

RESEARCH ARTICLE

Prevalence of symptom exaggeration among North American independent medical evaluation examinees: A systematic review of observational studies

Andrea J. Darzi^{1,2,3}, Li Wang^{2,3,4}, John J. Riva¹, Rami Z. Morsi⁵, Rana Charide¹, Rachel J. Couban^{1,6}, Samer G. Karam¹, Kian Torabiardakani^{1,2,3}, Annie Lok³, Shanal Ebrahim¹, Sheena Bance⁶, Regina Kunz^{1,7}, Gordon H. Guyatt¹, Jason W. Busse^{1,2,3,4*}

1 Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada, **2** Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, ON, Canada, **3** Department of Anesthesia, McMaster University, Hamilton, ON, Canada, **4** Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, ON, Canada, **5** Department of Neurology, University of Chicago, Chicago, Illinois, United States of America, **6** Centre for Addiction and Mental Health, Toronto, ON, Canada, **7** Division of Clinical Epidemiology, Evidence-based Insurance Medicine, University Hospital Basel, Basel, Switzerland

* bussejw@mcmaster.ca



OPEN ACCESS

Citation: Darzi AJ, Wang L, Riva JJ, Morsi RZ, Charide R, Couban RJ, et al. (2025) Prevalence of symptom exaggeration among North American independent medical evaluation examinees: A systematic review of observational studies. PLoS One 20(6): e0324684. <https://doi.org/10.1371/journal.pone.0324684>

Editor: Thiago P. Fernandes, Federal University of Paraíba, BRAZIL

Received: July 31, 2024

Accepted: April 28, 2025

Published: June 25, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0324684>

Copyright: © 2025 Darzi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/),

Abstract

Background

Independent medical evaluations (IMEs) are commonly acquired to provide an assessment of impairment; however, these assessments show poor inter-rater reliability. One potential contributor is symptom exaggeration by patients, who may feel pressure to emphasize their level of impairment to qualify for incentives. This study explored the prevalence of symptom exaggeration among IME examinees in North America, which if common may represent an important consideration for improving the reliability of IMEs.

Methods

We searched CINAHL, EMBASE, MEDLINE and PsycINFO from inception to July 08, 2024. We included observational studies that used a known-group design or multi-modal determination method. Paired reviewers independently assessed risk of bias and extracted data. We performed a random-effects model meta-analysis to estimate the overall prevalence of symptom exaggeration and explored potential subgroup effects for sex, age, education, clinical condition, and confidence in the reference standard. We used the GRADE approach to assess the certainty of evidence.

which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Results

We included 44 studies with 46 cohorts and 9,794 patients. The median of the mean age was 40 (interquartile range [IQR] 38–42). Most cohorts included patients with traumatic brain injuries (n=31, 67%) or chronic pain (n=11, 24%). Prevalence of symptom exaggeration across studies ranged from 17% to 67%. We found low certainty evidence suggesting that studies with a greater proportion of women ($\geq 40\%$) may be associated with higher rates of exaggeration (47%, 95%CI 36–58) vs. studies with a lower proportion of women ($< 40\%$) (31%, 95%CI 28–35; test of interaction $p=0.02$). Possible explanations include biological differences, greater bodily awareness, or higher rates of negative affectivity. We found no significant subgroup effects for type of clinical condition, confidence in the reference standard, age, or education.

Conclusion

Symptom exaggeration may occur in almost 50% of women and in approximately a third of men undergoing IMEs. The high prevalence of symptom exaggeration among IME attendees provides a compelling rationale for clinical evaluators to formally explore this issue. Future research should establish the reliability and validity of evaluation criteria for symptom exaggeration and develop a structured IME assessment approach.

Background

In 2022, Statistics Canada found that 8.0 million Canadian adults reported a disability [1] and in 2020, 64.4 million Americans reported living with disability [2]. Individuals suffering from a disabling injury or illness may be eligible to receive financial compensation and services based on their level of impairment. Determinations of impairment often rely on independent medical evaluations (IMEs), which are requested by a third party, such as an insurance company or employer, and conducted by a clinician who is not part of the patient's regular medical team [3]. Underlying this process is the concern that treating clinicians may have difficulty providing impartial assessments of their patients [4,5]. Such concerns are supported by a trial that randomized 5,888 individuals in Norway to an independent assessment or usual care and found 29% of IMEs recommended less sick leave than the treating physician (68% the same, and 3% a longer duration) [6].

Despite their widespread use and far-reaching consequences, the consistency and reliability of IMEs has been challenged. The most recent systematic review found that clinical experts assessing the same patients often dissented on whether they were disabled from working (median inter-rater reliability 0.45) [7]. Although this review suggested that standardization of the assessment process may improve the reliability of IMEs, [7] two subsequent studies failed to support this hypothesis [8]. Another potential source of variability in IME assessments is symptom exaggeration [3]. IME assessors may focus too narrowly on a biomedical model to explain symptoms,

without giving sufficient attention to psychosocial and work-related factors that may influence how individuals present their symptoms [3,9].

Patients referred for IMEs often present with subjective complaints (e.g., mental illness, chronic pain) and may feel pressure to emphasize their level of impairment to qualify for wage replacement benefits, receiving time off work, or other incentives [3,10,11]. Patients' presentation may also be affected if they perceive the assessor as representing the referring agency rather than their interests. Whether or not IME assessors consider symptom exaggeration has the potential to lead to very different conclusions; however, the prevalence of exaggeration among IME attendees is uncertain and individual studies report rates as low as 17% [12] or as high as 67% [13]. Also, terminology such as exaggeration, malingering, or over-reporting are defined inconsistently across studies, making it difficult to distinguish intentional deception from psychological amplification of distress [4,14]. We undertook the first systematic review of observational studies to explore the prevalence of symptom exaggeration among IME examinees in North America.

