

## La simulazione nella valutazione neuropsicologica come problema diagnostico: le statistiche bayesiane

### Malingering in the neuropsychological assessment as a diagnostic matter: the bayesian statistics

Andrea Bosco

#### Abstract

Neuropsychological assessments can intentionally be distorted threatening their results. Such events has called simulation or malingering. The purpose of this study is to compare two diagnostic approaches, adopting both SIMS and ToMM tests, and three levels of prevalence for malingering performing Bayesian statistics and Monte Carlo methods. The most conservative approach, according to which the tests must be both positive for the diagnosis of malingering, showed very good diagnostic accuracy for all considered levels of prevalence and excellent results in minimizing false positives. An example in which Bayesian statistics are applied to the diagnostic process is shown with some general considerations regarding the topic of malingering detection.

**Key words:** Malingering, bayesian statistics, Monte Carlo Study, SIMS, TOMM

#### Riassunto

I risultati delle valutazioni neuropsicologiche possono essere intenzionalmente distorti minacciandone gli esiti. Tale eventualità è chiamata simulazione o *malingering*. Lo scopo di questo studio è di mettere a confronto due approcci diagnostici che adottano entrambi il SIMS e il ToMM e tre livelli di prevalenza del malingering mediante statistiche bayesiane e metodo Monte Carlo. L'approccio più conservativo, secondo cui i test devono essere entrambi positivi per la diagnosi di malingering, ha risultati di accuratezza diagnostica ottimi per i livelli di prevalenza considerati e risulta eccellente nel minimizzare i falsi positivi. Lo studio si conclude con un esempio di procedura di valutazione in cui le statistiche bayesiane si applicano al processo diagnostico e con alcune riflessioni sulle procedure per intercettare il malingering.

**Parole chiave:** Simulazione, statistiche bayesiane, studio Monte Carlo, SIMS, TOMM

---

Per corrispondenza: ANDREA BOSCO, Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro" • e-mail: [andrea.bosco@uniba.it](mailto:andrea.bosco@uniba.it)

Andrea BOSCO, Professore Associato di Psicometria, Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università degli Studi di Bari "Aldo Moro"

## 1. Il malingering

La valutazione neuropsicologica è fondata sull'osservazione del comportamento in condizioni controllate e richiede che il paziente sia cosciente e collaborante. Il significato di cosciente è piuttosto chiaro e non desta difficoltà definitorie. Al contrario vi è maggiore incertezza sul concetto di collaborante soprattutto quando il valutando è inviato da altri alla valutazione neuropsicologica. Secondo Bianchi e La Mastra (2005 p. 298) "è collaborante chi è in grado di comprendere gli obiettivi, le richieste, i compiti proposti dal valutatore ed è intenzionato a condividere gli scopi della procedura proposta". Usualmente in contesto forense gli scopi del valutatore e del valutando divergono, dal momento che il valutatore è reclutato da terzi per ragioni connesse con un giudizio che avrà degli effetti sull'esistenza del valutando. Nelle cause civili per traumi o altri tipi di ingiurie che più frequentemente richiedono la valutazione del neuropsicologo, il valutando può voler apparire più malato di quanto lo sia in realtà perché intende conservare lo status di malato (i.e. disturbo fittizio), subisce l'effetto di fattori psicologici e sociali che aggravano i disturbi (i.e. disturbo somatoforme), finge di essere malato al fine di procurarsi vantaggi (i.e. simulazione). Questo ultimo caso è definito malingering e costituisce il focus di questo studio.

Una prima ed esaustiva definizione di malingering la dobbiamo ad un gruppo di lavoro in seno all'American Psychiatric Association che nel 1994 lo ha definito come una più o meno deliberata esagerazione o invenzione di disturbi psicologici e/o fisici, come risposta ad un incentivo esterno quali l'evitamento di circostanze negative (p.e. una condanna) oppure l'ottenere un guadagno tangibile (p.e. un risarcimento danni). La classificazione dei diversi tipi di malingering è ancora oggetto di discussione. Diversi contributi tra gli anni ottanta e novanta si sono concentrati sul grado di falsificazione. Secondo Resnick (1988, 1992) può essere totale (i.e. sintomi completamente inventati) o parziale (i.e. sintomi presenti ma esagerati). Rogers (1988) suggerisce una divisione in tre livelli: falsificazione lieve che prevede esclusivamente una esagerazione di sintomi, falsificazione moderata che prevede anche il ricorso all'invenzione di sintomi e una falsificazione severa che prevede il ricorso all'invenzione di sintomi che risultano impropri o assurdi per la patologia lamentata. Lo stesso Resnick (1992) propone inoltre una classificazione del malingering in funzione della sua direzione: la simulazione riguarderebbe l'esagerazione di sintomi non posseduti, al contrario la dissimulazione si riferirebbe al tentativo di minimizzare sintomi, disturbi o abitudini (p.e. il ricorso a droghe) che effettivamente spiegherebbero i comportamenti attribuiti al valutando. In terzo luogo esisterebbe un tipo di malingering che potremmo chiamare "falsa attribuzione" in cui il disturbo genuinamente lamentato, è fatto risalire a circostanze diverse da

quelle che lo hanno effettivamente causato, usualmente con lo scopo di ottenere un risarcimento o un vantaggio di qualche tipo. Una classificazione ormai classica che può essere attribuita ad un contributo di Slick, Sherman e Iverson (1999) associa il malingering a tre livelli di sicurezza nella sua attribuzione. Esso può essere *possibile*, *probabile* o *definito*. Tale classificazione ha senso soprattutto nella descrizione della simulazione di deficit cognitivi. La classificazione di probabile malingering richiede la coesistenza di alcune condizioni: deve esistere un incentivo concreto a simulare una disfunzione cognitiva, sono presenti indizi chiari che suggeriscono una volontaria esagerazione / creazione di sintomi, sono state escluse tutte le spiegazioni alternative. Sul piano economico i rimborsi assicurativi / risarcimenti ottenuti in modo fraudolento possono essere rilevanti così come lo sono i costi sanitari e assistenziali non dovuti. Al contempo l'accertamento del malingering può determinare la negazione di rimborsi / risarcimenti a persone genuinamente danneggiate ma erroneamente classificate come simulatrici. Questo costituisce un problema etico importante per le scienze forensi che non può essere sottovalutato.

