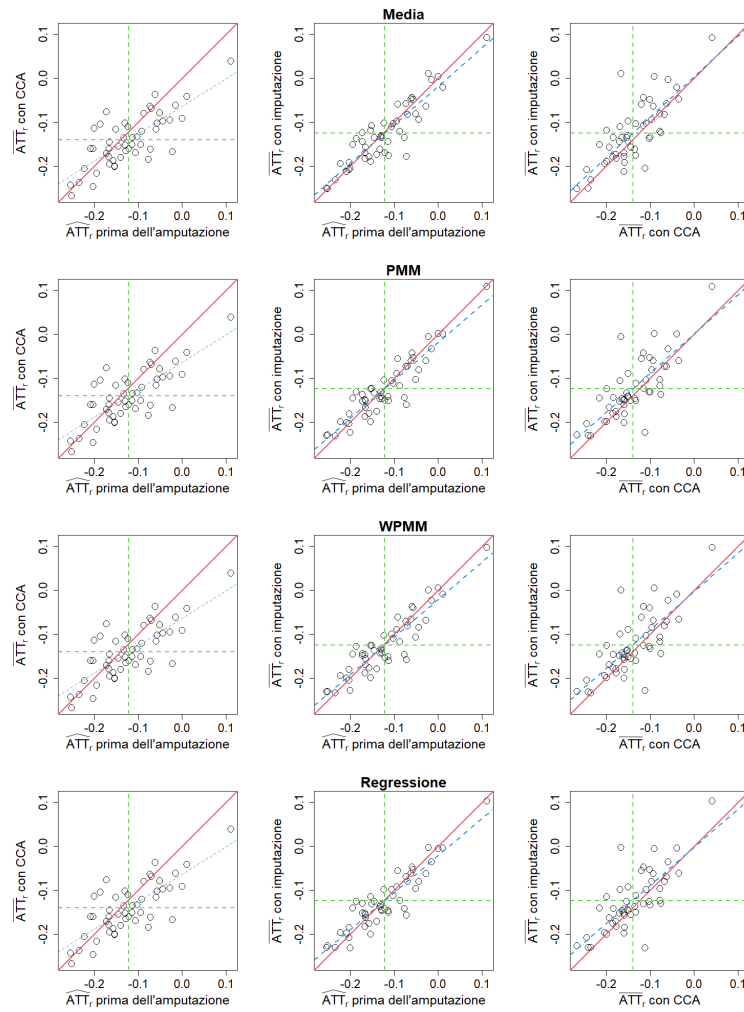


Figura A.105: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 7.

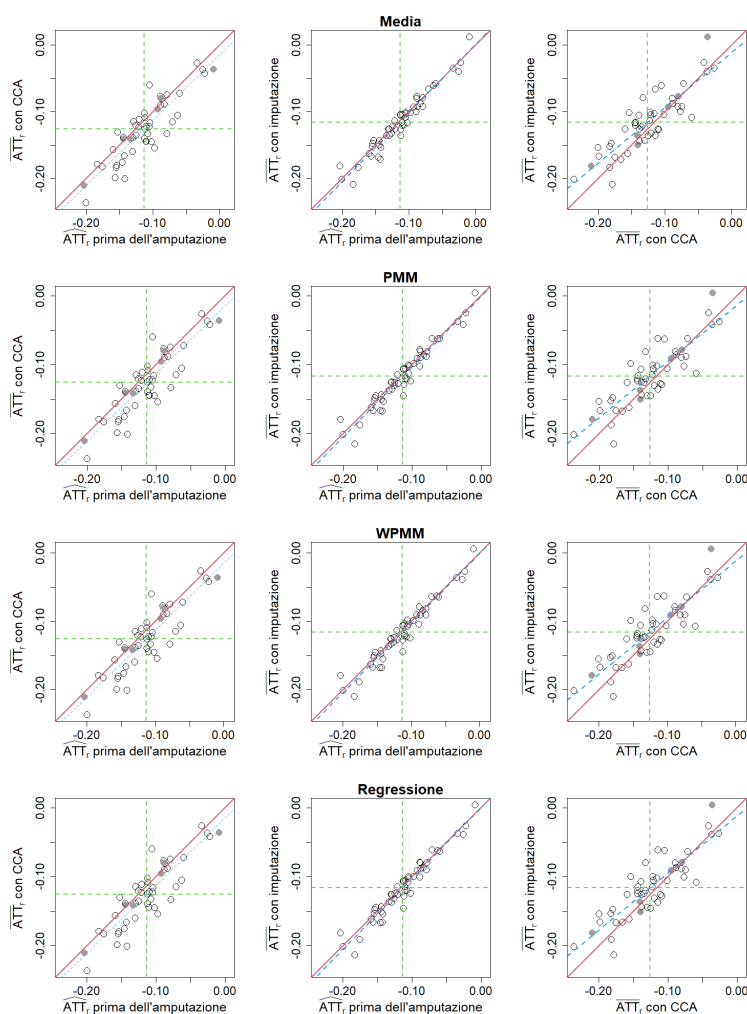


Figura A.106: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 7. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

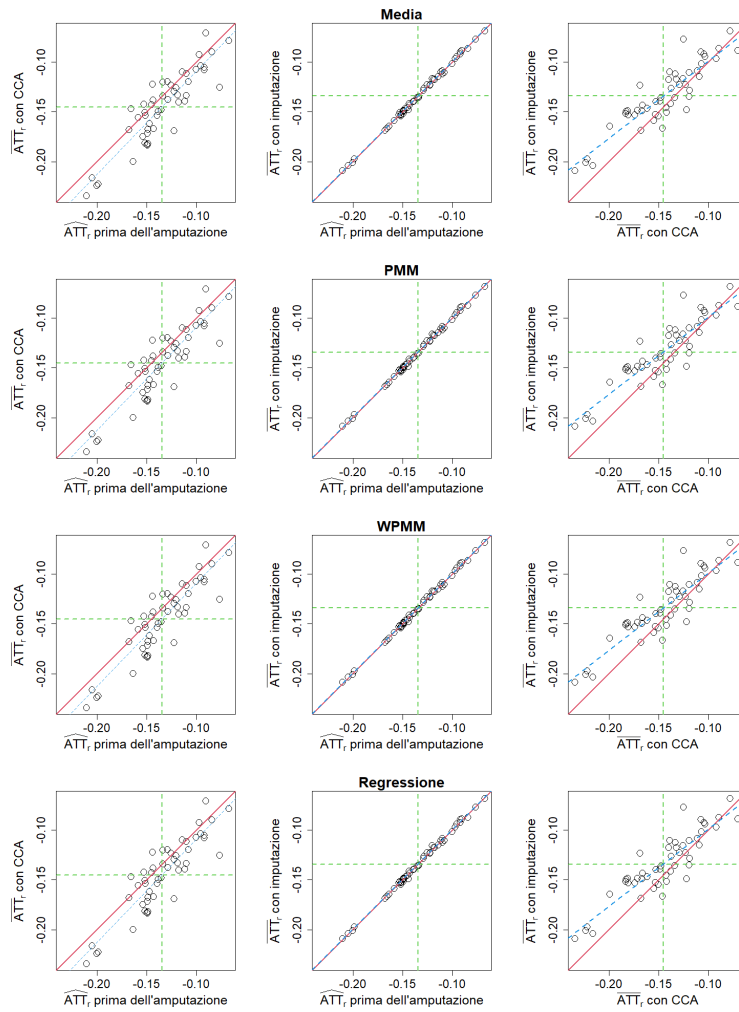


Figura A.107: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 7.

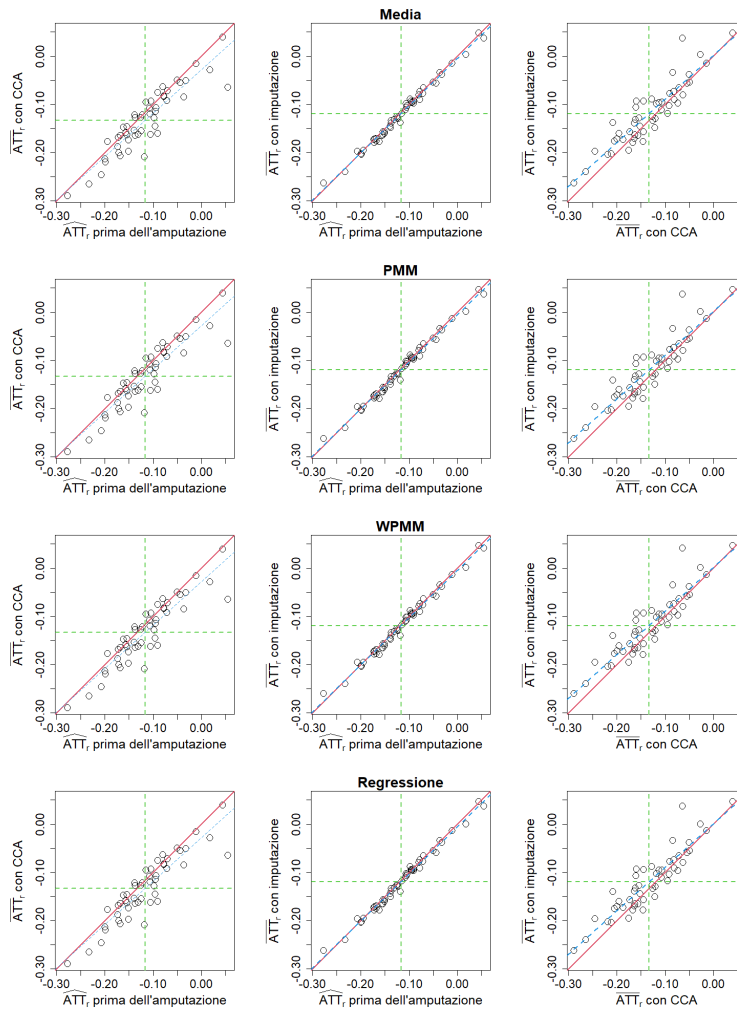


Figura A.108: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 7.

A.28 Simulazione 8 Scenario A

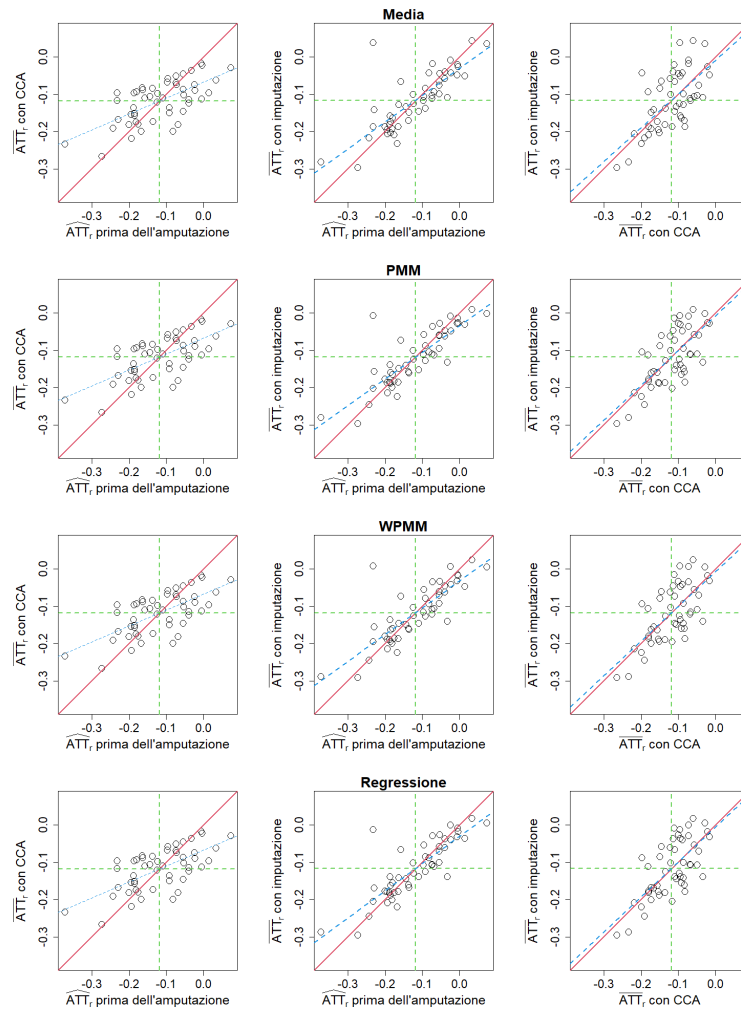


Figura A.109: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 8.