Methods

We conducted our systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Meta-analysis of Observational Studies in Epidemiology (MOOSE) checklists [15,16]. (See [S1](#) and [S2 Checklists](#) in the supplemental material) We registered our protocol on the Open Science Framework (Registration DOI: <https://doi.org/10.17605/OSF.IO/64V2B>) [17]. After registration but prior to data analysis, we included five meta-regressions/subgroup analyses to explore variability among studies reporting the prevalence of symptom exaggeration: (1) proportion of female participants, (2) older age, (3) level of formal education, (4) clinical condition, and (5) level of confidence in the reference standard used in the approach for evaluating symptom exaggeration.

Data sources and searches

An experienced medical librarian (RJC) developed database-specific search strategies ([S1 Table](#)) and conducted a systematic search in CINAHL, EMBASE, MEDLINE and PsycINFO, from inception through July 08, 2024. We included English, French or Spanish studies to reduce language bias. The search strategies were developed using a validation set of known relevant articles and included a combination of MeSH headings and free text key words, such as malingering* or litigation or litigant or "insufficient effort" and "Independent medical examination" or "independent medical evaluation" or "disability" or "classification accuracy". We did not use any filters for our searches to maximize sensitivity. We screened the reference lists of all included studies for additional eligible articles.

Study selection

Six reviewers screened the titles and abstracts of all retrieved citations, independently and in duplicate, and subsequently the full texts of potentially eligible studies, using standardized and pre-tested forms [18]. A third senior reviewer resolved disagreements when necessary.

Eligible studies: (i) enrolled individuals presenting for an IME in North America, (ii) in the presence of external incentive (e.g., insurance claims), and (iii) assessed the prevalence of symptom exaggeration using a known group design or multimodal determination method [19,20]. As there is no singular reliable and valid criteria (reference standard) in the literature that is used to assess for symptom exaggeration, we included known group study designs that defined their reference standard based on criteria incorporating both clinical findings and performance on psychometric testing to classify individuals as exaggerating (within diagnostic test terminology, the target positive group), or not exaggerating (the target negative group) their symptoms [21,22].

Examples of two commonly used known group designs are the Slick, Sherman, and Iverson criteria for malingered neurocognitive dysfunction [23] and the Bianchini, Greve, & Glynn criteria for malingered pain-related disability [24]. We excluded studies that used only beyond-chance scores on symptom validity tests as an indicator of symptom

exaggeration, since beyond-chance scores are infrequent and likely to result in underestimates [25–27]. We restricted our focus to North America as there may be important differences between IMEs conducted within North America where social insurance for disability is limited and Europe where social insurance is prominent. In cases where multiple studies had population overlap, we included only the study with the larger sample size.

Data extraction and risk of bias assessment

Teams of paired reviewers abstracted data independently and in duplicate from all eligible studies using standardized, pre-tested forms. We prefaced data abstraction with calibration exercises to optimize consistency and accuracy of extractions. For all identified studies, the reviewers abstracted the following data: name of first author, year of publication, participant demographics, referral source(s), criteria for establishing symptom exaggeration and reference standard, and the prevalence of symptom exaggeration. After completing training and calibration, pairs of reviewers independently evaluated risk of bias for each included study. They used key criteria tailored to known-group designs, which were developed and pre-tested in collaboration with research methodologists. These criteria included: (i) representativeness of the study population, (ii) validity of outcome assessment (including whether the index test was administered without knowledge of the reference standard, and confidence in the reference standard), (iii) whether those with and without symptom exaggeration were similar across age groups and education level, and (iv) loss to follow-up ($\geq 20\%$ was considered high risk of bias). The response options for all the above risk of bias items included “definitely yes”, “probably yes”, “probably no” and “definitely no”. Also, we evaluated whether the criteria for establishing symptom exaggeration had been shown reliable and valid. We resolved disagreements by consensus or with the help of a third senior reviewer.

We categorized the reference standard and rated our confidence in it as either: (i) ‘weak’ when the study declared a known-group design, however its only criterion for identifying symptom exaggeration was below-chance performance on forced-choice symptom validity testing without any corroborating clinical observations or inconsistencies in medical records. For example, a patient with a mild ankle sprain labeled as exaggerating exclusively because they failed a below-chance forced-choice test of pain threshold, with no clinical exam or review of documented pain or functional abilities; (ii) ‘moderate’ where most patients exaggerating symptoms were identified by forced symptom validity testing results, but some cases could be confirmed using other credible indicators. For example, a claimant insists they cannot remember simple details of their daily routine (e.g., the route to their kitchen), yet is casually observed navigating complex tasks with no apparent cognitive difficulty; or (iii) ‘strong’ where exaggeration was determined by either forced symptom validity testing results or other credible clinical evidence. For example, a clinical finding that would classify a patient presenting with persistent post-concussive complaints after a very mild head injury as exaggerating symptoms would include claims of remote memory loss (e.g., loss of spelling ability).

Data synthesis and analysis and certainty in the evidence assessment

We used a random-effects model to pool data for the prevalence of symptom exaggeration among IME examinees and a Freeman-Tukey double arcsine transformation to stabilize the variance [28,29]. This transformation avoids producing confidence intervals (CIs) that include values lower than 0% or greater than 100% [28,29]. We used the DerSimonian and Laird method [30] to pool estimates of symptom exaggeration based on the transformed values and their variances, and then the harmonic mean of sample sizes for back-transformation to the original units of proportions [31].