Una domanda urgente che ci viene dagli esperti di valutazione riguarda la messa a punto di studi epidemiologici sul malingering. Infatti il processo diagnostico ha bisogno di informazioni chiare sulla prevalenza di un fenomeno o di una caratteristica per poter procedere all'approfondimento e ottenere da questo una risposta appropriata. Non esistono molti studi che affrontino in modo sistematico questo tema. Un articolo particolarmente importante è il lavoro di Mittenberg, Patton, Canyock e Condit pubblicato nel 2002 su *Journal of Clinical and Experimental Neuropsychology*. La sua importanza è testimoniata dal numero di citazioni che ha ottenuto nel corso degli anni che ora è superiore a 600. Non esistono altri articoli che trattino in modo così ampio il tema della prevalenza del malingering pertanto mi baserò in questo studio proprio sulle informazioni derivate dall'eccellente studio di Mittenberg et al. (2002). Secondo questo studio la prevalenza di simulazione di un disturbo neurologico tra cui l'amnesia anterograda in seguito a trauma cerebrale, ipossia, avvelenamento, potrebbe variare approssimativamente tra 0,2 e 0,4, una prevalenza piuttosto elevata sulla quale esiste una buona convergenza tra studi indipendenti.

## 2. Cenni sul modello diagnostico bayesiano e sul metodo Monte Carlo

Per comprendere la ratio del presente studio è importante chiarire anche l'ambito statistico nel quale ci si muove. In primo luogo se nella psicomatria tradizionale il concetto di attendibilità è ancorato al principio della coerenza interna (si pensi all'Alpha di Cronbach), in ambito diagnostico il

concetto di attendibilità della misura è strettamente ancorato a quello della validità concorrente e, ancora di più, di criterio. Per un'introduzione in italiano a questa parte della statistica nel contesto della diagnosi psicologica si veda Lauriola (2007). Uno strumento può essere considerato affidabile solo se dimostra di rispondere adeguatamente alle esigenze della diagnosi che, in una parola, si riassumono nella necessità di minimizzare i falsi negativi (cioè persone che presentano la caratteristica o la patologia ma non vengono intercettati dal test) e i falsi positivi (persone che non presentano la caratteristica o la patologia eppure il test li identifica come in possesso della caratteristica). I parametri in grado di fornire queste informazioni sono la sensibilità (Se) e la specificità del test (Sp). La prima esprime la proporzione dei positivi su tutti quelli che possiedono la caratteristica (in precedenza individuati mediante diagnosi certa con gold standard) mentre la seconda esprime la proporzione dei negativi su tutti quelli che non possiedono la caratteristica (ancora dopo diagnosi certa). Tali parametri hanno significato se il valore è compreso tra 0,5 e 1, con 0,5 che corrisponde a fare diagnosi (presenza/assenza di malattia) lanciando una moneta non truccata, quindi a caso, mentre 1 indica l'assenza di errori diagnostici (nessun falso negativo o positivo). Ogni costruttore di test ambisce almeno idealmente a questa seconda condizione. Quello fin qui descritto sinteticamente corrisponde al c.d. "studio di I tipo". La relazione temporale tra diagnosi e test prevede che il secondo debba seguire la prima. In questo modo è lecito ottenere i parametri di sensibilità e specificità. Si tratta di studi "case-control" in cui a un numero di persone con malattia si appaia un gruppo equivalente (o maggiore) in termini di numerosità ma soprattutto per caratteristiche demografiche (esclusa quella oggetto di studio) cui si somministra il test da valutare. È evidente che in questo tipo di studi le prevalenze non sono rispettate dal momento che per effetto dell'appaiamento avremo quasi sempre un numero di "malati" e "sani" equivalente. Sono altrettanto importanti gli "studi di II tipo" che invertono l'ordine della valutazione: prima si somministra un test a un numero, generalmente grande, di persone provenienti dalla popolazione a rischio e solo dopo (e spesso a distanza di tempo) si procede alla diagnosi, tali studi hanno la tipica struttura delle indagini longitudinali che prevedono diverse fasi di follow-up nel corso del tempo. L'utilità di questi studi risiede nel fatto che il numero di malati (o con caratteristica) dovrebbe corrispondere alla vera prevalenza (sotto alcune condizioni di campionamento che devono essere rispettate, primariamente l'estrazione casuale del campione dalla popolazione). Il problema principale riguarda proprio il fatto che la numerosità dei malati potrebbe essere molto bassa (in funzione della reale prevalenza) inficiando la stabilità delle misure. Il vantaggio è la possibilità di calcolare altri due parametri che invece non hanno senso in studi di I tipo: il potere predittivo positivo (PPP) e negativo (PPN) che rappresentano rispettivamente la proporzione delle persone che possiedono la caratteristica tra tutte quelle risultate positive al test, e la proporzione delle persone che non possiedono la caratteristica tra tutte quelle negative al test. In buona sostanza si tratta di una stima della capacità predittiva del test. Questi sono gli indici più importanti per descrivere l'affidabilità di un test adottando procedure di ricerca differenti. Nella fattispecie degli obiettivi di questo lavoro, fare studi di I tipo

può risultare relativamente semplice: si istruisce una parte del gruppo di valutandi (anche variandone come si vuole la proporzione e quindi la prevalenza teorica) a simulare un certo profilo, mentre si dice all'altra parte del gruppo di rispondere al test in modo genuino. È evidente però che questa procedura per lavorare sulla simulazione non corrisponde esattamente a lavorare su campioni di persone che simulano "spontaneamente" e soprattutto con una motivazione forte, ma non sembrano esistere al momento sostituti validi a questa procedura. Su questi meccanismi e su questi indici di base si basano anche le c.d. statistiche bayesiane. Ci si riferisce alle statistiche bayesiane come ad una famiglia di tecniche basate sul teorema di Bayes (1):

$$p(A|B) = \frac{p(B|A)}{p(B)} \times p(A)$$

L'equazione esprime la probabilità condizionale a posteriori di un evento A (la causa, la condizione oggetto di studio) dato un evento B (l'effetto o il risultato al test) come il prodotto tra la probabilità a priori della causa (che corrisponde alla prevalenza del fenomeno) e il rapporto tra la probabilità dell'effetto data la presenza della causa e la probabilità dell'effetto indipendentemente dalla relazione con la causa. Il rapporto nell'equazione si può facilmente far corrispondere ai c.d. rapporti di verosimiglianza (RV anche detti fattori bayesiani). Tali rapporti esprimono il numero di volte per cui è più probabile (verosimile) che una diagnosi, positiva o negativa, provenga dal campione in cui la caratteristica è presente (il malingering nel nostro caso) piuttosto che dal campione in cui la caratteristica è assente (e.g. risposte genuine al test). I RV+ variano tra 1 e +∞ e gli RV- variano tra 0 e 1, ed è agevole dimostrare che il loro calcolo può essere semplificato mediante le formule che seguono:

$$RV+ = \text{Sensibilità} / (1 - \text{Specificità}) \quad (2)$$

$$RV- = (1 - \text{Sensibilità}) / \text{Specificità} \quad (3)$$

con 1 che rappresenta la condizione in cui una certa diagnosi si ottiene con medesima proporzione nei campioni in cui la caratteristica è presente e assente, rispettivamente (i.e. nessuna discriminazione). I RV moltiplicati per la prevalenza (o probabilità pre-test) di una caratteristica restituiscono la cosiddetta probabilità post-test che è il risultato del processo diagnostico e permette al valutatore di decidere se il risultato ottenuto dia sufficienti garanzie di sicurezza (fissando un criterio di accettazione, p.e.,  $p = 0,95$ ) per interrompere il processo diagnostico o piuttosto se sia necessario procedere con la somministrazione di un ulteriore test. In realtà non è possibile moltiplicare direttamente i RV per la prevalenza espressa in termini di probabilità, ma è necessario prima trasformare quest'ultima nel corrispondente ODDS, che si ottiene come rapporto tra la probabilità della presenza della caratteristica e la probabilità, reciproca, dell'assenza della caratteristica. La formula per trasformare una probabilità in ODDS è la seguente:

$$ODDS = \text{probabilità} / (1 - \text{probabilità}) \quad (4)$$

Una volta che si è proceduto con il calcolo del suddetto prodotto è possibile ottenere una probabilità a partire da un ODDS secondo la seguente formula inversa:

$$\text{Probabilità} = \text{ODDS} / (1 + \text{ODDS}) \quad (5)$$

In questo modo è possibile interpretare gli ODDS post test trasformandoli nelle più comuni probabilità. Questo processo ricorsivo parte da una probabilità pre-test, accerta mediante il test la positività/negatività del valutando alla caratteristica e quindi calcola una probabilità post-test. Si tratta del tipico processo circolare che il diagnosta è chiamato a terminare con un valore probabilistico in grado di suffragare la diagnosi e quindi la decisione. A questo punto del nostro discorso è chiaro come sia possibile affrontare la valutazione del malingering con un approccio molto simile. Ma è anche indispensabile individuare delle procedure basate su evidenze scientificamente fondate, che forniscano un sufficiente grado di sicurezza nella diagnosi e al contempo risultino “fair” nei confronti del valutando, in altri termini, che non si ostinino alla ricerca di una unica possibile conclusione: la “positività” del valutando.

A tal fine è necessario andare alla ricerca di strumenti che in più studi indipendenti abbiano dimostrato caratteristiche operative degne di questo delicato compito che spetta all'esperto di valutazione. Una volta individuati i migliori per il caso specifico è necessario selezionarne un numero ridotto, p.e. due, che chiameremo A e B, e decidere con quale approccio somministrarli: le strade da percorrere sono sostanzialmente due che, riferendosi alla logica booleana, risultano essere: (a) l'*approccio disgiunzione* basta che almeno uno dei due test sia positivo perché il valutando sia considerato positivo. Sul piano diagnostico questo significa dare maggiore importanza alla minimizzazione dei falsi negativi e (b) l'*approccio congiunzione*. È necessario che entrambi i test siano positivi affinché il valutando sia considerato positivo. Il secondo approccio rispetto al primo è orientato alla minimizzazione dei falsi positivi, che costituiscono un vero problema etico per il valutatore.

Infine è utile fornire una brevissima introduzione al metodo Monte Carlo che sarà alla base dei calcoli statistici per valutare la prestazione dei diversi approcci diagnostici nelle diverse condizioni a nostra disposizione. In generale il metodo Monte Carlo (per un'introduzione Mooney, 1997) si riferisce ad una famiglia molto ampia di tecniche che risolvono un problema deterministico mediante la generazione di numeri pseudo casuali, quindi attraverso un processo stocastico guidato da un algoritmo di generazione. In altri termini avendo a disposizione alcuni parametri di un problema derivati da studi su campioni reali, è possibile generare N campioni/unità statistiche (con N grande) in funzione di quei parametri cui sarà stata preventivamente applicata una fonte d'errore casuale (spesso una semplice distribuzione uniforme o normale). Mediante tale procedura è possibile studiare la distribuzione campionaria del parametro d'interesse, calcolare intervalli di confidenza non parametrici e di conseguenza confrontare le distribuzioni sotto le diverse condizioni imposte. Questo metodo risulta molto utile per stimare l'andamento stocastico delle distribuzioni sotto specifiche condizioni prima di intraprendere studi reali su persone.

In estrema sintesi l'obiettivo di questo studio è simulare

la distribuzione di variabili molto rilevanti per gli studi diagnostici: l'accuratezza diagnostica e in particolare la proporzione di falsi positivi in diverse condizioni di prevalenza bassa, media e alta (i.e. 0,2; 0,3; 0,4), ricorrendo a due approcci di somministrazione diversi e adottando due test per il malingering con eccellenti caratteristiche operative. La simulazione dovrebbe fornirci risposte su quale approccio adottare per individuare il malingering in funzione della prevalenza stimata. Si procederà inoltre a illustrare un caso specifico di applicazione della tecnica, passando dal contesto sperimentale simile agli studi “case-control”, al contesto tipico della clinica e della valutazione psicometrica e neuropsicologica caratterizzato dai c.d. studi “single-case”. Allo stato attuale non vi sono, per quanto in mia conoscenza, studi sistematici sul malingering che confrontino la prestazione dei due approcci, in termini di accuratezza diagnostica (cioè la media tra sensibilità e specificità), e in termini di proporzione di falsi positivi facendo ricorso a statistiche bayesiane e tecniche Monte Carlo.