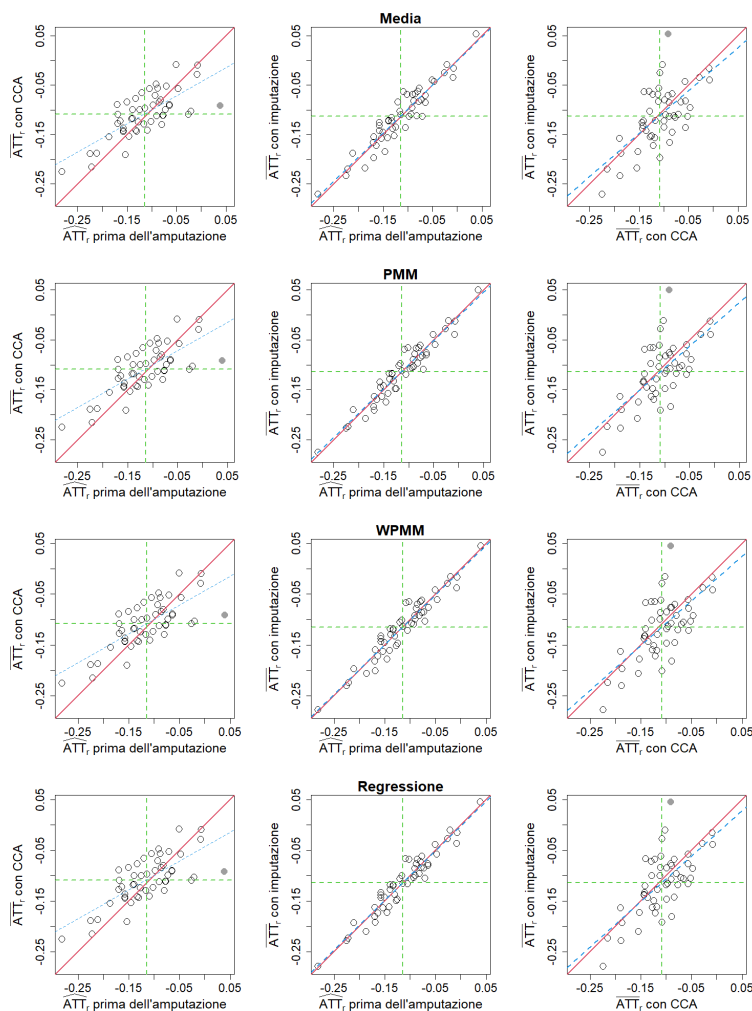


Figura A.110: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 8. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

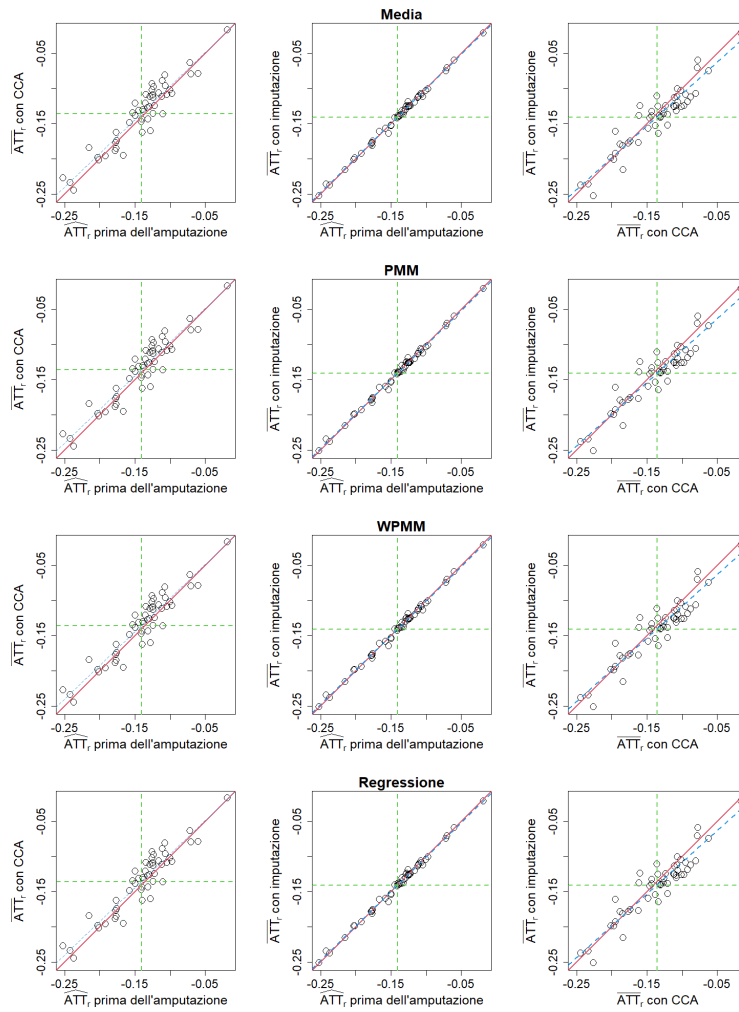


Figura A.111: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 8.

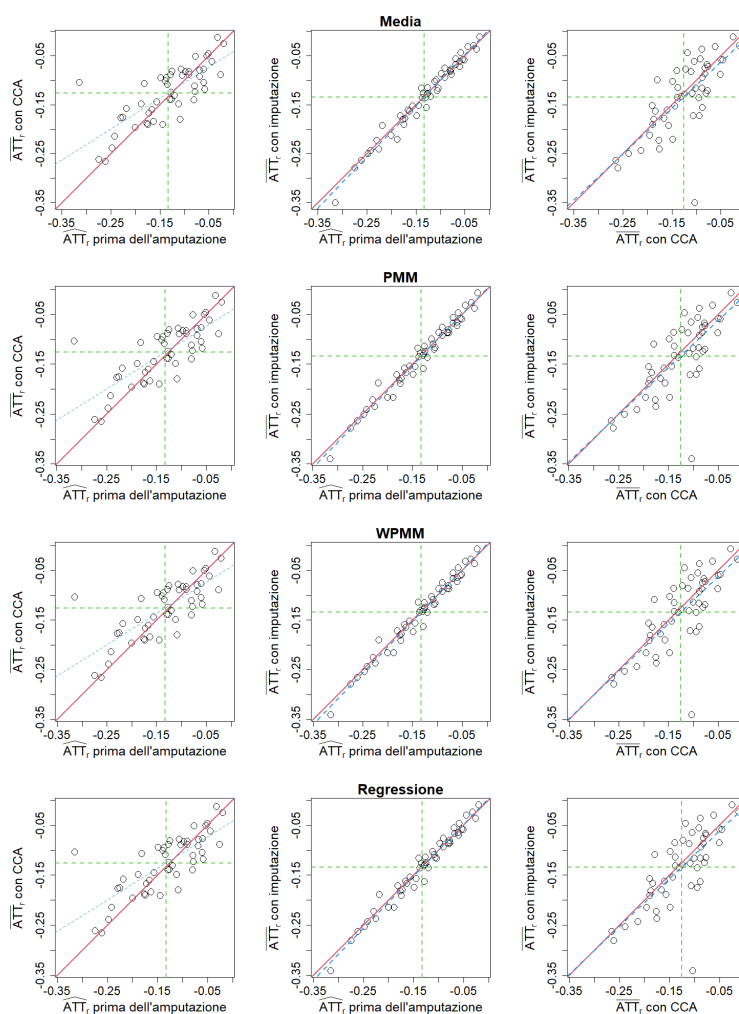


Figura A.112: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 8.

A.29 Simulazione 8 Scenario B

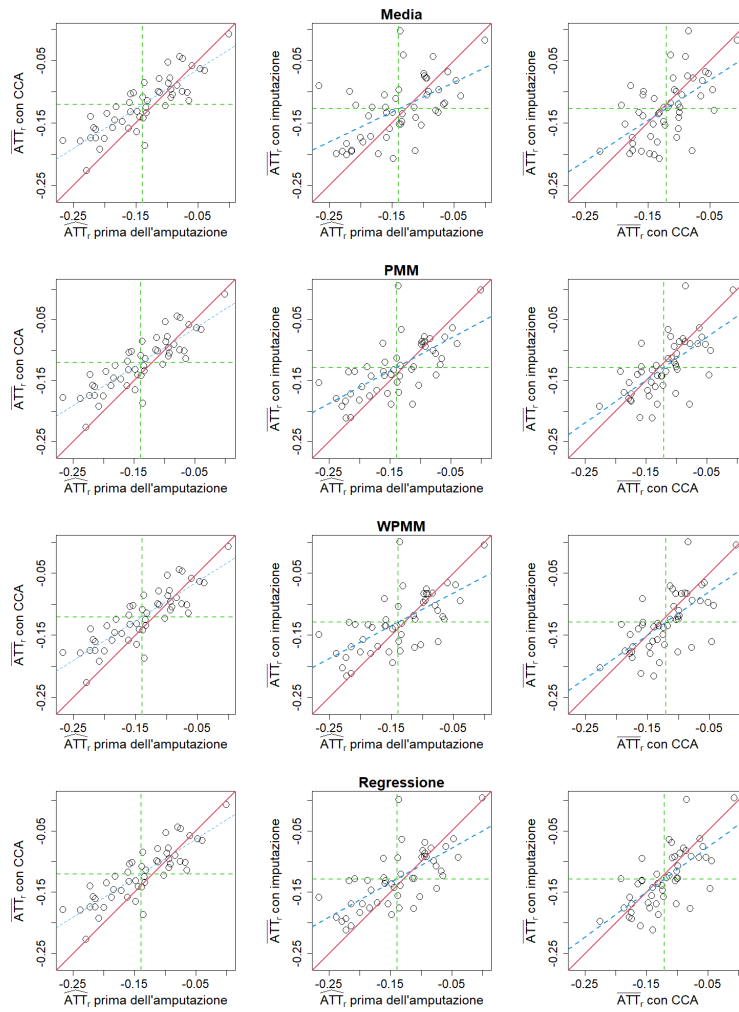


Figura A.113: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 8.

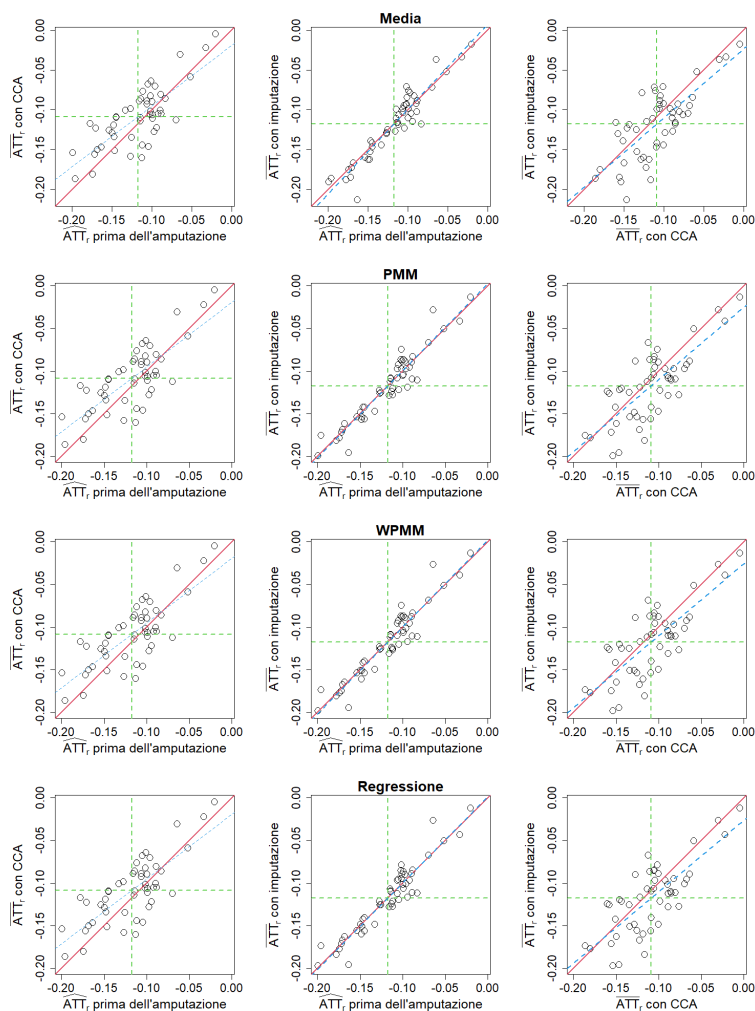


Figura A.114: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 8. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

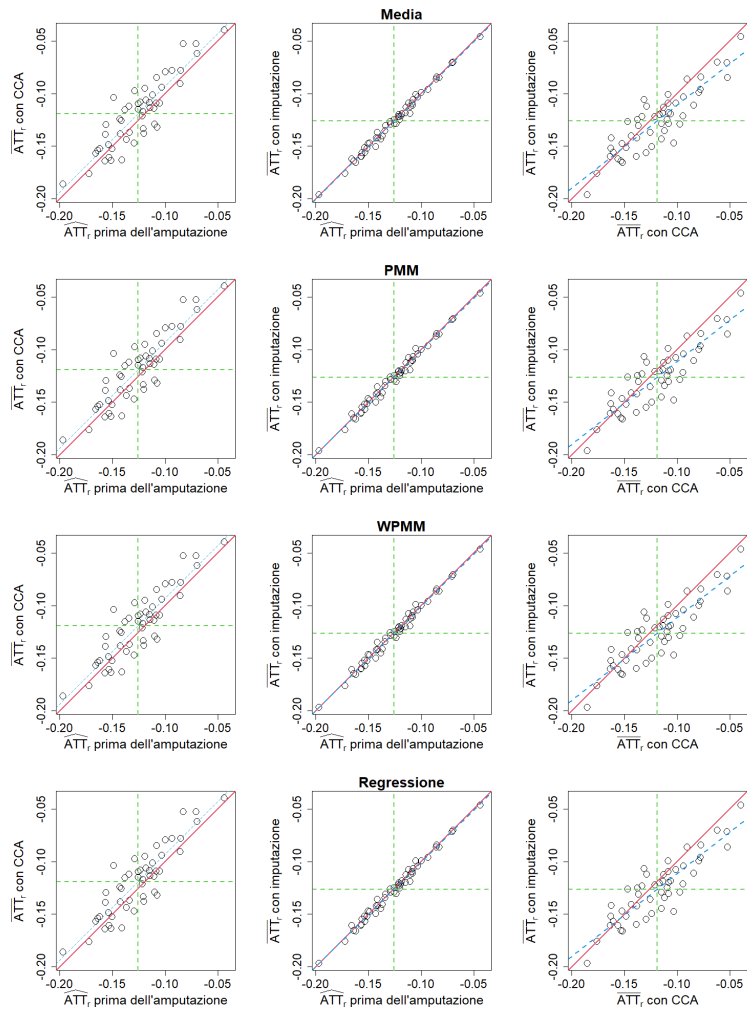


Figura A.115: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 8.