We assessed the certainty of evidence based on the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [32]. This approach considers risk of bias, indirectness, inconsistency, imprecision, and small study effects, to appraise the overall certainty of evidence as high, moderate, low, or very low [32]. We estimated that if 20% of IME attendees presented with symptom exaggeration, that would be sufficiently frequent to justify formal evaluation for exaggeration by IME evaluators. Therefore, we rated down for imprecision if the 95%CI associated with the

prevalence of symptom exaggeration included 20%. When there were at least 10 studies contributing to meta-analysis, we evaluated small study effects by visual inspection of the funnel plot for asymmetry and calculation of Egger's test [33].

Subgroup analyses, meta-regression, and sensitivity analyses

We assessed heterogeneity across studies contributing to our pooled estimate of symptom exaggeration using both a statistical test and visual inspection of forest plots. We did not calculate I^2 as it can be misleading in cases where the estimates of precision are very narrow due to large sample sizes. Instead, we estimated the between-study variance with tau-squared (τ^2), which provides an absolute measure of heterogeneity. We considered $\tau^2 < 0.05$ as low, between 0.05–0.1 as moderate, and > 0.1 as substantial heterogeneity [34].

We assessed the variability between studies based on five hypotheses. We assumed a higher prevalence of symptom exaggeration with: (1) greater strength of the reference standard, (2) higher proportion of female participants, (3) older age, (4) lower level of formal education, and (5) higher risk of bias on a component-by-component basis. We also explored for subgroup effects based on type of clinical condition but did not pre-specify an anticipated direction of association. We conducted subgroup analyses if there were two or more studies in each subgroup, and evaluated credibility of significant subgroup effects using ICEMAN criteria [35].

We performed meta-regression to explore the relationship between the proportion of women, severity of the presenting complaint, mean age, and years of formal education, with the prevalence of symptom exaggeration. If meta-regression suggested an association, we used visual inspection of the associated scatterplot to estimate a threshold and conducted subgroup analysis. We performed all analyses using Stata software version 16.0 [36]. All comparisons were two-tailed, with a threshold P -value of 0.05.

Ethics approval and consent to participate

We did not require ethics approval for this systematic review and meta-analysis due to our sole use of already published data.

Systematic review update

Considering the speed at which studies exploring the prevalence of symptom exaggeration among IME attendees are published, we plan to update this review within the next five years [37].

Results

Of 20,405 unique citations identified in our search, 44 English-language studies that reported on 46 cohorts and 9,794 patients were eligible for review. (Fig 1). None of the studies had overlapping cohorts. In S5 Table we detail the included and excluded studies with reasons at full text screening. Of the 46 cohorts, 67% (n=31) reported on patients with traumatic brain injuries (TBI) with or without mixed neurological diseases, 24% (n=11) on chronic pain patients, and 9% (n=4) on other populations including toxic exposure (n=1) [38], personal injury claimants that were not described (n=1) [39], patients with memory impairment (n=1) [13] and claimants reporting cognitive dysfunction following exposure to occupational and environmental substances (n=1) [40]. In terms of criteria used to identify individuals who were exaggerating symptoms, 61% (n=28) of cohorts relied on the Slick, Sherman, and Iverson criteria for probable malingered neurocognitive dysfunction [23], 24% (n=11) on the Bianchini criteria [24], and 15% (n=7) used other criteria such as those proposed by Greiffenstein, Gola, and Baker [41], Nies and Sweet [22] or Lees-Haley methods [42] (Table 1).

Risk of bias

Of the 32% of studies that described their sampling method (14 of 44), 13 used consecutive sampling and one used random sampling methods to identify IME referrals. All studies reported minimal missing data (<5%). Most studies (n=29, 64%) showed similar age and education characteristics between exaggerating and non-exaggerating groups. No study