### 3. Metodo

Materiali. Prima di introdurre sinteticamente gli strumenti sviluppati per individuare il comportamento di malingering, è utile fare una premessa. La combinazione tra approccio clinico (colloquio + osservazione) e psicometrico (test psicologici oggettivi) sembra in grado di fornire le massime garanzie di rigore scientifico e affidabilità nel dimostrare la validità della valutazione (Bianchi e La Mastra, 2005). È quindi improprio escludere dall'intervista gli strumenti oggettivi come improprio è credere che questi ultimi siano, da soli, in grado di fornire una valutazione dell'attendibilità del profilo del valutando. È quindi la diversificazione delle fonti che permette di avere maggiori garanzie di affidabilità del profilo stesso. Tenendo in mente tale principio generale noi qui ci occuperemo della sola parte testistica, nella convinzione che questa parte della valutazione meriti una più approfondita riflessione da parte di operatori delle diverse discipline e un contributo specifico della psicomatria nel cercare di individuare metodiche di calcolo in grado di sostenere le scelte del valutatore. Gli strumenti che ho scelto di considerare sono due eccellenti esempi di due famiglie diverse di strumenti: a) i questionari sui sintomi basati sulle dichiarazioni spontaneamente prodotte dai valutandi e dall'impressione clinica che ne deriva, e b) i compiti comportamentali che al contrario dei primi si basano esclusivamente sul livello e le caratteristiche della risposta comportamentale direttamente osservabile da parte del valutatore. La prima tipologia (che comprende anche M-FAST, SIRS, SADS) è rappresentata qui dal SIMS, acronimo di Structured Inventory of Malingered Symptomatology (versione italiana Widows & Smith, 2011) mentre la seconda tipologia (che comprende anche test a scelta forzata, Dot Counting Test, 15-item memorization test, Test di Barash sulle 15 Parole di Rey) ha uno dei suoi più diffusi esempi nel TOMM. Procederemo a descriverli brevemente entrambi.

Il SIMS è un questionario self-report che può essere utilizzato per acquisire dati di simulazione nella sfera dei disturbi neuropsichiatrici. È composto da 75 domande in formato dicotomico che hanno lo scopo di identificare

eventuali fenomeni di distorsione peggiorativa (Q), di psicosi (P) valutando la presenza di sintomi psicotici infrequenti, bizzarri o inusuali, di disturbi affettivi (AF) tra i quali sintomi depressivi e ansiosi inusuali. Ma quello che più ci interessa qui sono le scale di danno neurologico (NI) basato su checklist di sintomi neurologici impropri o inusuali e la scala di disturbi amnesici (AM) che si riferisce a disturbi di memoria incoerenti con il quadro clinico. Infine è disponibile anche una scala di bassa intelligenza (LI) che ha come obiettivo di smascherare deficit intellettivi esagerati attraverso prestazioni molto scadenti in domande di conoscenza / cultura generale. Esistono diversi studi sulla validità-attendibilità dello strumento condotti anche in Italia (Badiani, Girotto, Giannini, Gori, 2008), gli altri provengono per lo più dagli Stati Uniti e dall'Olanda (p.e. Poythress, Edens e Watkins, 2001; Heinze e Purisch, 2001; Lewis, Simcox e Berry, 2002; Merckelbach e Smith, 2003; Jelicic, Hessels e Merckelbach, 2006).

In questi studi si dimostra che il SIMS ha una buona coerenza interna con valori di Alpha di Cronbach compresi tra 0,82 e 0,88. Quello che è più importante per gli scopi di questo studio è che esistono diverse ricerche che hanno dimostrato la validità dello strumento nel discriminare persone che erano state istruite a simulare i disturbi da persone che invece rispondevano in modo genuino. In uno studio di questo tipo, Smith e Burger (1997) hanno dimostrato che i falsi negativi erano solo il 4% mentre i falsi positivi erano tre volte maggiori (12%). Quindi un risultato positivo al test era 8 volte più probabile nel campione dei simulatori rispetto al campione dei rispondenti onesti, mentre un risultato negativo era 22 volte più probabile nel gruppo dei rispondenti onesti rispetto al campione dei simulatori. Inoltre i risultati al SIMS erano congruenti con i risultati di altre scale di controllo in particolare con la scala F ( $r=0,84$ ) e con la scala F-K ( $r=0,81$ ) del test MMPI. In uno studio successivo di Merckelbach e Smith (2003) sempre con soggetti istruiti a simulare confrontati con rispondenti genuini sono stati ottenuti indici di sensibilità e specificità soddisfacenti con il 7% di falsi negativi e solo 2% di falsi positivi. Questo studio permette di concludere che un risultato positivo al test era circa 46 volte più probabile nel campione dei simulatori confrontato col campione dei rispondenti onesti e che un risultato negativo al test era circa 14 volte più probabile nel campione dei rispondenti onesti se confrontati con i simulatori. Nell'insieme il SIMS sembra uno strumento adeguato alla valutazione della simulazione in generale e anche nella simulazione di disturbi neurologici e nelle amnesie. Resta che la percentuale di falsi positivi facendo ricorso a questo strumento non sembra essere uguale a zero. In altri termini ci troveremmo comunque nella condizione di indicare come probabili simulatrici persone che genuinamente lamentano il disturbo per il quale sono in valutazione.

Il Test of Memory Malinger (ToMM, Tombaugh, 1996), basato su ricerche in neuropsicologia e psicologia cognitiva, è un test di riconoscimento visivo di 50 item costruito per aiutare a distinguere il malinger dai sintomi genuini dei disturbi della memoria in persone con almeno 16 anni di età che per caratteristiche tecniche lo inserisce nei cosiddetti Symptom Validity Test. Il ToMM consiste in due prove di apprendimento e una di ritenzione e fornisce

due punteggi di cut-off: 1) sotto il livello del caso, e 2) un criterio basato su popolazioni di pazienti con traumi cranici e decadimento cognitivo. Esistono studi (e.g. Teichner, & Wagner, 2004; Tombaugh, 1997; Greve, Bianchini, & Doane, 2006) di attendibilità e validità del test che prevedono il confronto tra persone cognitivamente intatte e diversi campioni clinici (decadimento cognitivo, afasia, traumi cranici e demenza). Inoltre, Powell, Gfeller, Hendricks, e Sharland (2004) hanno dimostrato un livello generale di accuratezza diagnostica pari al 96% per lo strumento confrontando diversi gruppi di soggetti. Mentre, secondo Bianchi e La Mastra (2005, p. 316) la sensibilità e la specificità del test dovrebbero attestarsi a 0,91 e 0,95, rispettivamente per un cut-off di 45 su 50.