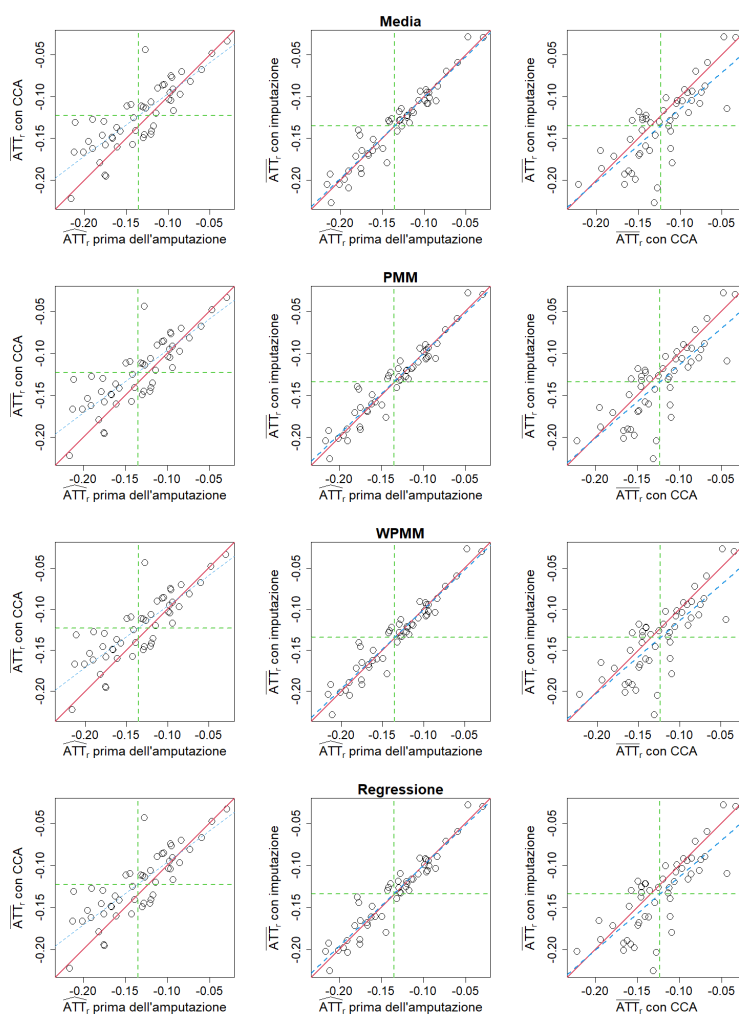


Figura A.116: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 8.

A.30 Simulazione 8 Scenario C

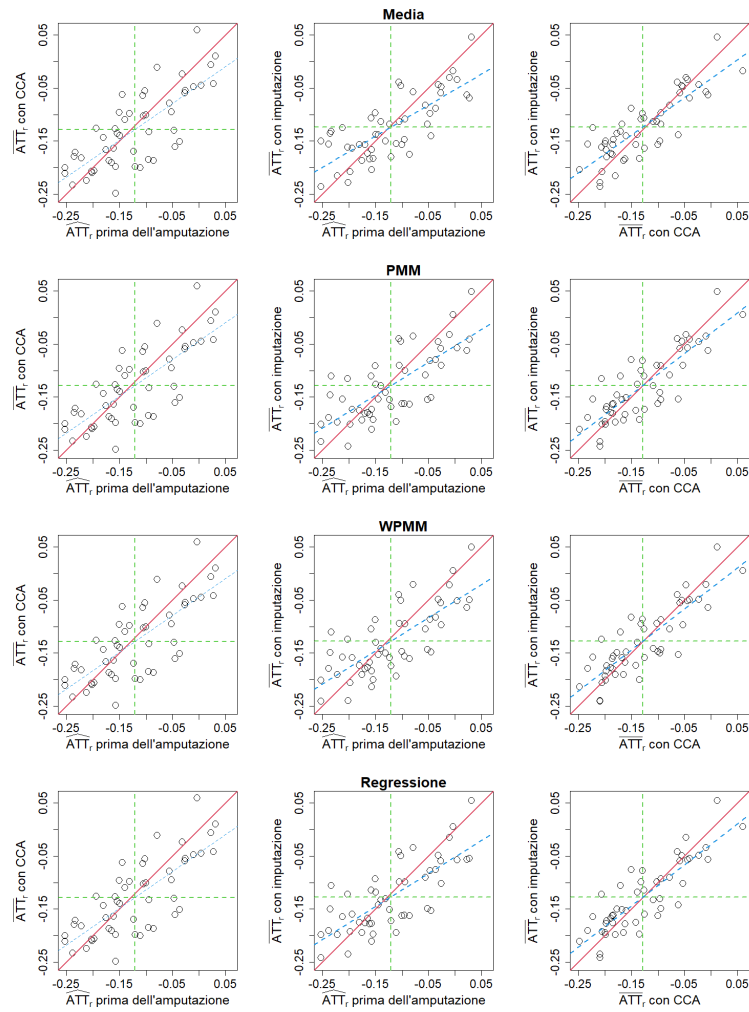


Figura A.117: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 8.

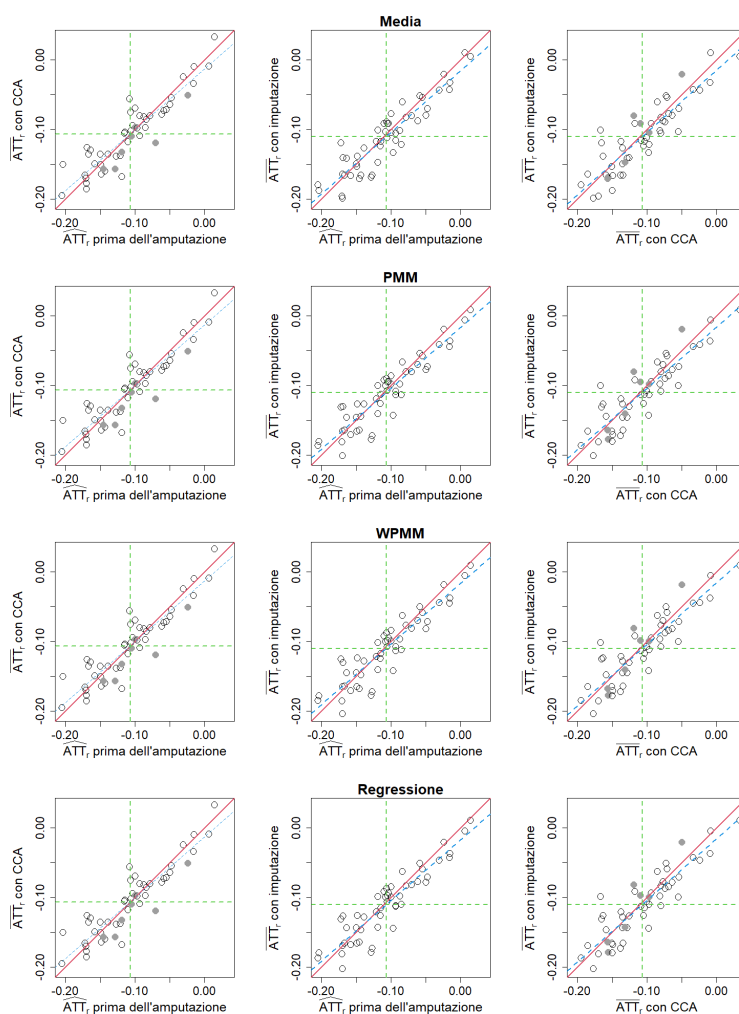


Figura A.118: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 8. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

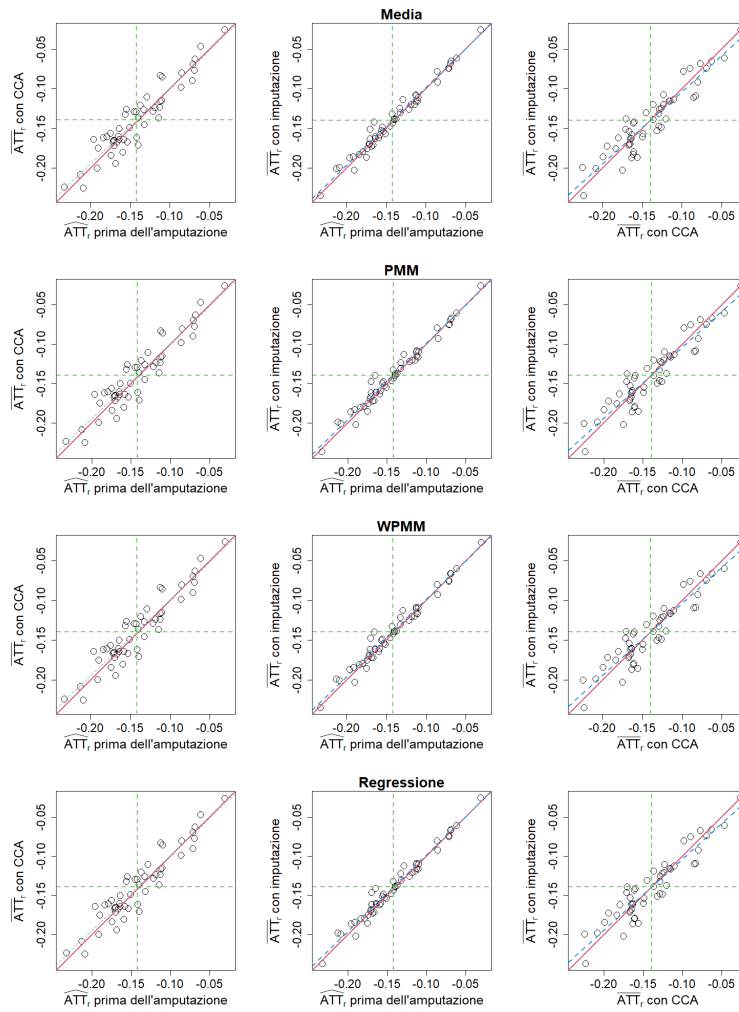


Figura A.119: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 8.

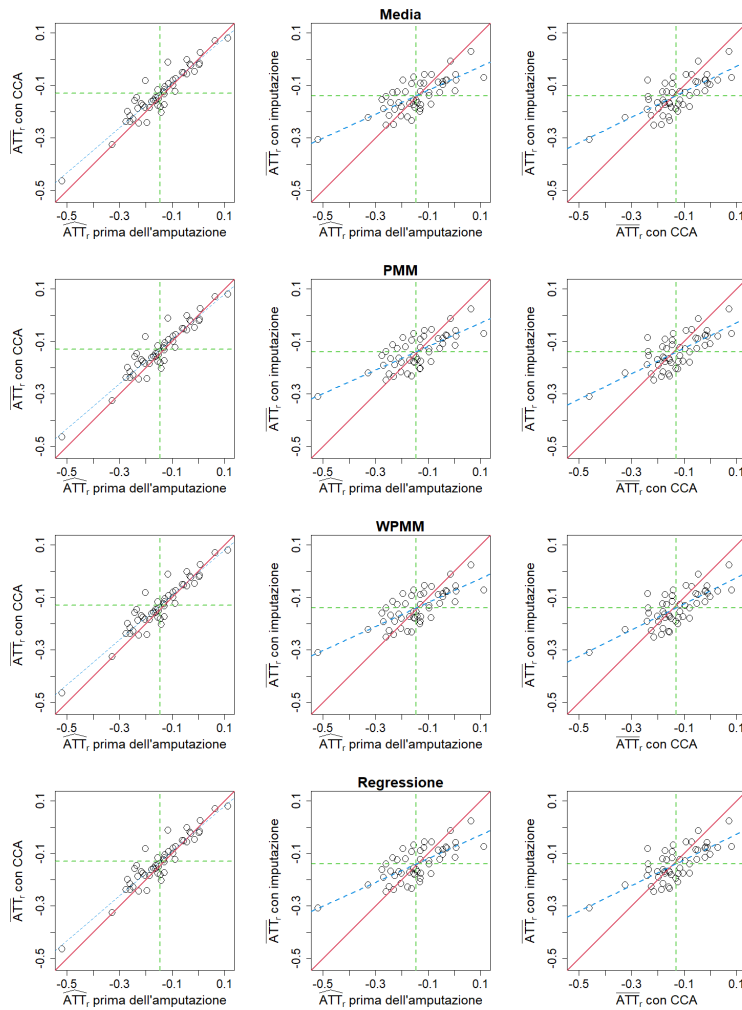


Figura A.120: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 8.