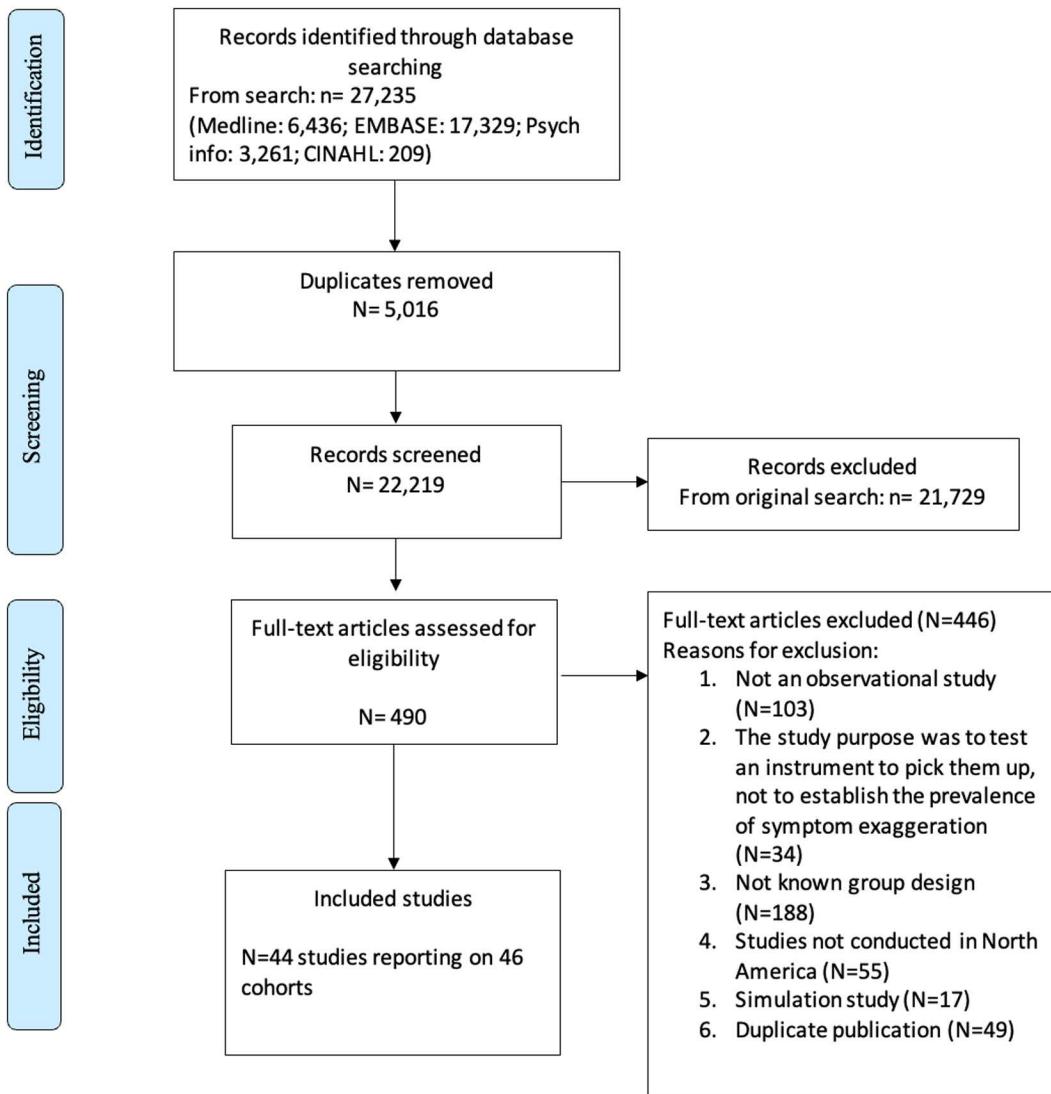


Fig 1. PRISMA flow chart.

<https://doi.org/10.1371/journal.pone.0324684.g001>

explicitly stated that IME assessors administered the index test without knowledge of the reference standard. We had moderate confidence in the reference standard used by most studies (n=35, 80%). None of the known group designs used to evaluate symptom exaggeration provided evidence of reliability and validity testing; however, there has been formal evaluation of psychometric properties of forced-choice tests that were administered in eligible studies (See [S4 Table](#) in supplementary material for details). ([S2 Table](#)).

Prevalence of symptom exaggeration and additional analyses

The prevalence of symptom exaggeration ranged from 17% to 67%, median 33% (inter-quartile range: 25–44), and the pooled prevalence was 35% (95% confidence interval [CI]: 31–39) (low certainty evidence) ([Fig 2](#)). However, we found a significant subgroup effect, of low to moderate credibility, that studies with a higher proportion of women (≥40% vs. <40%) may be associated with higher rates of exaggeration: 47% (95%CI 36–58) vs. 31% (95%CI 28–35) (test of interaction

Table 1. Study Characteristics.

First author, year	Study Design (N)	Sampling Method	Study population	Age (mean)	% Female	Education (mean/years)	Method used to assess symptom exaggeration
Lees-Haley, 1991 [39]	Prospective cohort (N=45)	NR	Personal Injury claimants	37.8	58%	NR	Credibility scale: 100-item true-false scale that provides a sample of the claimant's behavior during the evaluation session and compares that behavior with both the professional experience of the clinician and with the scores of normative samples.
Greiffenstein, 1995 [41]	Retrospective cohort (N=177)	Consecutive	TBI patients	35.4	NR	12.1	4 Criteria: (1) improbable symptom histories, (2) improbably poor neuropsychological test scores not accounted for by physical or sensory limitations, (3) claims of subjective remote memory loss and (4) total disability in at least one major social role * Two or more of the features had to be present.
Suhr, 1997 [43]	Retrospective cohort (N=96)	NR	TBI patients	35.3	46%	13.2	Criteria by Greiffenstein et al. (1994) ^a [41] with a modification that poor performance on neuropsychological tests could not be used solely to make a definitive criterion for symptom exaggeration.
Costa, 1999 [13]	Retrospective cohort (N=42)	Consecutive	Patients with memory impairment	40.7	55%	11.5	Criteria: (1) below chance scores on Victoria Test, (2) less than two rows on Rey-15, (3) digits forward ≤ 3 and digits backward ≥ 4 , (4) endorsement of one or more of the following improbable procedural memory deficits: forgetting the order of walking, chewing, and swallowing movements, forgetting how to speak, constantly forgetting the way home or in own home, (5) endorsement of one or more implausible/incorrect items judged to be easy even for the moderately impaired patients with genuine amnesia, and (6) evidence of current, gainful employment * One or more of the criteria had to be present.
Van Gorp, 1999 [44]	Retrospective cohort (N=81)	NR	TBI patients	36.6	NR	13.3	Criteria: (1) improbable symptom history; (2) total disability in work or a major social role after 1 year from a mild closed-head injury in which loss of consciousness was less than 1 hr; (3) claims of remote or autobiographical memory loss; and (4) at least one failure on one or more neuropsychological malingering tests * One or more of the criteria had to be present.
Sweet, 2000 [45]	Retrospective cohort (N=63)	NR	TBI patients	37.4	NR	13.8	Criteria: (1) poor effortful performance on MDMT and/or Rey 15 Item, (2) evidence of insufficient effort on one or more traditional neuropsychological measures for which valid criteria have been established, (3) plus absence of credible history of neurotrauma, blatant discrepancy between potential injury and patient complaints, (4) blatant discrepancy between type of potential disorder and neuropsychological presentation, and (5) exaggerated patient presentation within a context of litigation or disability application * More than one of the above criteria and lacking a plausible alternative explanation for patient behavior had to present
Greve, 2003 [46]	Retrospective cohort (N=151)	NR	TBI patients	36.6	34%	12.8	Criteria by Slick et al. (1999) ^b [23]