Nella prima prova il valutatore mostrerà 50 immagini, una alla volta, per circa tre/quattro secondi ognuna, mostrerà subito dopo 50 coppie di immagini una delle quali appartiene al set di stimoli appresi in precedenza. Nella seconda prova i 50 stimoli sono ripresentati in un nuovo ordine e nuovamente il valutatore mostrerà 50 coppie di figure, con una immagine dell'iniziale set di stimoli più una immagine nuova estratta da un nuovo set di distrattori. La stessa procedura vale per la prova di ritenzione differita senza la fase di apprendimento. In tutte le prove il compito del valutando è indicare l'immagine proveniente dall'iniziale set di stimoli. Ogni risposta corretta equivale ad un punto con un massimo di 50 per ogni prova. Punteggi inferiori a 45 costituiscono già una indicazione di malinger.

Procedura. Come già chiarito nel paragrafo introduttivo, il metodo Monte Carlo consiste nel generare unità statistiche che hanno una componente deterministica (quella legata ai parametri di simulazione le cui condizioni di partenza sono indicate nella tabella 1) e una componente stocastica (garantita da appositi generatori di funzioni pseudo-casuali, ad alta efficienza). In questo caso sono state simulate le frequenze di 1000 tavole diagnostiche di contingenza  $2 \times 2$  (veri e falsi positivi e veri e falsi negativi) di 100 casi ognuna (dimensione campionaria mantenuta costante, relativamente alta e compatibile con studi simili su soggetti reali). Complessivamente sono state effettuate 100.000 estrazioni indipendenti per ognuna delle condizioni generate dalle due variabili variate sistematicamente: la prevalenza prevista, a tre livelli (i.e.  $p = 0,2; 0,3; 0,4$ ) e l'approccio di valutazione, a due livelli (*disgiunzione, congiunzione*) ottenendo sei condizioni differenti. Due parametri sono invece rimasti fissi: la sensibilità e la specificità del SIMS (0,93 e 0,98, rispettivamente secondo Merckelbach e Smith, 2003), la sensibilità e la specificità del ToMM (0,91 e 0,95 rispettivamente secondo Bianchi & La Mastra, 2005, p. 316). Entrambi gli strumenti sono in grado di predire il malinger grazie alle loro ottime caratteristiche operative (tutti i valori sono superiori a 0,9). Il ciclo di estrazione è estremamente semplice:

- Il sistema costruisce un vettore E di possibili esiti (nel nostro caso gli esiti sono quattro: vero positivo, falso negativo, vero negativo, falso positivo) la cui frequenza corrisponde ad una delle distribuzioni di probabilità della Tabella 1.
- Ad ogni valore del vettore E è associato un valore di un vettore C di numeri casuali con distribuzione uniforme

- (ogni valore, compreso tra 0 e 1, ha eguale probabilità di essere estratto).
- I due vettori vengono ordinati secondo il vettore C, il sistema quindi estrae il primo valore del vettore E e lo inserisce nel vettore definitivo di ampiezza campionaria n.
  - I punti 2) e 3) si ripetono fin quando non è saturata una matrice n x N, dove n è l'ampiezza campionaria prevista

- (nel nostro caso n = 100) e N è il numero di campioni previsto (nel nostro caso N = 1000). Tale matrice prevede pertanto 100.000 unità casuali.
- Il processo dal punto 1) al 4) si ripete per tutte le distribuzioni di probabilità a disposizione (sei nel nostro caso). L'intera procedura è stata realizzata mediante funzioni customizzate di MS EXCEL.

			Esiti possibili			Prevalenza		
						0,2	0,3	0,4
A+U B+	Simulatori	Veri positivi	A+B+	A+B-	A-B+	0,19874	0,29811	0,39748
		Falsi Negativi	A-B-			0,00126	0,00189	0,00252
	Genuini	Veri negativi	A-B-			0,74480	0,65170	0,55860
		Falsi positivi	A+B+	A+B-	A-B+	0,05520	0,04830	0,04140
A+∩ B+	Simulatori	Veri positivi	A+B+			0,16926	0,25389	0,33852
		Falsi Negativi	A+B-	A-B+	A-B-	0,03074	0,04611	0,06148
	Genuini	Veri negativi	A+B-	A-B+	A-B-	0,79920	0,69930	0,59940
		Falsi positivi	A+B+			0,00080	0,00070	0,00060

Tab. 1: Le sei serie di probabilità di partenza per lo studio Monte Carlo in funzione di approccio di somministrazione, dei tre livelli di prevalenza con valori di sensibilità e specificità dei due test tenuti costanti. Sono pure indicati i possibili esiti delle somministrazioni dei due test, in funzione delle caratteristiche dei valutandi "Simulatori" e "Genuini".

#### 4. Risultati e discussione

Di seguito sono riportati gli esiti della simulazione Monte Carlo per i due approcci diagnostici in funzione della capacità di massimizzare l'accuratezza diagnostica e minimizzare i falsi positivi (i.e. positivo se almeno uno dei due test è positivo: *approccio disgiunzione*, positivo se entrambi i test sono simultaneamente positivi: *approccio congiunzione*) e per tre livelli di prevalenza di simulazione per  $p = 0,2; 0,3; 0,4$ .