A.31 Simulazione 8 Scenario D

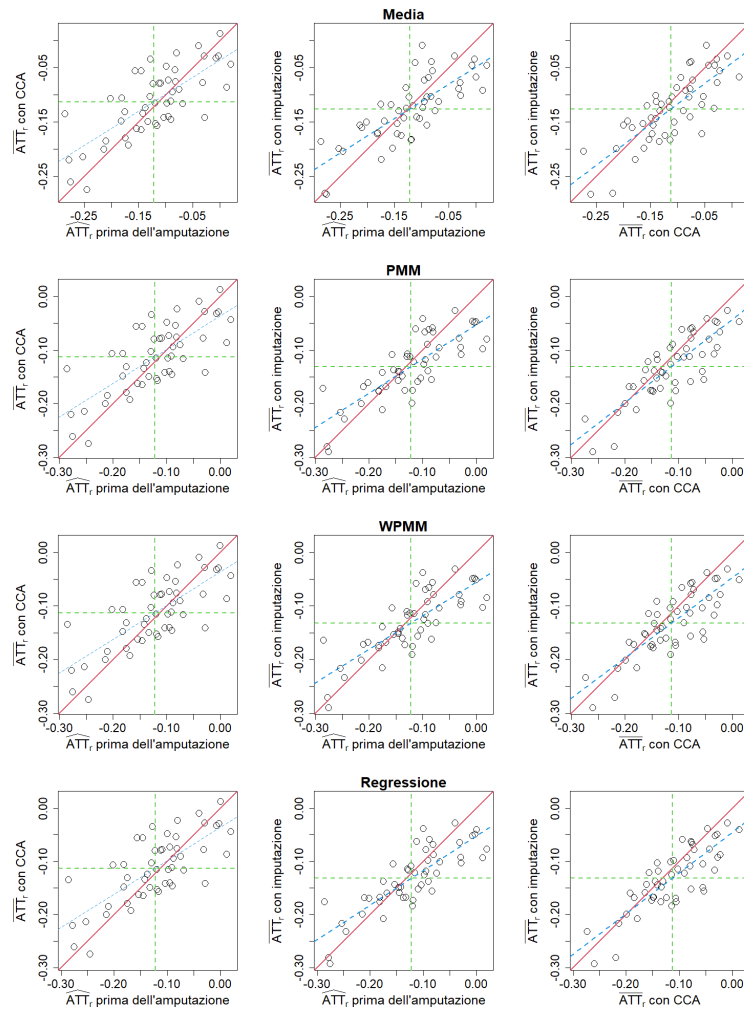


Figura A.121: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 8.

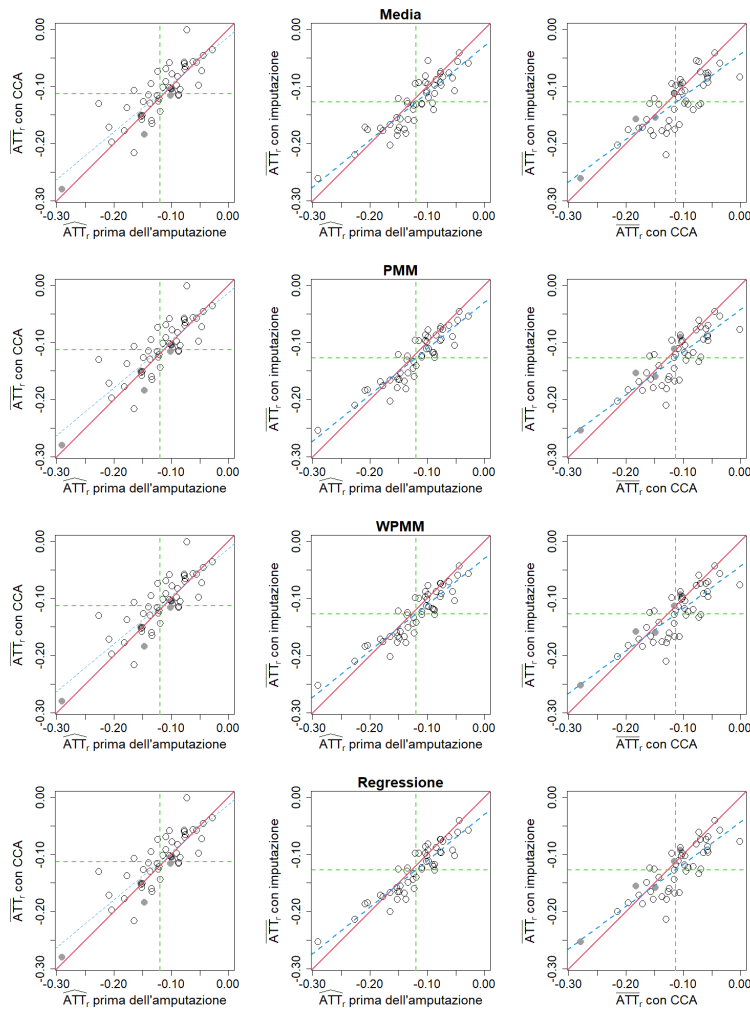


Figura A.122: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 8. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

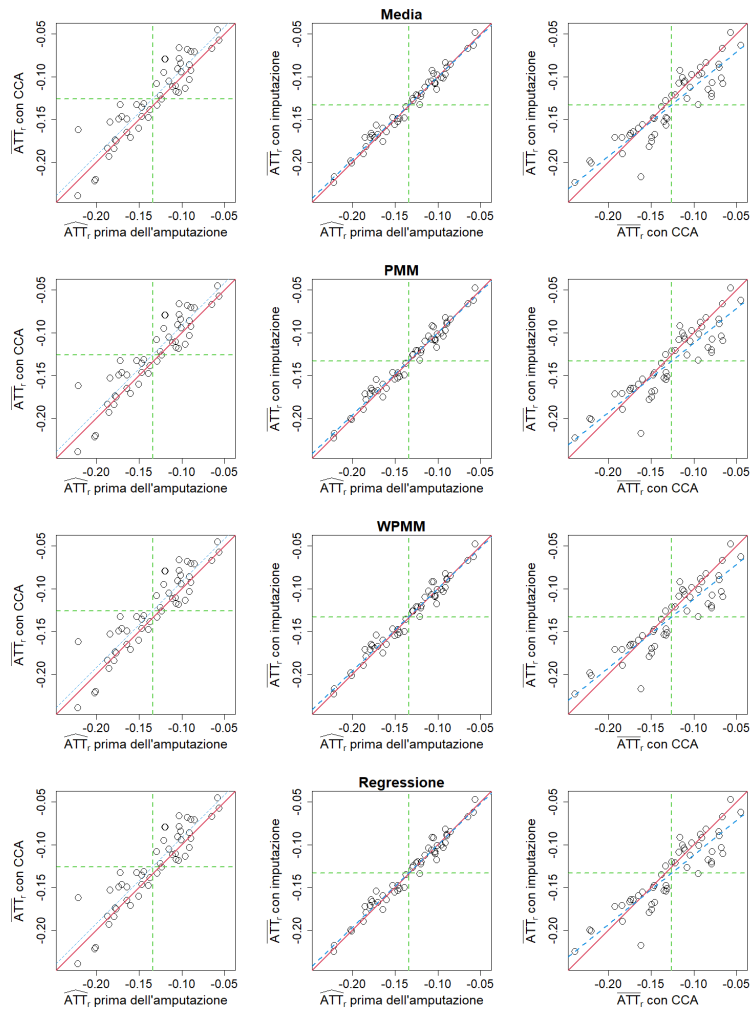


Figura A.123: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 8.

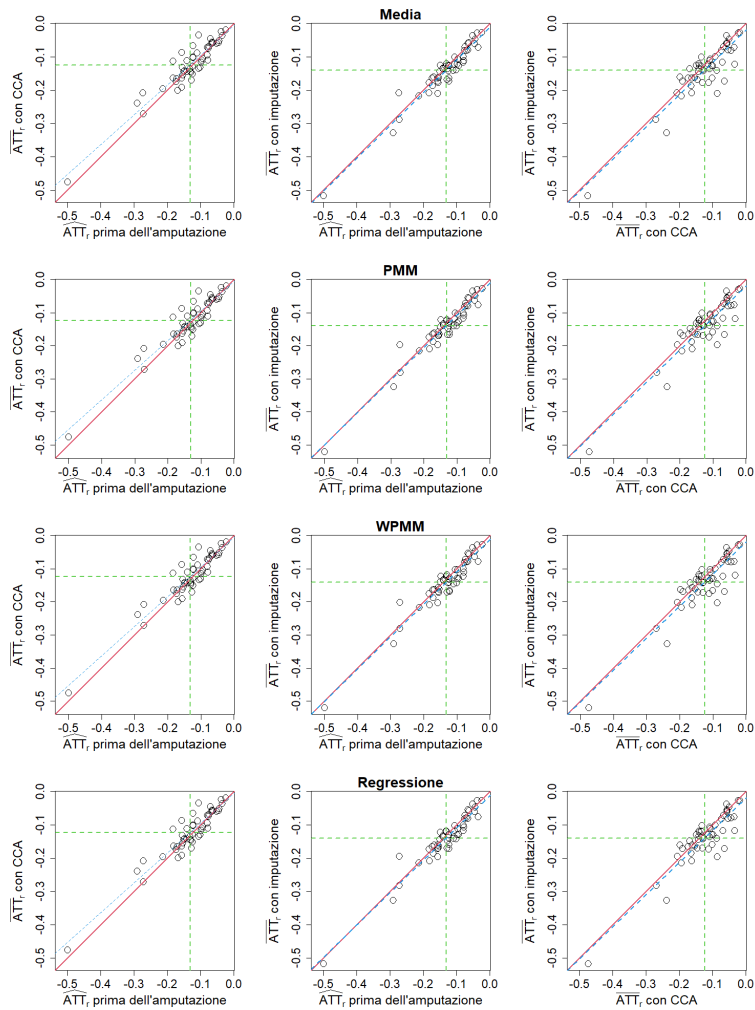


Figura A.124: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 8.

A.32 Simulazione 9 Scenario A

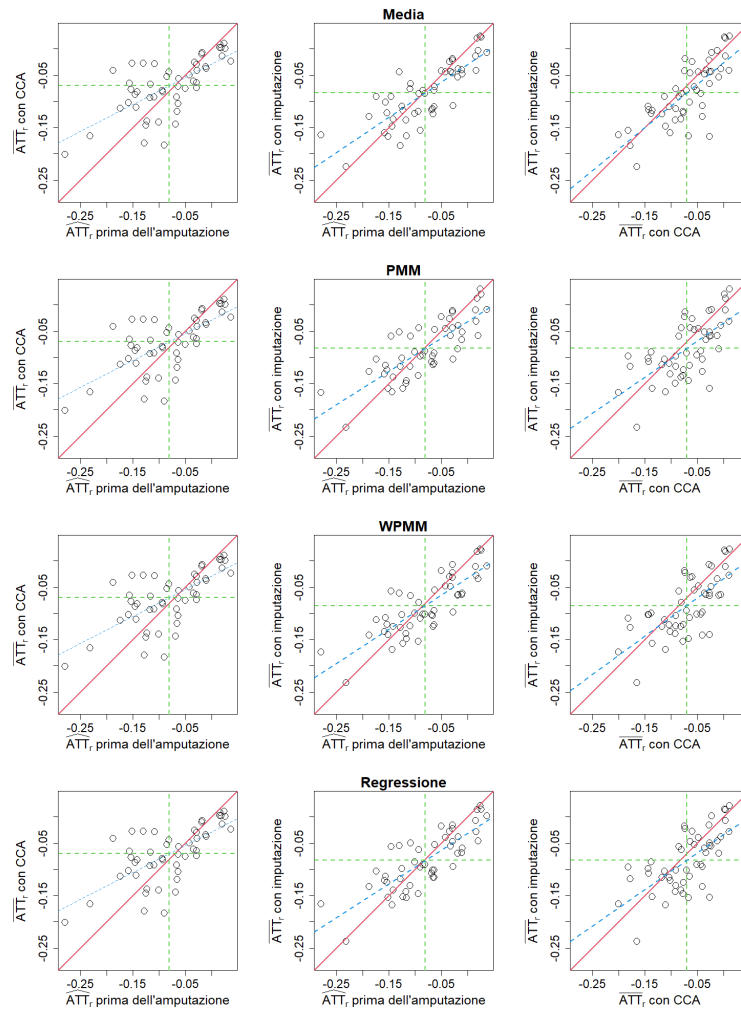


Figura A.125: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 9.