(Continued)

Table 1. (Continued)

First author, year	Study Design (N)	Sampling Method	Study population	Age (mean)	% Female	Education (mean/years)	Method used to assess symptom exaggeration
Lu, 2003 [47]	Retrospective cohort (N=128)	Consecutive	TBI and mixed neurological conditions	42.5	44%	12.9	Criteria by Greiffenstein et al. 1994 [41] and van Gorp et al., 1999 [44]: (1) involvement in litigation or seeking to obtain or maintain disability benefits for reported symptoms and impairments at the time of evaluation, (2) evidence of noncredible cognitive symptoms drawn from at least two of six tests designed to discreetly assess motivation and cooperation, and (3) at least one of six "external" criteria or behavioral presentations that are often observed by clinicians as signs of noncredible symptomatology * All three criteria had to be present.
Barrash, 2004 [48]	Retrospective cohort (N=108)	NR	TBI and mixed neurological conditions	45.2	54%	12.9	Criteria by Greiffenstein et al. 1995 [41]: (1) Minimal brain injury (loss of consciousness and post traumatic amnesia <5min; No CT/MRI indications of brain injury and Glasgow Coma Scale>=13), (2) Clear issues of secondary gain (financial compensation, formal accommodations in work or school setting and adjudication issues), and (3) Evidence of dissimulation independent of neuropsychological performances, as indicated by at least two of the following: marked disability in a major psychosocial role, contradiction between patient and collateral sources of information, complaints of remote memory loss or other symptoms that are rarely seen as a consequence of mild head injury.
Heinly, 2005 [49]	Retrospective cohort (N=344)	NR	TBI patients	39.6	30%	12.1	Criteria by Slick et al. (1999) ^b [23]
Curtis, 2006 [50]	Retrospective cohort (N=275)	NR	TBI patients	38.7	28%	12.3	Criteria by Slick et al. (1999) ^b [23]
Etherton, 2006a [51]	Retrospective cohort (N=81)	NR	Chronic pain patients	43.3	36%	11.9	Criteria by Bianchini et al. (2005) ^c [24]
Greve, 2006a [52]	Not reported (N=259)	NR	TBI patients	38.7	29%	12.5	Criteria by Slick et al. (1999) ^b [23]
Greve, 2006b [53]	Not reported (N=161)	NR	TBI patients	39.3	27%	12.3	Criteria by Slick et al. (1999) ^b [23]
Greve, 2006c [54]	Not reported (N=262)	NR	TBI patients	38.3	27%	12.2	Criteria by Slick et al. (1999) ^b [23]
Greve, 2006d [40]	Retrospective cohort (N=128)	NR	Cognitive dysfunction upon exposure to occupational and environmental substances	40.8	28%	12	Criteria by Slick et al. (1999) ^b [23]
Ardolf, 2007 [55]	Retrospective cohort (N=105)	NR	TBI and mixed neurological conditions	40.1	0%	10.5	Criteria by Slick et al. (1999) ^b [23]

(Continued)

Table 1. (Continued)

First author, year	Study Design (N)	Sampling Method	Study population	Age (mean)	% Female	Education (mean/years)	Method used to assess symptom exaggeration
Greve, 2007 [56]	Prospective cohort (N=206)	NR	TBI patients	39.0	30%	12.6	Criteria by Slick et al. (1999) ^b [23]
Henry, 2007 [57]	Retrospective cohort (N=54)	NR	TBI and mixed neurological conditions	39.8	46%	14.3	Criteria by Slick et al. (1999) ^b [23]
O'Bryant, 2007 [58]	Retrospective cohort (N=329)	Consecutive	TBI and mixed neurological conditions	41	33%	12.7	Criteria by Slick et al. (1999) ^b [23]
Greve, 2007a [38]	Retrospective cohort (N=123)	NR	Toxic exposure patients	41.3	29%	12.0	Criteria by Slick et al. (1999) ^b [23]
Aguerrev-ere, 2008 [59]	Retrospective cohort (N=185)	NR	TBI patients	37.8	28%	12.4	Criteria by Slick et al. (1999) ^b [23]
Curtis, 2008 [60]	Prospective cohort (N=204)	NR	TBI patients	39.6	29%	12.3	Criteria by Slick et al. (1999) ^b [23]
Greve, 2008 [61]	Prospective cohort (N=211)	NR	TBI patients	38.3	28%	12.1	Criteria by Slick et al. (1999) ^b [23]
Ord, 2008 [62]	Not reported (N=93)	NR	TBI patients	36.2	36%	12.7	Criteria by Slick et al. (1999) ^b [23]
Greve, 2008b [63]	Not reported (N TB=109; N Chronic pain=228)	NR	TBI and Chronic pain patients	TBI: 40.35 Chronic pain: 42.5	TBI: 24% Chronic pain: 35%	TBI: 12.27 Chronic pain: 11.8	Criteria by Slick et al. (1999) ^b [23]
Henry, 2009 [64]	Retrospective cohort (N=161)	Consecutive	TBI and mixed neurological conditions	42.0	41%	13.83	Criteria by Slick et al. (1999) ^b [23]
Greve, 2009 [12]	Retrospective cohort (N=282)	NR	TBI patients	37.7	27%	11.7	Criteria by Slick et al. (1999) ^b [23]
Greve, 2009a [65]	Retrospective cohort (N=318)	Random	Chronic pain patients	41.2	35%	11.8	Criteria by Bianchini et al. (2005) ^c [24]
Greve, 2009b [66]	Prospective cohort (N TB=442; N Chronic pain=378)	NR	TBI and Chronic pain patients	TBI: 38.7 Chronic pain: 42.4	TBI: 29% Chronic pain: 37%	TBI: 12.3 Chronic pain: 11.6	Criteria by Bianchini et al. (2005) ^c [24]
Greve, 2009c [67]	Retrospective cohort (N=604)	Consecutive	Chronic pain patients	42.3	36%	11.7	Criteria by Bianchini et al. (2005) ^c [24]
Greve, 2009d [68]	Retrospective cohort (N=508)	Consecutive	Chronic pain patients	42.1	35%	11.6	Criteria by Bianchini et al. (2005) ^c [24]