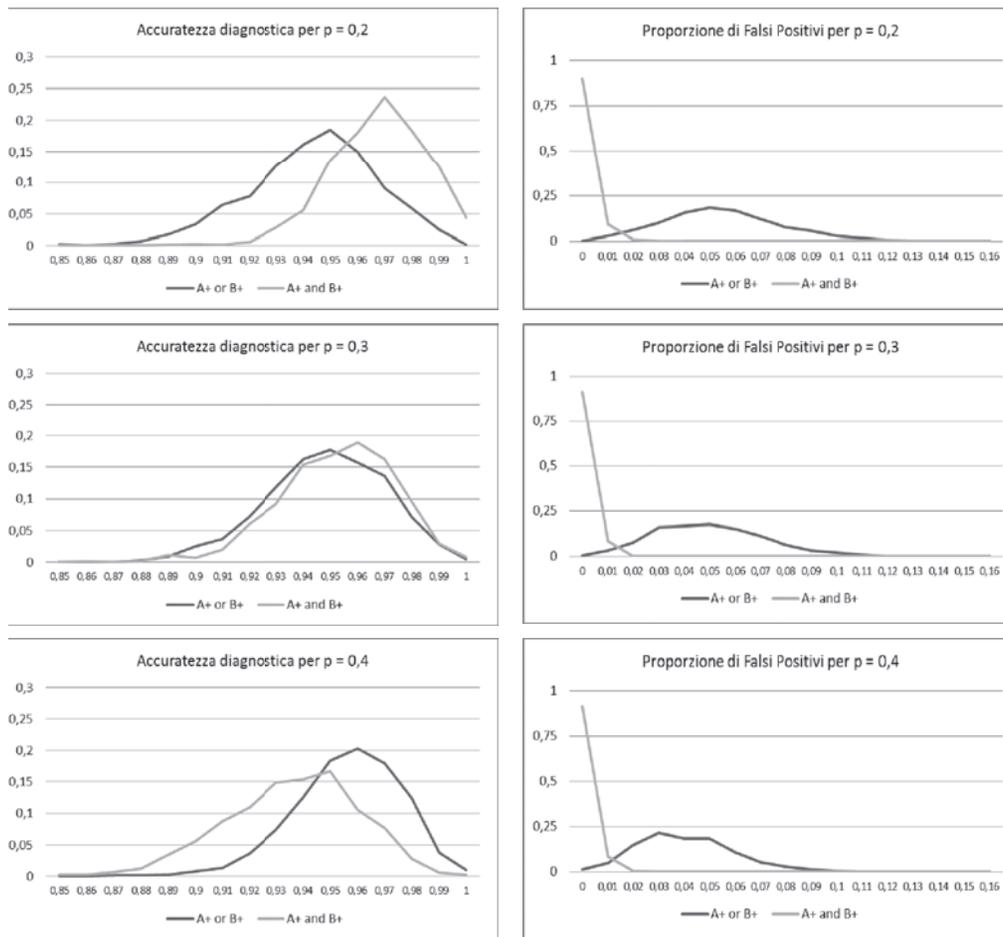
In Figura 1 sono riportate le rappresentazioni grafiche delle distribuzioni relative ai due parametri delle due misure di outcome: l'accuratezza diagnostica (i.e. (veri positivi + veri negativi)/dimensione campionaria, prima colonna), e la proporzione di falsi positivi nei campioni estratti (i.e. proporzione di persone autenticamente compromesse scambiate per simulatrici, seconda colonna). In Tabella 2 sono riportati i parametri di media, deviazione standard, l'intervallo di confidenza non parametrico del 90% e il risultato del confronto tra i due approcci nelle tre condizioni di prevalenza mediante una serie di U di Mann-Whitney per il confronto delle distribuzioni. Da tutti i test emergono risultati di differenza significativa delle distribuzioni, ma vediamo nel dettaglio in quale direzione.

Accuratezza diagnostica. Dallo studio della prima variabile emerge che nella simulazione con prevalenza bassa (0,2) l'approccio congiunzione ha una media più elevata sebbene la differenza tra le due distribuzioni, rilevata dal test U di Mann-Whitney non si rifletta in una differenza anche tra gli intervalli di confidenza che appaiono parzialmente sovrapposti. Lo stesso vale per il livello di prevalenza medio

(0,3), in cui il grado di sovrapposizione tra le distribuzioni dei due approcci è ancora superiore. Lo stesso accade, ma a parti invertite, per il livello alto di prevalenza (0,4) in cui la media della distribuzione calcolata per l'approccio disgiunzione è più elevata di quella relativa all'approccio congiunzione. Quindi, la variabile prevalenza ha una sua importanza nel problema diagnostico presentato. Quando la prevalenza è relativamente bassa (per il fenomeno del malingering) diciamo al 20%, l'approccio che richiede la positività di entrambi i test ha un'efficienza diagnostica migliore dell'approccio disgiunzione (i.e. quando basta un solo test positivo per diagnosticare il malingering). I due approcci sostanzialmente pareggiano intorno alla prevalenza del 30%, poi diventa più efficace il metodo disgiunzione quando la prevalenza è intorno al 40%.

Proporzione di falsi positivi. Per quanto riguarda la seconda variabile di outcome, la proporzione di falsi positivi, al contrario, non sussistono dubbi. Infatti, alla significatività statistica dei test U di Mann-Whitney corrispondono anche risultati molto chiari dal punto di vista degli intervalli di confidenza che non risultano mai sovrapposti tra i due approcci a vantaggio di quello congiunzione per tutti i livelli di prevalenza. L'approccio congiunzione mostra proporzioni medie di falsi positivi molto basse, intorno a uno su 1000, contro 5 su 100 dell'approccio disgiunzione. Inoltre il primo approccio mostra un valore pari a 1 su 100 contro 10 su 100 del secondo approccio, in corrispondenza del 95° percentile delle rispettive distribuzioni.

Da questo studio Monte Carlo quindi abbiamo alcune risposte importanti: è utile procedere a studi di prevalenza



**Fig. 1: Andamenti delle distribuzioni di accuratezza diagnostica e proporzione di falsi positivi per i campioni composti mediante metodo Monte Carlo per due approcci di somministrazione e tre livelli di prevalenza.**

del malingering in contesti reali per chiarire le effettive dimensioni di questo fenomeno. Ne costituisce un ottimo esempio il lavoro di Conson, Sansone, Ferrantino, Duretto e Crisci (2014) apparso su questa rivista i quali hanno confrontato persone con danni cerebrali provenienti dal contesto medico-legale – quindi ad alto rischio di malingering – con pazienti provenienti da contesti di cura che gli autori considerano a basso rischio di malingering e discutono le implicazioni connesse alla selezione di strumenti efficaci per la misurazione del danno neuropsicologico e alla simultanea esigenza di difendersi dal malingering.

Il presente studio invece si occupa di chiarire come procedendo agli accertamenti con strumenti dotati di eccellenti caratteristiche operative – mai inferiori al 90% – anche adottando differenti approcci e per differenti prevalenze, la qualità diagnostica resta adeguata. Inoltre dal presente studio emerge che uno dei due approcci presi in esame consente quasi di azzerare il rischio di falsi positivi: l’approccio congiunzione. Questo significa che, presi insieme, i risultati dello studio Monte Carlo convergono nel farci preferire l’approccio congiunzione perché minimizza il rischio di falsi positivi e risulta adeguato fino a prevalenze intorno a 0,30, per diventare leggermente meno efficace con prevalenze intorno a 0,4, pur conservando livelli di accuratezza del tutto adeguati.