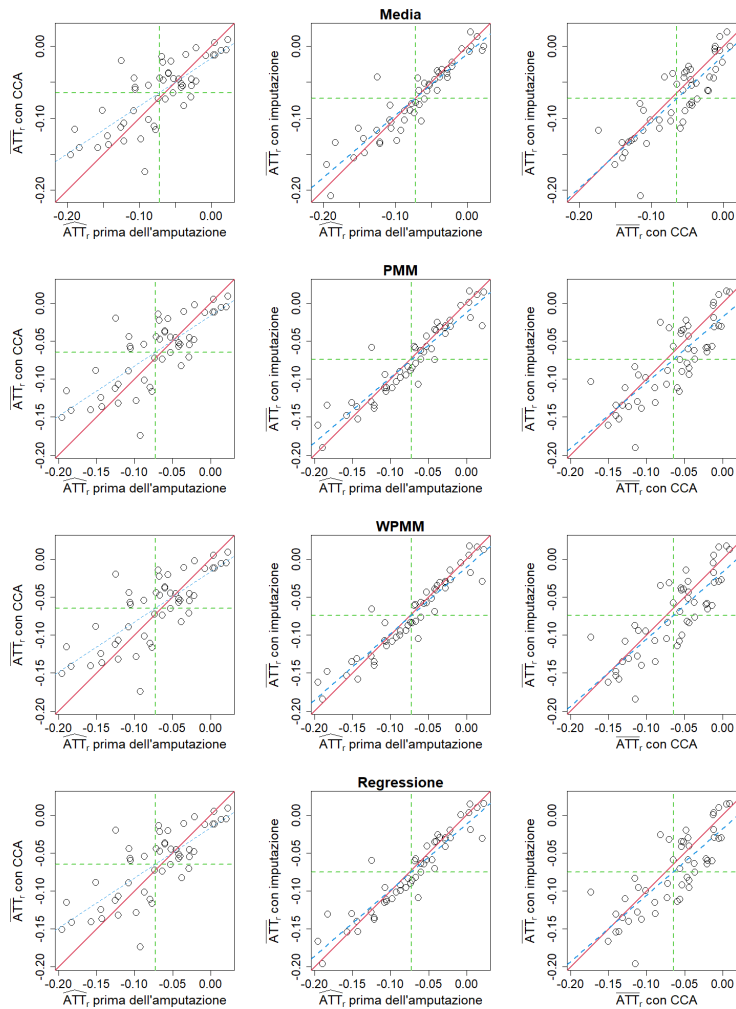


Figura A.126: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 9. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

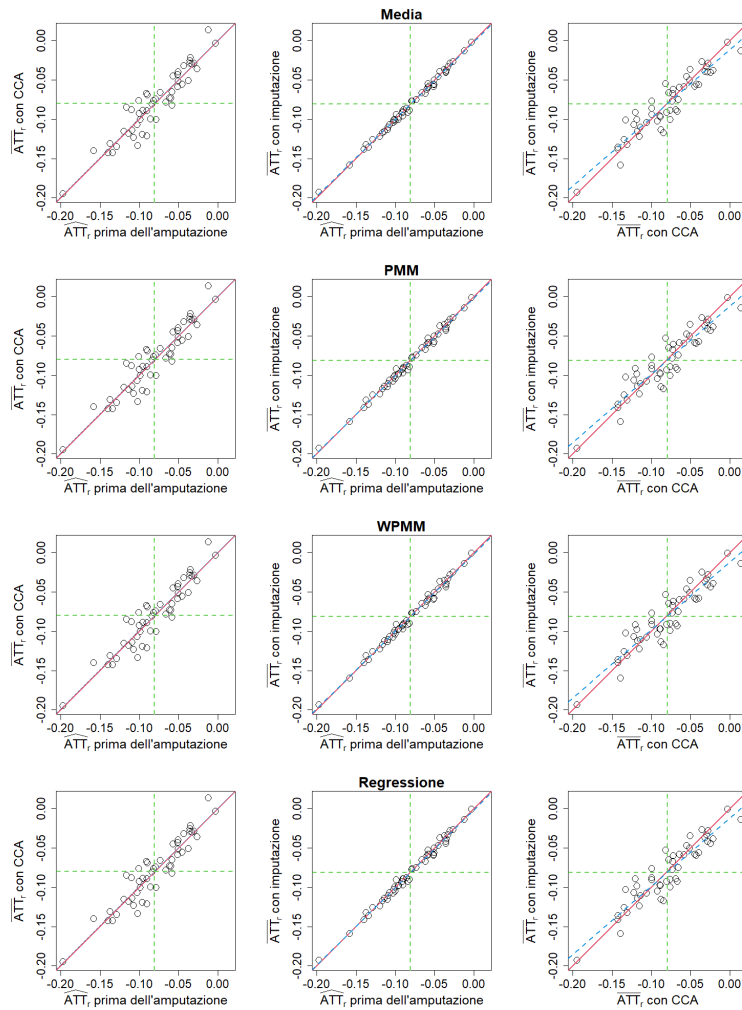


Figura A.127: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 9.

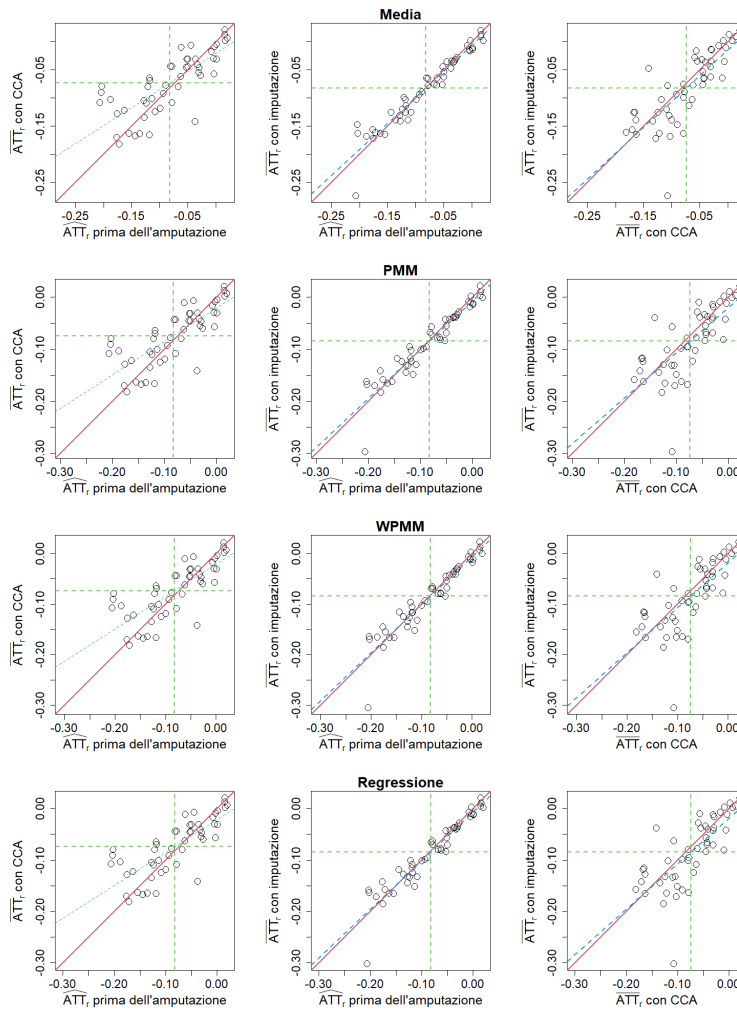


Figura A.128: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario A della Simulazione 9.

A.33 Simulazione 9 Scenario B

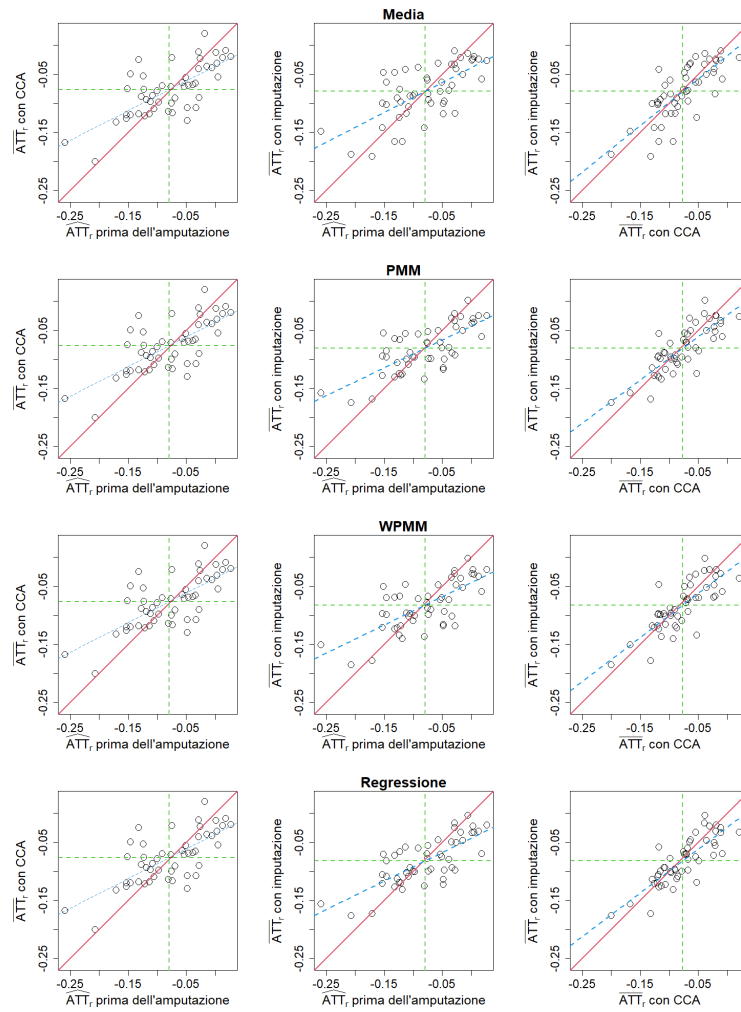


Figura A.129: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 9.

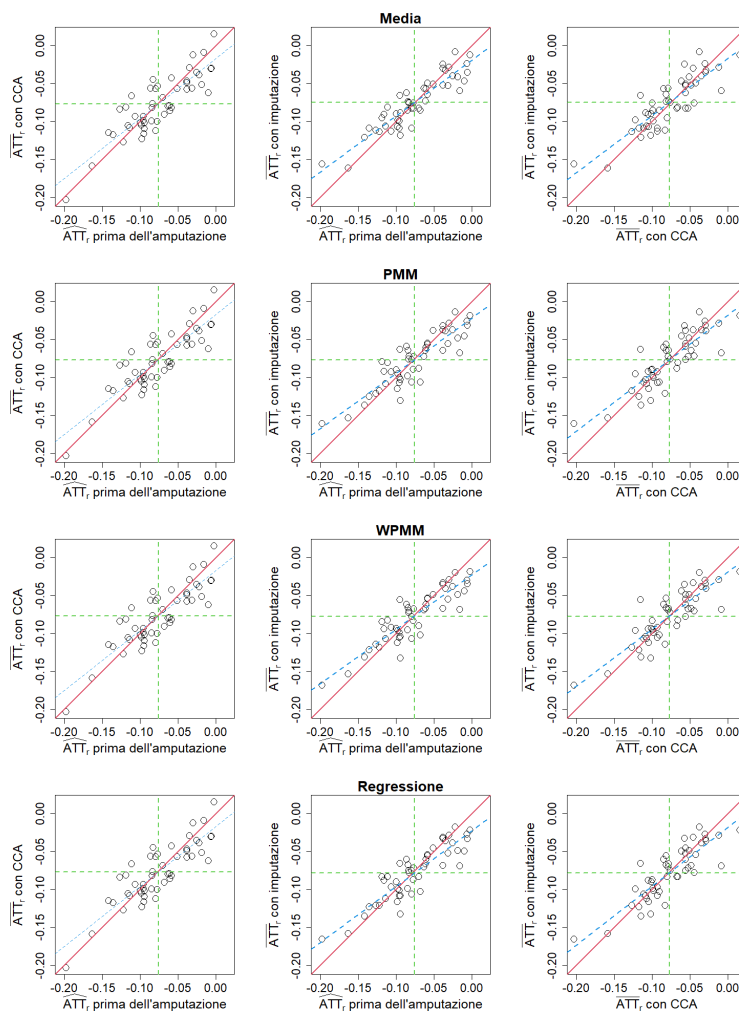


Figura A.130: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 9. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

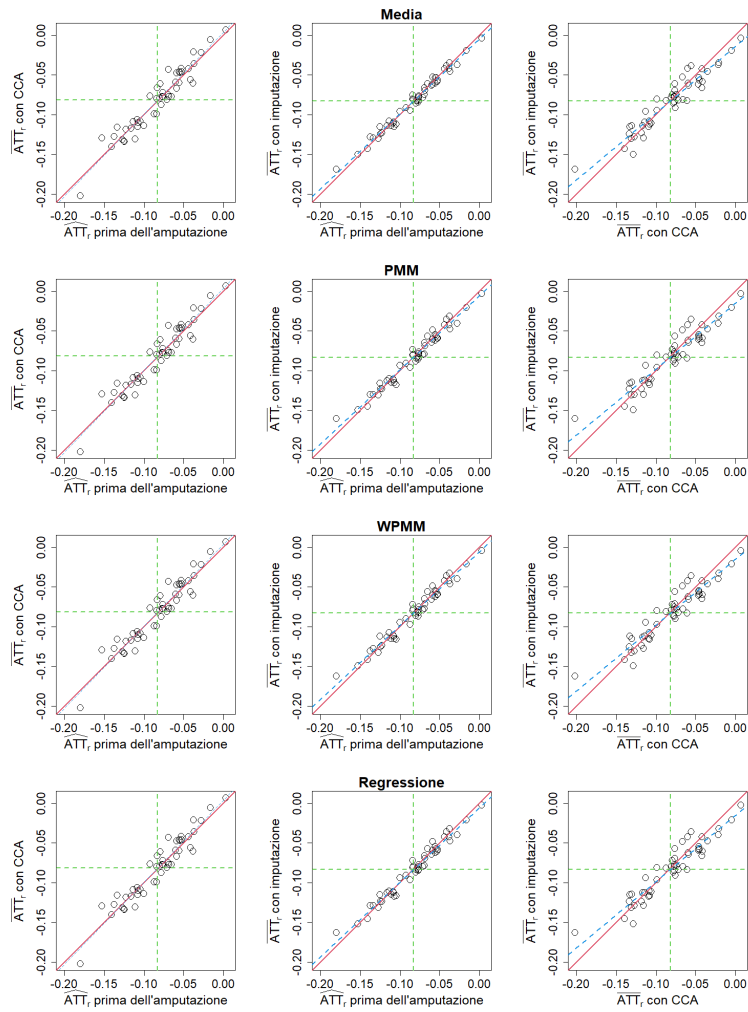


Figura A.131: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 9.