(Continued)

Table 1. (Continued)

First author, year	Study Design (N)	Sampling Method	Study population	Age (mean)	% Female	Education (mean/years)	Method used to assess symptom exaggeration
Bortnik, 2010 [69]	Retrospective cohort (N= 188)	NR	TBI and mixed neurological conditions	42.7	49%	11.9	Criteria by Slick et al. (1999) ^b [23]
Curtis, 2010 [70]	Retrospective cohort (N= 74)	Consecutive	TBI and mixed neurological conditions	36.3	35%	13	Criteria by Slick et al. (1999) ^b [23]
Greve, 2010 [71]	Retrospective cohort (N= 612)	Consecutive	Chronic pain patients	41.1	35%	11.7	Criteria by Bianchini et al. (2005) ^c [24]
Ord, 2010 [72]	Retrospective cohort (N= 84)	NR	TBI patients	39.4	37%	13.0	Criteria by Slick et al. (1999) ^b [23]
Aguerrev-ere, 2011 [25]	Prospective cohort (N= 108)	Consecutive	TBI patients	39.9	26%	12.4	Criteria by Slick et al. (1999) ^b [23]
Rober-son, 2013 [73]	Retrospective cohort (N= 315)	NR	TBI and mixed neurological conditions	43.1	44%	13.1	Criteria by Slick et al. (1999) ^b [23]
Bianchini, 2014 [74]	Retrospective cohort (N= 328)	NR	Chronic pain patients	43.3	35%	12.1	Criteria by Bianchini et al. (2005) ^c [24]
Guise, 2014 [75]	Retrospective cohort (N= 119)	Consecutive	TBI patients	38.3	31%	12.6	Criteria by Slick et al. (1999) ^b [23]
Patrick, 2014 [76]	Retrospective cohort (N= 52)	Consecutive	TBI and mixed neurological conditions	43.5	17%	13.1	Criteria by Slick et al. (1999) ^b [23]
Aguerrev-ere, 2017 [77]	Retrospective cohort (N= 348)	NR	Chronic pain patients	43.1	36%	11.7	Criteria by Bianchini et al. (2005) ^c [24]
Bianchini, 2018 [78]	Retrospective cohort (N= 501)	Consecutive	Chronic pain patients	42.3	NR	11.3	Criteria by Bianchini et al. (2005) ^c [24]
Curtis, 2019 [79]	Retrospective cohort (N= 219)	NR	Chronic pain patients	43.5	32%	12.2	Criteria by Bianchini et al. (2005) ^c [24]

NR = not reported; MDMT = Medical Symptom Validity Test; PDRT = Portland Digit Recognition Test; TOMM = Test of Memory Malingering; RDS = Reliable Digit Span; FBS = Fake Bad Scale; MI = Malingering Index; WMI = Working Memory Index; PSI = Processing Speed Index

^a Criteria by Greffeinstein et al. 1994 [41] include (1) improbable poor performance on more than two neuropsychological measures, (2) total disability in a major social role, (3) contradiction between collateral sources and symptom history, and (4) remote memory loss

^b Criteria by Slick et al. (1999) [23] include (A) presence of substantial external incentive, (B) evidence from neuropsychological testing, (C) evidence from self-report, and (D) behaviors meeting the necessary B and C criteria are not fully accounted for by psychiatric, neurological, or developmental factors. External incentive (Criterion A) plus Criterion B and/or C evidence had to present for a diagnosis of malingering. Criterion B behaviors are sufficient for a diagnosis of malingering on their own.

^c Criteria by Bianchini et al. (2005) [24] reflects a modification of the criteria of Slick et al. (1999) [23] and includes external incentive and meeting one of the following four conditions: (1) positive findings on either [PDRT or TOMM or RDS] and positive findings on either [FBS or MI]; (2) positive findings on [WMI and PSI] and positive findings on either [FBS or MI]; (3) positive findings on either [PDRT or TOMM] and [WMI and PSI]; or (4) significantly below chance on either [PDRT or TOMM].