Prevalenza	Approccio	Accuratezza diagnostica				Test U di Mann-Whitney p
		media	dev. St.	5%	95%	
p=0,2	$A+ \cup B+$	0,944	0,024	0,900	0,980	< 0,001
	$A+ \cap B+$	0,968	0,017	0,940	0,990	
p=0,3	$A+ \cup B+$	0,948	0,022	0,910	0,980	< 0,001
	$A+ \cap B+$	0,952	0,021	0,920	0,980	
p=0,4	$A+ \cup B+$	0,956	0,020	0,920	0,980	< 0,001
	$A+ \cap B+$	0,936	0,025	0,890	0,970	
		Proporzione di falsi positivi (per singolo campione)				
p=0,2	$A+ \cup B+$	0,0550	0,0240	0,0200	0,1000	< 0,001
	$A+ \cap B+$	0,0010	0,0030	0,0000	0,0100	
p=0,3	$A+ \cup B+$	0,0500	0,0220	0,0200	0,0900	< 0,001
	$A+ \cap B+$	0,0010	0,0030	0,0000	0,0100	
p=0,4	$A+ \cup B+$	0,0410	0,0190	0,0100	0,0700	< 0,001
	$A+ \cap B+$	0,0010	0,0030	0,0000	0,0100	

**Tab. 2: Parametri di media, deviazione standard, intervallo di confidenza non parametrico del 90%, esito dei Test U di Mann-Whitney per il confronto delle distribuzioni in funzione dei due approcci di somministrazione disgiunzione e congiunzione, rispettivamente.**

## 5. Statistiche bayesiane e valutazione di malingering nel caso singolo

Come abbiamo detto sopra, lo studio Monte Carlo ci fornisce un quadro d'insieme sull'efficienza dei nostri parametri sotto diverse condizioni di utilizzo e ci ha aiutato a comprendere come ottimizzare i nostri sforzi per ottenere una stima giusta della probabilità di malingering nel contesto della valutazione forense. Nella pratica come dovrebbe comportarsi il valutatore e come può procedere per ottenere una stima probabilistica del rischio di malingering adottando i due strumenti fin qui considerati e dovendo valutare un solo individuo? Può ricorrere alla procedura di calcolo della probabilità a posteriori desunta dal teorema di Bayes e descritta nell'introduzione, che qui tratteggerò sinteticamente immaginando di dover procedere alla valutazione di un caso specifico.

Immaginiamo di dover procedere alla valutazione del Sig. A.D.G. che lamenta un'amnesia anterograda (disturbo della memoria per gli eventi successivi al trauma, in pratica un disturbo nei processi di apprendimento) post esposizione a monossido di carbonio per cause professionali con perdita di conoscenza per un periodo lungo, ipossia cerebrale, ricovero e degenza. Dopo il recupero dall'ipossia che non lascia chiari segni all'esame oggettivo, il valutando intenta una causa in cui l'amnesia costituisce l'elemento fondante. La prevalenza di malingering in questa casistica è pari a 29,49% secondo lo studio di Mittenberg et al. (2002). Fissiamo il nostro livello di sicurezza al conservativo valore di 99%, quindi concluderemo che il valutando è positivo al malingering se entrambi in test che intendiamo somministrare risulteranno positivi e la probabilità post-test alla fine del percorso avrà raggiunto il 99%.

Somministriamo prima il SIMS che è il nostro miglior strumento dal punto di vista delle caratteristiche operative. Il test restituisce un risultato positivo.

In primo luogo, applicando la (2) calcoliamo il Rapporto di Verosimiglianza del test mediante valori di sensibilità e specificità in nostro possesso:

$$RV+ = \text{sensibilità}/(1-\text{specificità}) = 0,93/(1-0,98) = 46,5$$

Otteniamo un valore davvero molto alto di RV+ che ci dice come un risultato positivo al test è circa 46 volte più probabile nel gruppo dei simulatori piuttosto che nel gruppo dei rispondenti genuini. Quindi, calcoliamo l'Odds della prevalenza, applicando la (4):

$$\text{Odds}(M+) = P(M+)/[1-P(M+)] = 0,2949/0,7051 = 0,4182$$

A questo punto non ci resta che applicare la (1) avendo cura di sostituire al rapporto presente nella formula il nostro RV+ appena calcolato:

$$\text{Odds}(M+ | T+) = 46,5 \times 0,4182 = 19,446$$

Il valore così ottenuto rappresenta l'odds post-test. Per semplicità d'interpretazione e adottando la (5) convertiremo nuovamente l'odds nella corrispondente probabilità, sebbene nel processo diagnostico ricorsivo che prevede la som-

ministrazione di un secondo test, il valore che utilizzeremo sarà ancora l'odds appena calcolato.

$$P(M+ | T+) = \text{Odds}(M+ | T+)/[\text{Odds}(M+ | T+)+1] = 19,446 / 20,446 = 0,951$$

Questo risultato significa che dopo la somministrazione del SIMS il valutando è passato dalla probabilità pre-test di meno del 30% a più del 95%. Già questo valore potrebbe risultare sufficiente per affermare che il malingering è *probabile*. Come si ricorderà il criterio fissato all'inizio del processo era 99% quindi procediamo con la seconda somministrazione che qualora risultasse negativa determinerebbe la riduzione della probabilità post-test e di conseguenza la diagnosi di disturbo genuino.