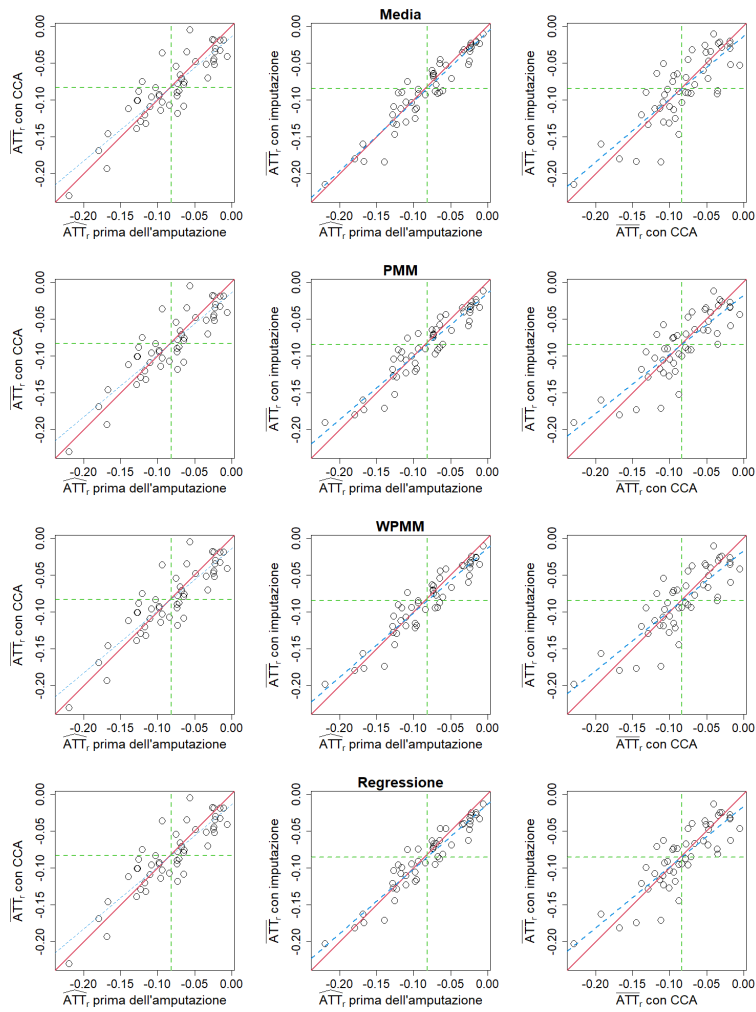


Figura A.132: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario B della Simulazione 9.

A.34 Simulazione 9 Scenario C

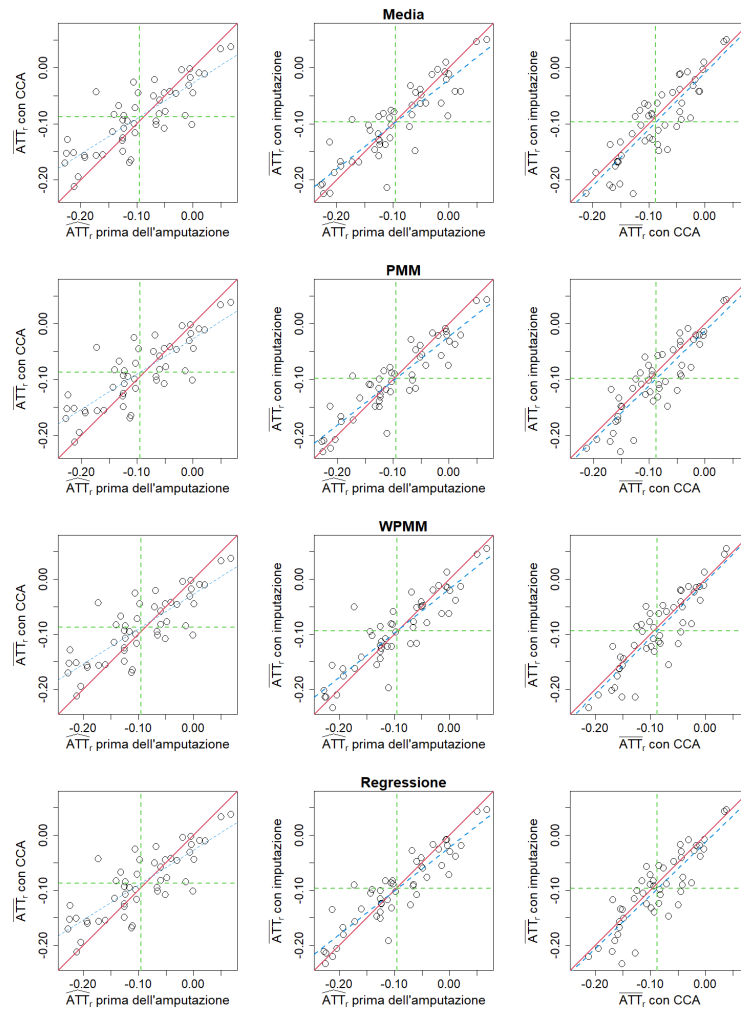


Figura A.133: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 9.

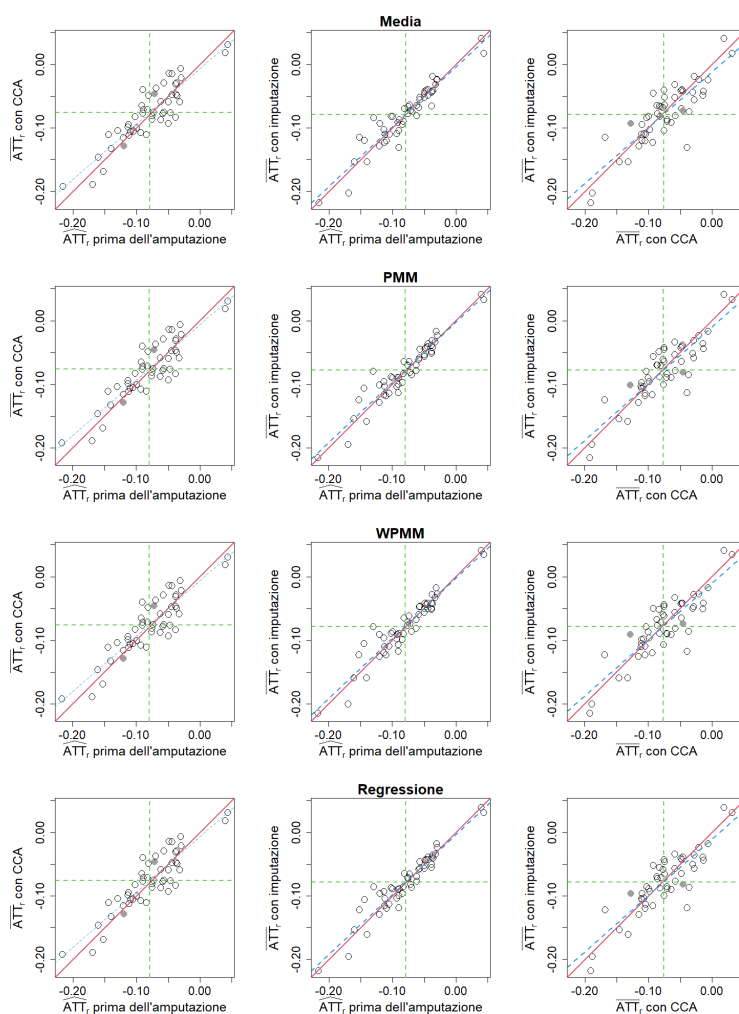


Figura A.134: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 9. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

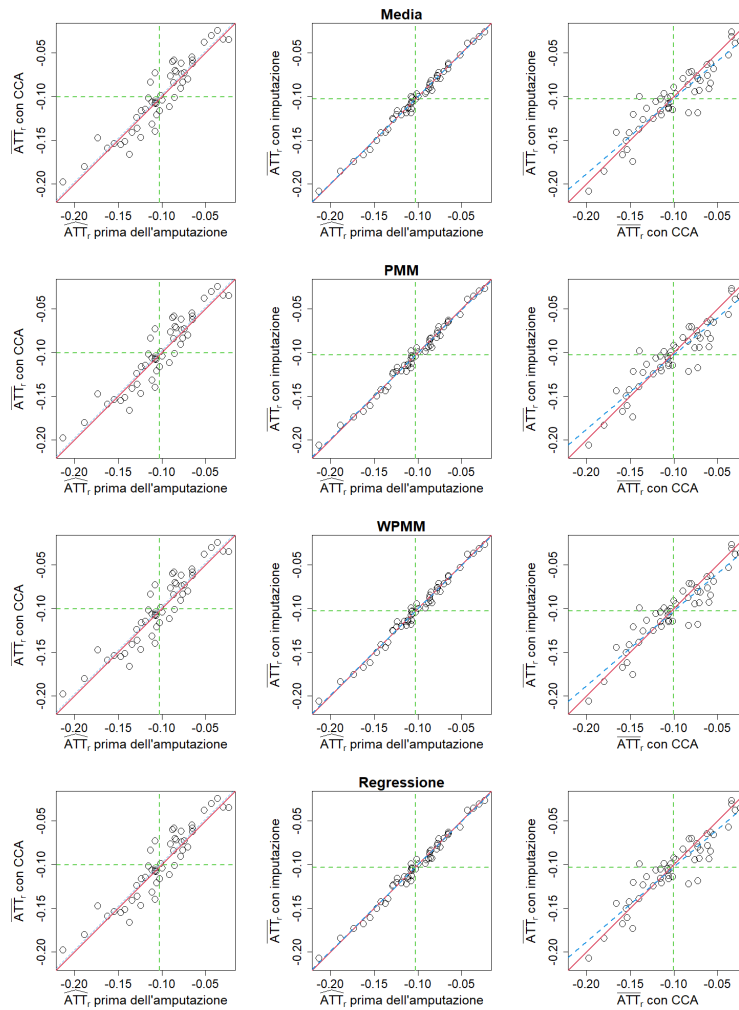


Figura A.135: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 9.

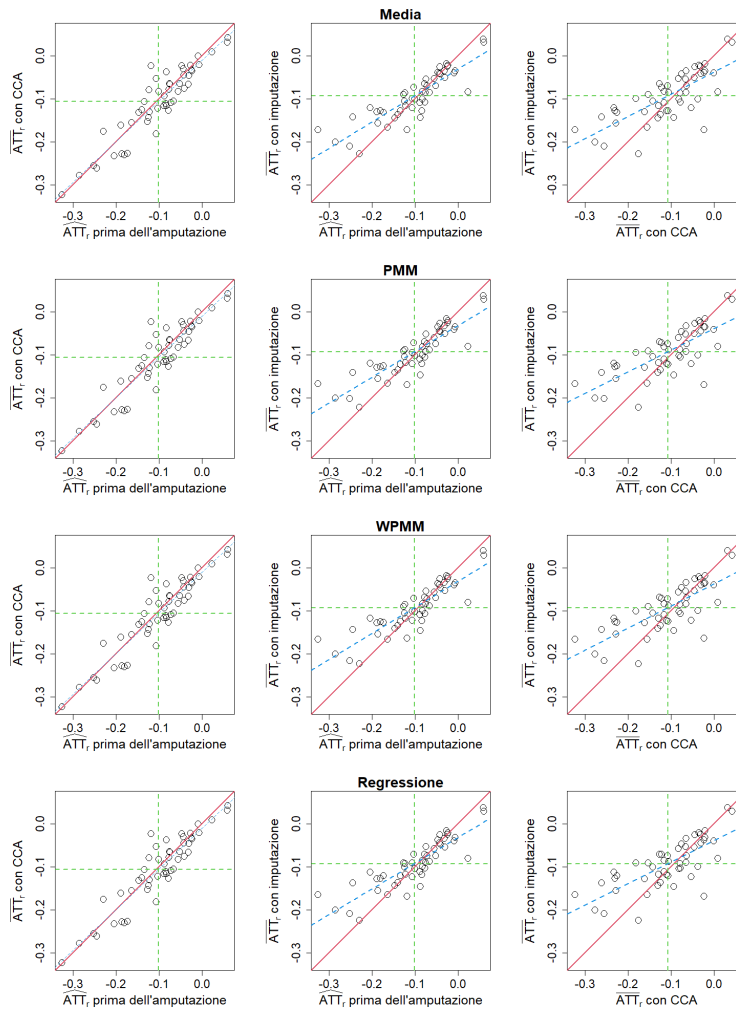


Figura A.136: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario C della Simulazione 9.