<https://doi.org/10.1371/journal.pone.0324684.t001>

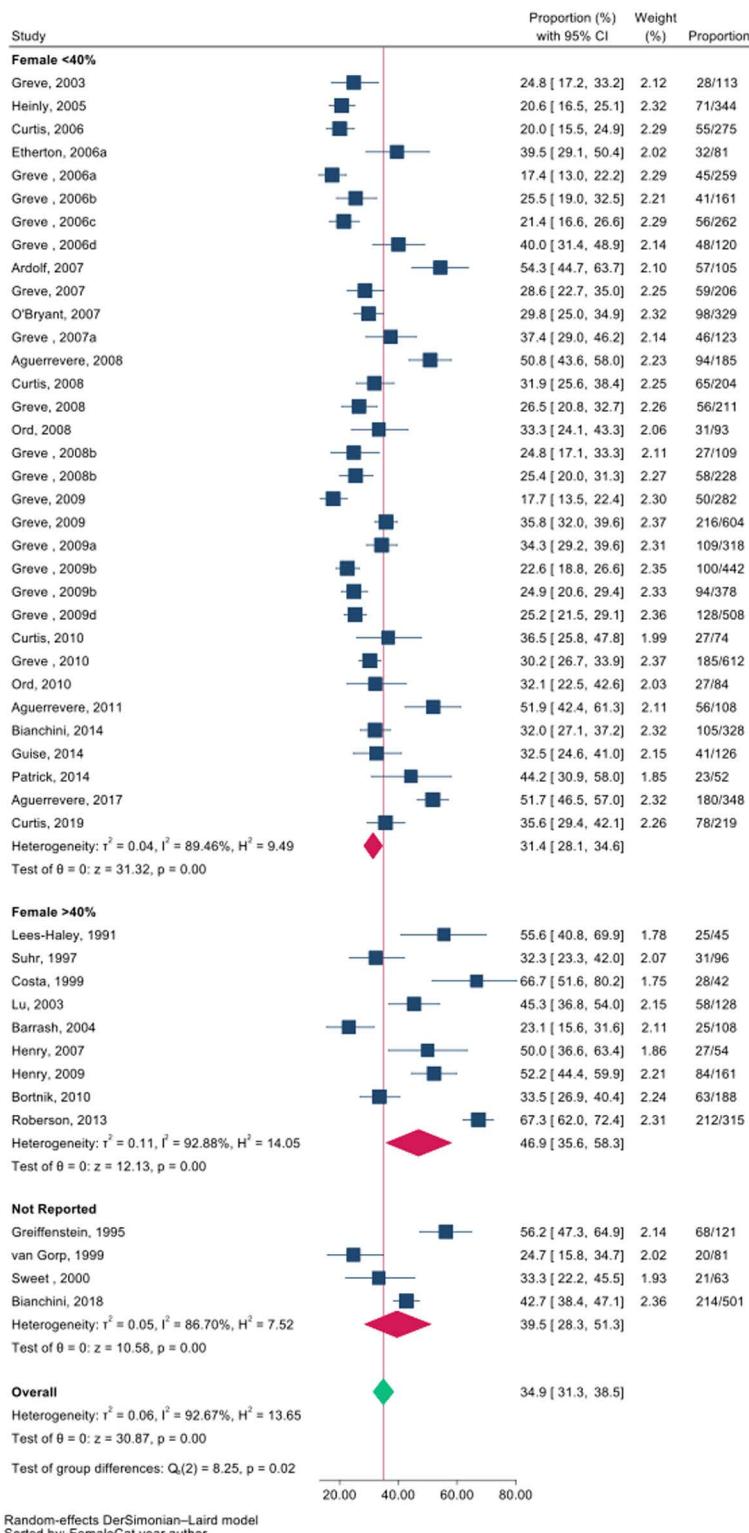


Fig 2. Forest plot for prevalence by proportion of females (P=0.02).

<https://doi.org/10.1371/journal.pone.0324684.g002>

Table 2. GRADE evidence profile: prevalence of symptom exaggeration among IME attendees in North America.

# of studies	# of patients	Risk of bias	Inconsistency ^a	Indirectness ^b	Imprecision ^c	Publication bias	Prevalence (95% CI)	Overall certainty in the evidence
Prevalence of symptom exaggeration in studies with >40% female participants								
9	1,137	Serious ^d	Serious ^e	Not serious	Not serious	Could not be assessed as number of included studies <10	46.9% (95% CI: 35.6–58.3)	Low
Prevalence of symptom exaggeration in studies with <40% female participants								
33	7,891	Serious ^d	Serious ^e	Not serious	Not serious	Undetected; symmetric funnel plot; Egger's test P=0.16	31.4% (95% CI: 28.1–34.6)	Low

^a Inconsistency refers to variability in effect estimates across studies (i.e., heterogeneity) that could not be adequately explained.

^b Indirectness results if the intervention, patients or outcomes are different from the research question under investigation

^c In this review, serious imprecision is based on the position of the confidence interval relative to a 20% threshold for symptom exaggeration and if the effect on the patient, or clinical action, would differ depending on whether the upper or the lower boundary of the confidence interval represented the truth.

^d We downgraded 1 level for risk of bias as of none the criteria used to evaluate symptom exaggeration have been formally validated.

^e We downgraded one level for inconsistency due to a wide range in prevalence among eligible studies (17.4% to 67.3%), which was partially explained by participant sex. Specifically, studies enrolling a higher proportion of women (≥40% vs. <40%) were associated with higher rates of symptom exaggeration: 47% (95%CI 36–58) vs. 31% (95%CI 28–35; test of interaction p=0.02).

<https://doi.org/10.1371/journal.pone.0324684.t002>

p=0.02; [Fig 2](#), [Tables 2](#) and S3). We did not detect any evidence of small study effects for the overall prevalence of symptom exaggeration (Egger's test P=0.13; S2 Fig) nor for the subgroup of studies with <40% women (Egger's test P=0.16; S2 Fig).

We found no significant subgroup effects for type of clinical condition (mild TBI versus chronic pain versus other conditions), confidence in the reference standard, age, or education ([S3–S5 Figs](#)). Meta-regression showed no association between prevalence of symptom exaggeration and age, level of education, or severity of presenting complaint, but did suggest an association with the proportion of female participants ([S1](#), [S6](#) and [S7 Figs](#)). We present all extracted data per study in [S6 Table](#).

Discussion

Our systematic review and meta-analysis of observational studies found low certainty evidence, rated down due to risk of bias and inconsistency, that symptom exaggeration may be common among individuals attending for IMEs in North America, affecting approximately 1 in 3 assessments. The prevalence of symptom exaggeration was higher in studies that enrolled a greater proportion of female attendees (47%) vs. a lower proportion of female attendees (31%).