Somministriamo il ToMM con i ben noti valori di sensibilità e specificità già utilizzati in precedenza e associati ad un cut-off di 45 su 50. Il Sig. A.D.G. ottiene nei trial 1, 2 e nel ricordo ritardato rispettivamente: 31, 34, 37 sempre ben al di sotto del cut-off di 45. Risulta pertanto positivo anche al secondo test. Vogliamo a questo punto procedere con il calcolo della probabilità post-test dopo la somministrazione del ToMM e per fare questo non partiremo più dalla prevalenza giacché la nostra probabilità a priori sarà quella calcolata grazie alla somministrazione del SIMS. Procediamo al calcolo del RV+ del ToMM:

$$RV+ = \text{sensibilità}/(1-\text{specificità}) = 0,91/(1-0,95) = 18,2$$

Calcoliamo ora l'Odds post-test moltiplicando il RV+ del ToMM con l'Odds post-test della precedente somministrazione:

$$\text{Odds}(M+ | T+) = 18,2 \times 19,446 = 353,917$$

Non ci resta che convertire per l'ultima volta l'Odds post-test nella corrispondente probabilità post-test per poi procedere con la decisione diagnostica.

$$P(M+ | T+) = \text{Odds}(M+ | T+)/[\text{Odds}(M+ | T+)+1] = 353,917 / 354,917 = 0,997$$

Il valore di probabilità supera ampiamente il criterio del 99%. Il processo diagnostico si può chiudere con una valutazione che potrebbe essere di *malingering definito*. Esiste l'incentivo a simulare, la descrizione dei sintomi non è del tutto convincente (SIMS positivo), la prestazione comportamentale al test di memoria ToMM è peggiore di quella di persone effettivamente compromesse cognitivamente, la probabilità di simulare è altissima. Una coesistenza di indici di questo tipo depone decisamente a favore della simulazione, con probabilità di falso positivo vicina allo zero.

## 6. Considerazioni conclusive

Intendo concludere questo studio con alcune considerazioni sul tema della valutazione del malingering.

Caratteristiche operative adeguate. Il processo di valutazione del malingering è un percorso ad ostacoli in cui il valutatore si trova nella difficoltà di dover decidere l'affida-

bilità del valutando e di doverne dare conto ad un decisore che intraprenderà delle iniziative con effetti sull'esistenza dello stesso. Pertanto è davvero molto importante che: (a) il processo diagnostico sia portato a termine da esperti in grado di valutare in modo scientificamente rigoroso gli esiti della valutazione, (b) gli strumenti adottati per la valutazione del malingering posseggano indici di specificità e sensibilità utili per il caso in esame e alti, superiori a 0,9. In caso contrario non dovrebbero essere utilizzati per questi scopi, e (c) è bene verificare che le caratteristiche operative siano elevate ma che in particolare vi sia elevata specificità giacché i costi etici per i falsi positivi di malingering (i.e. persone genuinamente compromesse scambiate per simulatrici) sono decisamente alti se si tiene conto che la simulazione è un reato penalmente perseguibile. Anche nei casi in cui vi siano pochi dubbi sulla diagnosi di malingering le conseguenze sociali, psicologiche e penali di una simile accusa non possono essere sottovalutate.

Valutazione "fair". Gli strumenti utilizzati sono tutti costruiti per individuare la simulazione. Essi possiedono rapporti di verosimiglianza positivi che fanno, di solito, aumentare di molto la probabilità a posteriori di simulazione anche perché le probabilità a priori (le prevalenze) sono molto alte per questo fenomeno. Pertanto se utilizzassimo un numero elevato di strumenti otterremmo quasi sicuramente risultati che suggeriscono la presenza di malingering. Un approccio che adotta molti strumenti sulla simulazione non è, a mio avviso, "fair" nei confronti del valutando. Sarebbe quindi corretto interrompere la procedura dopo aver utilizzato non più di due strumenti con alta sensibilità e specificità così come suggerito nel presente studio.

Il paradosso del "malato non troppo malato". La collaborazione sulla messa a punto di strategie per la valutazione del malingering tra approccio medico-legale, psicométrico, neuropsicologico sarà indispensabile in futuro per risolvere un paradosso epistemologico a mio avviso importante che pretende di considerare genuinamente malato soltanto colui i cui sintomi non siano tali da farlo sembrare *troppo malato*. È evidente che il limite sottile tra effettiva gravità dei sintomi e la loro esagerazione porta le nostre discipline a muoversi su un terreno paludoso che richiederà in futuro il ricorso ad equipaggiamenti sempre più peculiari, raffinati e genuinamente multidisciplinari.

## Bibliografia

- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.
- Badiani, S., Giroto, G., Giannini, M., & Gori, A. (2008). Measuring the malingering: psychometric properties and discriminant ability of the Italian version of the structured inventory of malingered symptomatology (SIMS). *Italian Journal of Criminology*, 2(1), 163-178.
- Bianchi, A. & La Mastra, C. (2005). L'esame neuropsicologico forense: una visione d'insieme. In A. Bianchi (Ed.), *La valutazione neuropsicologica del danno psichico ed esistenziale* (pp. 279-318). Padova: Cedam.
- Conson, M., Sansone, N., Ferrantino, T., Duretto, G., & Crisci, A. (2014). Neuropsychological assessment of feigned cognitive symptoms in brain-damaged patients. *Italian Journal of Criminology*, 8(1), 12-17.
- Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1176-1190.
- Heinze, M. C., & Purisch, A. D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology Practice*, 1(4), 23-52.
- Jelicic, M., Hessels, A., & Merckelbach, H. (2006). Detection of feigned psychosis with the Structured Inventory of Malingered Symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavioral Assessment*, 28(1), 19-22.
- Lauriola, M. (2007). Validità e uso diagnostico dei test. In A.P. Ercolani (Ed.), *Strumenti statistici per la ricerca, la valutazione e la diagnosi in psicologia* (pp. 71-144). Milano: Raffaello Cortina.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, 14(2), 170.
- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, 18(2), 145-152.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of clinical and experimental neuropsychology*, 24(8), 1094-1102.
- Mooney, C. Z. (Ed.). (1997). *Monte Carlo simulation* (No. 116). Sage.
- Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, 19(5), 693-702.
- Poythress, N. G., Edens, J. F., & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior*, 25(6), 567.
- Resnick, P.J. (1988). Malingering of post-traumatic disorders. In R. Rogers (ed.), *Clinical Assessment of Malingering and Deception* (pp. 84-103). New York: Guilford.
- Resnick, P.J. (1992). *The detection of malingered mental illness*. Unpublished manuscript. Presented at the American Psychology-Law Society, Division 41 Mid-Year Conference, San Diego, California.
- Rogers, R. (1988). Introduction. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception*, (pp. 1-9). New York: Guilford.
- Slick, D.J., Sherman, E.M., & Iverson, G.L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545-561.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online*, 25(2), 183-189.
- Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, 19(3), 455-464.
- Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. North Tonawanda, NY: Multi-Health Systems.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260.
- Widows M.R., Smith G.P. (2011). *SIMS Structured Inventory of Malingered Symptomatology - Versione italiana di La Marca, Rigoni, Sartori & Lo Priore*. Firenze: Giunti O.S.