A.35 Simulazione 9 Scenario D

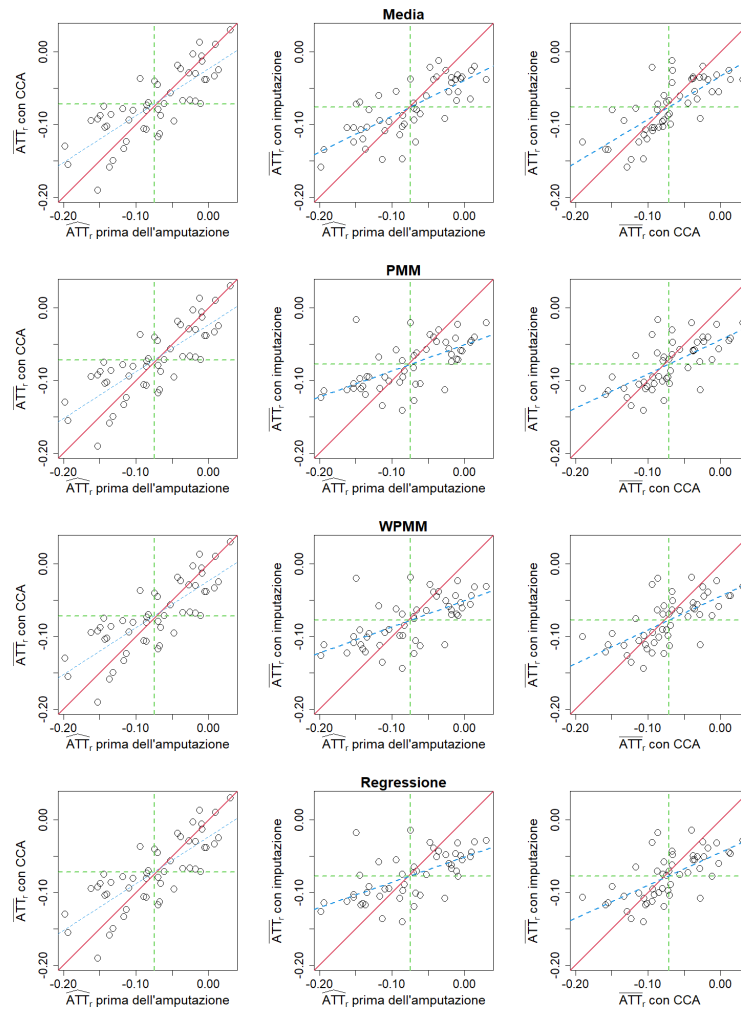


Figura A.137: Grafici di dispersione delle stime dell'ATT ottenute con il *matching* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 9.

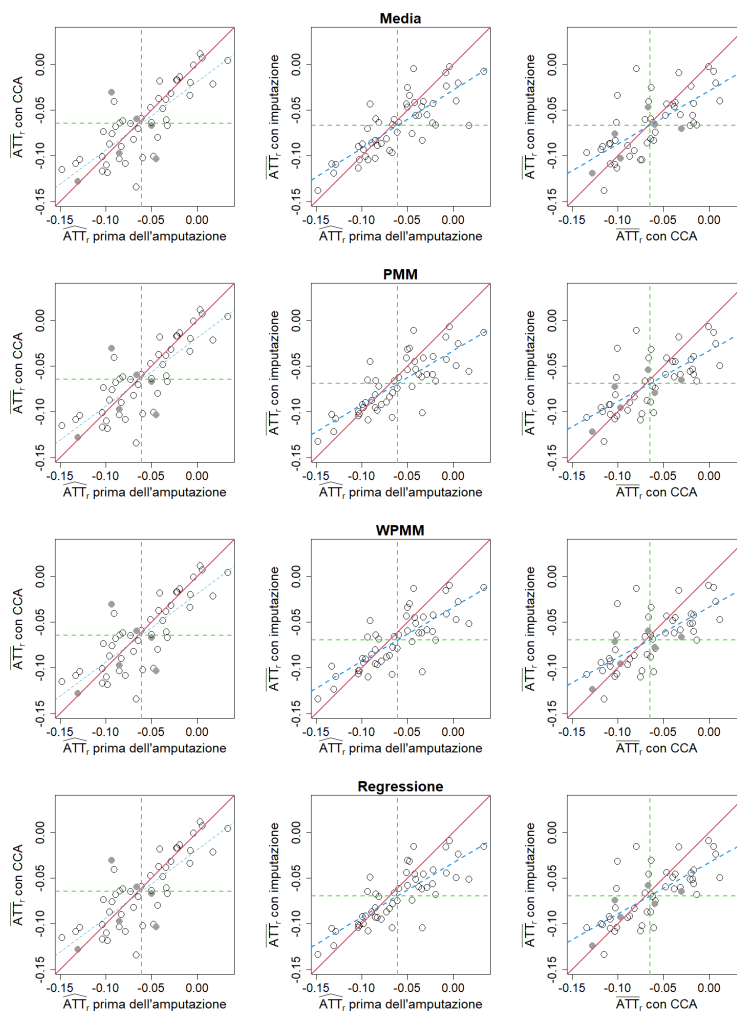


Figura A.138: Grafici di dispersione delle stime dell'ATT ottenute con la stratificazione attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 9. I punti grigi indicano le repliche in cui per almeno uno dei 10 dataset imputati non si è potuta ottenere la stima dell'ATT per la presenza di strati con nessun soggetto trattato o non trattato.

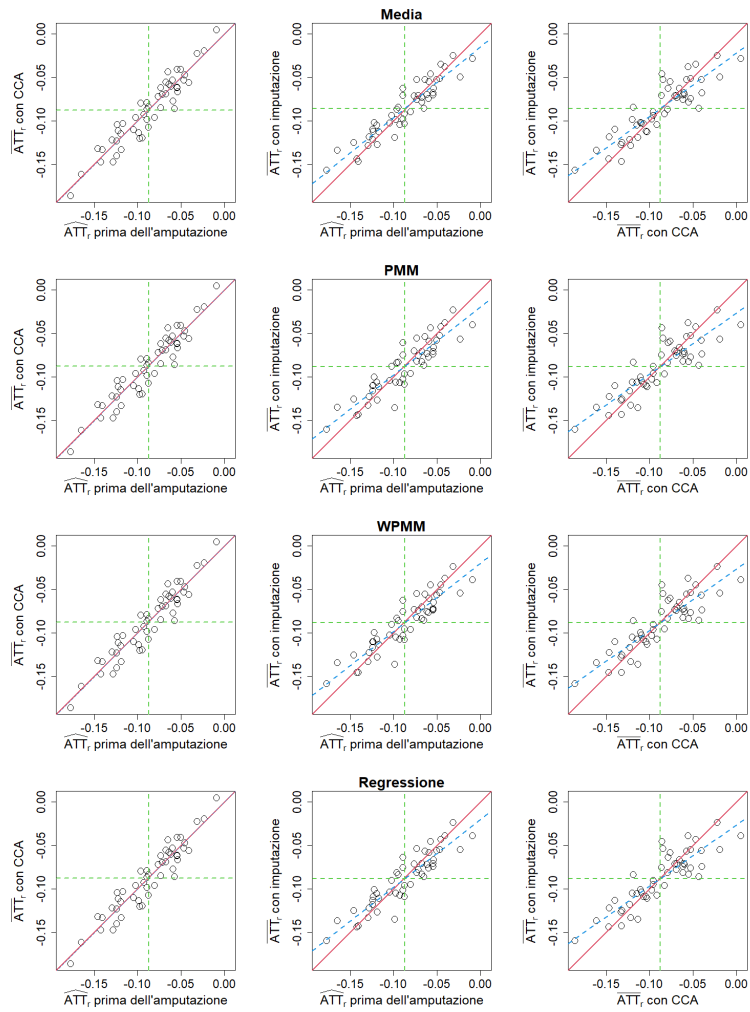


Figura A.139: Grafici di dispersione delle stime dell'ATT ottenute con il *covariate adjustment* attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 9.

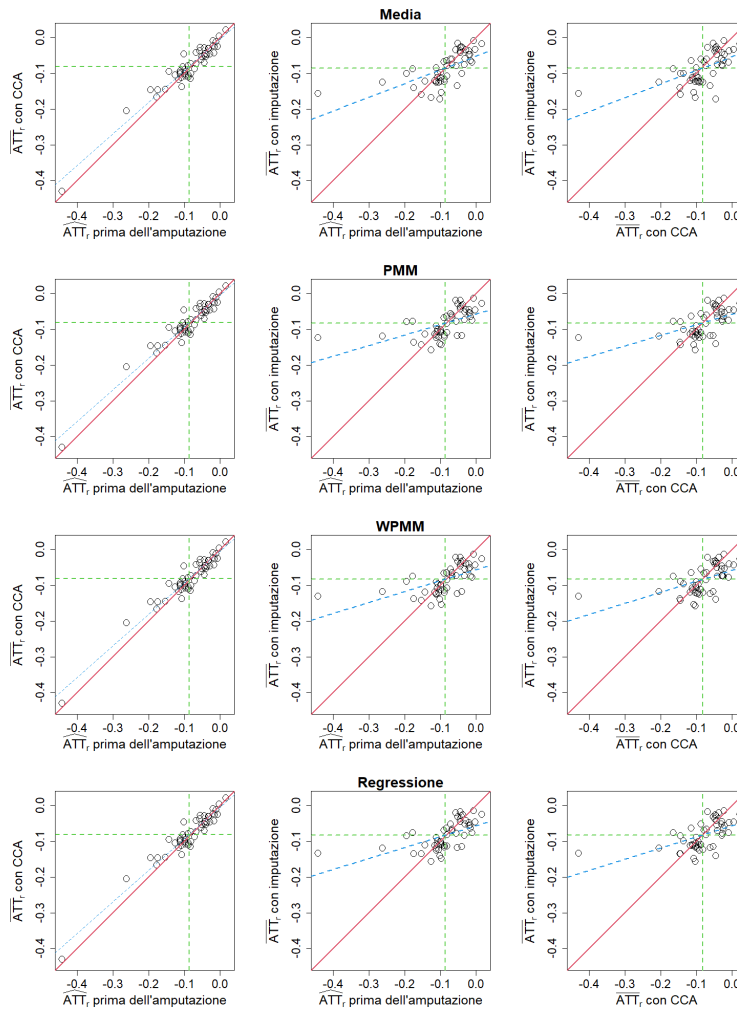


Figura A.140: Grafici di dispersione delle stime dell'ATT ottenute con l'IPTW attraverso i diversi metodi di gestione dei dati mancanti per lo Scenario D della Simulazione 9.

Bibliografia

- Abdia, Younathan et al. (2017). «Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study». *Biometrical Journal* 59.5, 967–985.
- Anderson, Andy B., Alexander Basilevsky e Derek P.J. Hum (1983). «Missing data: a review of the literature». *Handbook of Survey Research*. A cura di Andy B. Anderson Peter H. Rossi, James D Wright. New York: Elsevier, pp. 415–494.
- Andridge, Rebecca R. e Roderick J.A. Little (2010). «A review of hot deck imputation for survey non-response». *International Statistical Review* 78.1, 40–64.
- Austin, Peter C. (2006). «The performance of different propensity score methods for estimating marginal odds ratios». *Statistics in Medicine* 26, 3078–3094.
- Austin, Peter C. (2007). «The performance of different propensity score methods for estimating marginal odds ratios». *Statistics in Medicine* 26.16, 3078–3094.
- Austin, Peter C (2008a). «Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score». *Pharmacoepidemiology and Drug Safety* 17.12, 1202–1217.

- Austin, Peter C. (2008b). «The performance of different propensity-score methods for estimating relative risks». *Journal of Clinical Epidemiology* 61.6, 537–545.
- Austin, Peter C. (2009a). «Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and monte carlo simulations». *Biometrical Journal* 51.1, 171–184.
- Austin, Peter C. (2009b). «The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies». *Medical Decision Making* 29.6, 661–677.
- Austin, Peter C. (2010). «The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies». *Statistics in Medicine* 29.20, 2137–2148.
- Austin, Peter C. (2011a). «A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality». *Statistical Methods in Medical Research* 46.1, 119–151.
- Austin, Peter C. (2011b). «An introduction to propensity score methods for reducing the effects of confounding in observational studies». *Multivariate Behavioral Research* 46.3, 399–424.
- Austin, Peter C. (2011c). «Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples». *Statistics in Medicine* 30.11, 1292–1301.

- Austin, Peter C. (2011d). «Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies». *Pharmaceutical Statistics* 10.2, 150–161.
- Austin, Peter C. (2013). «The performance of different propensity score methods for estimating marginal hazard ratios». *Statistics in Medicine* 32.16, 2837–2849.
- Austin, Peter C. (2014). «A comparison of 12 algorithms for matching on the propensity score». *Statistics in Medicine* 33.6, 1057–1069.
- Austin, Peter C., Paul Grootendorst e Geoffrey M. Anderson (2007). «A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study». *Statistics in Medicine* 26.4, 734–753.
- Austin, Peter C. e Muhammad M. Mamdani (2006). «A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use». *Statistics in Medicine* 25.12, 2084–2106.
- Austin, Peter C. e Tibor Schuster (2016). «The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study». *Statistical Methods in Medical Research* 25.5, 2214–2237.
- Azur, Melissa J. et al. (2011). «Multiple imputation by chained equations: What is it and how does it work?». *International Journal of Methods in Psychiatric Research* 20.1, 40–49.
- Brand, J. P. L. (1999). «Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets».
- Brand, J. P. L. et al. (2003). «A toolkit in SAS for the evaluation of multiple imputation methods». *Statistica Neerlandica* 57.

- Brookhart, M. Alan et al. (2006). «Variable selection for propensity score models». *American Journal of Epidemiology* 163.12, 1149–1156.
- Casella, George e Edward I. George (1992). «Explaining the gibbs sampler». *The American Statistician* 46.3, 167–174.
- Cham, Heining e Stephen G. West (2016). «Propensity score analysis with missing data». *Psychological Methods* 21.3, 427–445.
- Choi, Jungyeon, Olaf M. Dekkers e Saskia le Cessie (2019). «A comparison of different methods to handle missing data in the context of propensity score analysis». *European Journal of Epidemiology* 34.1, 23–36.
- Cochran, W. G. (1968). «The effectiveness of adjustment by subclassification in removing bias in observational studies». *Biometrics* 24.2, 295–313.
- Dagenais, Marcel G. (1973). «The use of incomplete observations in multiple regression analysis. A generalized least squares approach». *Journal of Econometrics* 1.4, 317–328.
- D’Agostino, Ralph B. (1998). «Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group». *Statistics in Medicine* 17, 2265–2281.
- Donders, A. Rogier T. et al. (2006). «Review: a gentle introduction to imputation of missing values». *Journal of Clinical Epidemiology* 59.10, 1087–1091.
- Drake, Christiana (1993). «Effects of misspecification of the propensity score on estimators of treatment effect». *Biometrics* 49.4, 1231–1236.
- Efron, Bradley e Robert J. Tibshirani (1993). *An introduction to the bootstrap*. A cura di Chapman and Hall/CRC. Springer US.
- Elze, Markus C. et al. (2017). «Comparison of Propensity Score Methods and Covariate Adjustment». *Journal of the American College of Cardiology* 69.3, 345–357.

- Eulenburg, Christine et al. (2016). «Propensity scoring after multiple imputation in a retrospective study on adjuvant radiation therapy in lymph-node positive vulvar cancer». *PLoS ONE* 11.11, 1–12.
- Franklin, Jessica M. et al. (2017). «Comparing the performance of propensity score methods in healthcare database studies with rare outcomes». *Statistics in Medicine* 36.12, 1946–1963.
- Gelfand, Alan E. e Adrian F. M. Smith (1990). «Sampling-based approaches to calculating marginal densities». *Journal of the American Statistical Association* 85.410, 409.
- Granger, Emily, Jamie C. Sergeant e Mark Lunt (2019). «Avoiding pitfalls when combining multiple imputation and propensity scores». *Statistics in Medicine* 38.26, 5120–5132.
- Greenland, Sander e William D. Finkle (1995). «A critical look at methods for handling missing covariates in epidemiologic regression analyses». *American Journal of Epidemiology* 142.12, 1255–1264.
- Gu, Xing Sam e Paul R. Rosenbaum (1993). «Comparison of multivariate matching methods: structures, distances, and algorithms». *Journal of Computational and Graphical Statistics* 2.4, 405–420.
- Hansen, Ben B. (2004). *Full matching in an observational study of coaching for the SAT*.
- Heckman, James J., Hidehiko Ichimura e Petra E. Todd (1997). «Matching as an econometric evaluation estimator: evidence from evaluating a job-training programme». *The Review of Economic Studies* 64.4, 605–654.
- Heijden, Geert J.M.G. van der et al. (2006). «Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example». *Journal of Clinical Epidemiology* 59.10, 1102–1109.

- Heitjan, Daniel F. e Roderick J. A. Little (1991). «Multiple imputation for the fatal accident reporting system». *Journal of the Royal Statistical Society* 40.1, 29.
- Hill, Jennifer (2004). «Reducing bias in treatment effect estimation in observational studies suffering from missing data».
- Holland, Paul W. (1986). «Statistics and causal inference». *Journal of the American Statistical Association* 81.396, 945–960.
- Horton, Nicholas J. e Ken P. Kleinman (2007). «Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models». *The American Statistician* 61.1, 79–90.
- Horvitz, D. G. e D. J. Thompson (1952). «A generalization of sampling without replacement from a finite universe». *Journal of the American Statistical Association* 47.260, 663–685.
- Imbens, Guido W (2004). «Nonparametric estimation of average treatment effects under exogeneity: A review». *The Review of Economics and Statistics* 86.1, 4–29.
- Kara, Yusuf et al. (2021). «Covariate balance as a quality indicator for propensity score analysis». *Journal of Measurement and Evaluation in Education and Psychology* 12.4, 374–387.
- Knol, Mirjam J. et al. (2010). «Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example». *Journal of Clinical Epidemiology* 63.7, 728–736.
- Lee, Brian K., Justin Lessler e Elizabeth A. Stuart (2010). «Improving propensity score weighting using machine learning». *Statistics in Medicine* 29.3, 337–346.
- Lee, Brian K., Justin Lessler e Elizabeth A. Stuart (2011). «Weight trimming and propensity score weighting». *PLoS one* 6.3.

- Leyrat, Clémence et al. (2019). «Propensity score analysis with partially observed covariates: How should multiple imputation be used?» *Statistical Methods in Medical Research* 28.1, 3–19. eprint: 1608.05606.
- Ling, A et al. (2020). «How to apply multiple imputation in propensity score catching with partially observed confounders: A simulation study and practical recommendations». *Journal of Modern Applied Statistical Methods* 19.1, eP3439.
- Little, Roderick (2009). «Selection and pattern-mixture models». *Longitudinal Data Analysis*. A cura di Garrett Fitzmaurice et al. Boca Raton, FL: CRC Press. Cap. 18, pp. 409–431.
- Little, Roderick J A (1992). «Regression with missing X's: a review». *Journal of the American Statistical Association* 87.420, 1227–1237.
- Little, Roderick J. A. e Donald B. Rubin (2002). *Statistical analysis with missing data*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Little, Roderick J.A. (1988). «Missing-data adjustments in large surveys». *Journal of Business and Economic Statistics* 6.3, 287–296.
- Lunceford, Jared K. e Marie Davidian (2004). «Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study». *Statistics in Medicine* 23.19, 2937–2960.
- Ma, Sophie et al. (2021). «Comparisons of propensity score methods for time to event outcomes: evaluation through simulations and oral squamous cell carcinoma case study». *University of Toronto Journal of Public Health* 2.2.
- Marshall, Andrea, Douglas G. Altman e Roger L. Holder (2010). «Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: A resampling study». *BMC Medical Research Methodology* 10.1, 112.

- Marshall, Andrea et al. (2010). «Comparison of techniques for handling missing covariate data within prognostic modelling studies: A simulation study». *BMC Medical Research Methodology* 10.1, 7.
- McCaffrey, Daniel F., Greg Ridgeway e Andrew R. Morral (2004). «Propensity score estimation with boosted regression for evaluating causal effects in observational studies». *Psychological Methods* 9.4, 403–425.
- Miettinen, Olli S. (1985). «Regression analysis». *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*. Wiley, pp. 216–243.
- Ming, Kewei e Paul R. Rosenbaum (2000). «Substantial gains in bias reduction from matching with a variable number of controls». *Biometrics* 56.1, 118–124.
- Mitra, Robin e Jerome P. Reiter (2011). «Propensity score matching with missing covariates via iterated, sequential multiple imputation». Southampton.
- Mitra, Robin e Jerome P. Reiter (2016). «A comparison of two methods of estimating propensity scores after multiple imputation». *Statistical Methods in Medical Research* 25.1, 188–204.
- Package 'Matching' (2022). URL: <https://cran.r-project.org/web/packages/Matching/Matching.pdf>.
- Package 'mice' (2021). URL: <https://cran.r-project.org/web/packages/mice/mice.pdf>.
- Penning De Vries, Bas B.L. e Rolf H.H. Groenwold (2017). «A comparison of approaches to implementing propensity score methods following multiple imputation». *Epidemiology Biostatistics and Public Health* 14.4, e12630–1–e12630–21.

- Raghunathan, Trivellore E. et al. (2001). «A multivariate technique for multiply imputing missing values using a sequence of regression models». *Survey Methodology* 27, 85–95.
- Rosenbaum, Paul R. (1987). «Model-based direct adjustment». *Journal of the American Statistical Association* 82.398, 394.
- Rosenbaum, Paul R. (1991). «A characterization of optimal designs for observational studies». *Journal of the Royal Statistical Society: Series B (Methodological)* 53.3, 597–610.
- Rosenbaum, Paul R. e Donald B. Rubin (1983). «The central role of the propensity score in observational studies for causal effects». *Biometrika* 70.1, 41–55.
- Rosenbaum, Paul R. e Donald B. Rubin (1984). «Reducing bias in observational studies using subclassification on the propensity score». *Journal of the American Statistical Association* 79.387, 516–524.
- Rosenbaum, Paul R. e Donald B. Rubin (1985). «Constructing a control group using multivariate matched sampling methods that incorporate the propensity score». *The American Statistician* 39.1, 33–38.
- Rubin, Donald B (1973). «The use of matched sampling and regression adjustment to remove bias in observational studies». *Biometrics* 29.1, 185–203.
- Rubin, Donald B (1976). «Inference and missing data». *Biometrika* 63.3, 581–592.
- Rubin, Donald B (1978). «Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse». *American Statistical Association* 1, 20–34.

- Rubin, Donald B (1979). «Using multivariate matched sampling and regression adjustment to control bias in observational studies». *Matched Sampling for Causal Effects* 74.366, 318–328.
- Rubin, Donald B (1986). «Statistical matching using file concatenation with adjusted weights and multiple imputations». *Journal of Business & Economic Statistics* 4.1, 87–94.
- Rubin, Donald B. (1987). *Multiple imputation for nonresponse in surveys*. A cura di Inc. John Wiley & Sons. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc.
- Rubin, Donald B. (2004). «The design of a general and flexible system for handling nonresponse in sample surveys». *American Statistician* 58.4, 298–302.
- Saar-Tsechansky, Maytal (2007). «Handling missing values when applying classification models». 8, 1625–1657.
- Schafer, Joseph L. (1997). *Analysis of incomplete multivariate Data*. New York: Chapman e Hall/CRC.
- Schafer, Joseph L. e Maren K. Olsen (1998). «Multiple imputation for multivariate missing-data problems: A data analyst's perspective». *Multivariate Behavioral Research* 33.4, 545–571.
- Schouten, Rianne, Gerko Vink e Peter Lugtig. *ampute: Generate missing data for simulation purposes*. URL: <https://rdrr.io/cran/mice/man/ampute.html>.
- Schouten, Rianne et al. (2020). *Generate missing values with ampute*. URL: https://rianneschouten.github.io/mice_ampute/vignette/ampute.html.
- Schouten, Rianne Margaretha, Peter Lugtig e Gerko Vink (2018). «Generating missing values for simulation purposes: a multivariate amputation

- procedure». *Journal of Statistical Computation and Simulation* 88.15, 2909–2930.
- Schouten, Rianne Margaretha e Gerko Vink (2021). «The dance of the mechanisms: how observed information influences the validity of missingness assumptions». *Sociological Methods & Research* 50.3, 1243–1258.
- Sekhon, Jasjeet (2011). «Multivariate and propensity score matching». *Journal of Statistical Software* 42.7, 52. arXiv: arXiv:1501.0228.
- Setoguchi, Soko et al. (2008). «Evaluating uses of data mining techniques in propensity score estimation: A simulation study». *Pharmacoepidemiology and Drug Safety* 17.6, 546–555.
- Tang, Lingqi et al. (2005). «A comparison of imputation methods in a longitudinal randomized clinical trial». *Statistics in Medicine* 24.
- van Buuren, S e CGM Oudshoorn (2000). *Multivariate Imputation by Chained Equations: MICE V1.0 User's manual TNO Prevention and Health TNO Prevention and Health TNO Prevention. Technical Report PG/VG-Z/00.038*. Rapp. tecn. TNO Prevention e Health.
- van Buuren, S. et al. (2006). «Fully conditional specification in multivariate imputation». *Journal of Statistical Computation and Simulation* 76.12, 1049–1064.
- van Buuren, Stef (2007). «Multiple imputation of discrete and continuous data by fully conditional specification». *Statistical Methods in Medical Research* 16.3, 219–242.
- van Buuren, Stef (2018). *Flexible Imputation of Missing Data*. A cura di P. van der Heijden N. Keiding, B.J.T. Morgan, C.K. Wikle. Second Edition. Chapman & Hall /CRC.

- van Buuren, Stef e Karin Groothuis-Oudshoorn (2011). «mice: Multivariate imputation by chained equations in R». *Journal of Statistical Software* 45.3, 1–67.
- Westreich, Daniel, Justin Lessler e Michele Jonsson Funk (2010). «Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression». *Journal of Clinical Epidemiology* 63.8, 826–833.
- Williamson, Elizabeth et al. (2012). «Propensity scores: From naïve enthusiasm to intuitive understanding». *Statistical Methods in Medical Research* 21.3, 273–293.
- Wyss, Richard et al. (2014). «The role of prediction modeling in propensity score estimation: An evaluation of logistic regression, bcart, and the covariate-balancing propensity score». *American Journal of Epidemiology* 180.6, 645–655.
- Yu, Liu, Andrea Burton e Oliver Rivero-Arias (2007). «Evaluation of software for multiple imputation of semi-continuous data». *Statistical Methods in Medical Research* 16, 243–258.
- Yu, Youfei et al. (2021). «A comparison of parametric propensity score-based methods for causal inference with multiple treatments and a binary outcome». *Statistics in Medicine* 40.7, 1653–1677.