Relation to other studies

This is the first systematic review to summarize the extent of symptom exaggeration among IME attendees in North America. A previous survey of 131 US board-certified neuropsychologists conducting forensic work found that, on average, they estimated 30% of examinees claiming personal injury, disability, or workers' compensation presented with symptom exaggeration. However, estimated prevalence ranged considerably by diagnosis – from an average of 41% for mild head injuries to 2% for vascular dementia [80]. Our review found no evidence for differences in the prevalence of symptom exaggeration based on clinical condition, but most patients among studies eligible for our review presented with either mild TBI or chronic pain.

Although our review focused on IMEs in North America, data from other regions also suggest high rates of symptom exaggeration. An observational study in Spain reported that of 1,003 participants (61.5% female), drawn from unselected undergraduates, advanced psychology students, the general population, forensic psychologists, and forensic/legal medicine physicians, one-third reported having feigned symptoms or illness [81]. Data from Germany and the Netherlands

suggest that one-fifth to one-third of clients in forensic or insurance contexts exhibit symptom overreporting [82]. Further, a Swiss study found that 28% to 34% of individuals undergoing medico-legal evaluations demonstrated probable or definite symptom exaggeration [83].

Our finding suggesting that women are more likely to exaggerate symptoms vs. men is supported by a systematic review of 175 studies that found women report more bodily distress and more numerous, more intense, and more frequent somatic symptoms than men [84]. Reasons for this discrepancy are uncertain, but may include biological differences, greater bodily vigilance and awareness, and higher rates of negative affectivity vs. men [84]. When symptoms are disproportionate to objective pathology, clinicians should inquire about other factors. For example, women are more likely to experience intimate partner violence than men [85,86], and pain patients who report lifetime traumatic events experience greater pain severity [87].

Studies eligible for our review used different strategies and approaches for assessing the prevalence of symptom exaggeration. The National Academy of Neuropsychologists (NAN) and American Academy of Clinical Neuropsychology (AACN) have emphasized the use of a multimethod approach to assess symptom and performance validity. These include clinical interviews, medical records, medical investigations in certain cases, behavioural observations, and symptom and performance validity tests [88]. Specific guidance is not provided on which symptom and performance validity tests should be used, when they should be conducted, and how they should be interpreted [89].

Strengths and limitations

Our study has several methodological strengths including (1) restricting our eligibility criteria to studies employing a known group design or multi-modal approach to assess symptom exaggeration, (2) subgroup analysis and assessment consistent with current best practices [35,90], and (3) use of the GRADE approach to evaluate the certainty of evidence.

In terms of limitations, we restricted our review to IMEs conducted in North America and eligible studies focused mainly on chronic pain and TBI. The generalizability of our findings to other jurisdictions, contexts, and clinical conditions, is uncertain. We were unable to explore the effect of cultural variability on the prevalence of symptom exaggeration as we found no studies within our inclusion criteria that addressed this issue. We did not find evidence for a subgroup effect based on confidence in the reference standard; however, there may have been insufficient variability to identify an association as almost all studies used a reference standard in which we rated moderate confidence. Another limitation of our review is the absence of a compelling reference standard for symptom exaggeration. Furthermore, even within the same reference standard, operationalization can be variable, which may affect prevalence. Another limitation of the primary studies is the lack of stratification of prevalence of symptom exaggeration according to possible effect modifiers, such as sex. Doing so would facilitate within-study subgroup analysis, which are less subject to confounding than between-study subgroup analysis. Another major limitation of the current evidence is that none of the known group approaches for evaluating symptom exaggeration have undergone reliability and validity testing.

Implications for future research and practice

Failure to identify the contribution of symptom exaggeration towards examinee's complaints not only compromises the reliability and validity of independent assessments but may also adversely impact patient care by medicalizing psychosocial issues [91–93]. Our findings suggest that symptom exaggeration is common among patients attending for IMEs; however, we rated down the certainty of evidence due to uncertain psychometric properties of the criteria used to evaluate exaggeration. An urgent research priority is the evaluation of inter-rater reliability of known group and multi-modal systems to appraise symptom exaggeration. Validation of such assessment systems is also critical and extremely challenging, but indirect evidence of validity could be acquired by evaluating accuracy in distinguishing between volunteers who were or were not exaggerating symptoms.

Future research should investigate how cultural factors affect IME outcomes, with attention to language barriers, health beliefs, and potential biases among both examinees and assessors. Another research priority is the development and validation of a structured and comprehensive approach to identify symptom exaggeration in IME assessments. Such an approach should consider observed versus reported abilities, findings of other providers, self-reported history that is discrepant with documented history, and administration of validated tests. A further consideration for research and practice is the use of symptom validity tests that focus on malingering (e.g., Test of Memory Malingering [TOMM], Lees-Haley Fake Bad Scale [FBS]), which imply intent. Clinicians are, understandably and appropriately, hesitant to assign a label of malingering; reasons include the challenges associated with determining intent and the risk of litigation [94]. To circumvent these issues, we would suggest the use of the less value-laden term 'symptom exaggeration'.

Conclusion

Symptom exaggeration may occur in almost 50% of women and in approximately a third of men undergoing IMEs. Assessors should evaluate symptom exaggeration when conducting IMEs using a multi-modal approach that includes both clinical findings and validated tests of performance effort, and avoid conflation with malingering which presumes intent. Priority areas for future research include establishing the reliability and validity of current evaluation criteria for symptom exaggeration, and development of a structured IME assessment approach that includes consideration of symptom exaggeration.

Supporting information

S1 Table. Search strategies.

(DOCX)

S2 Table. Risk of bias assessment.

(DOCX)

S3 Table. ICEMAN criteria to assess credibility of subgroup effect of female % and prevalence.

(DOCX)

S4 Table. Psychometric properties of tests included in symptom exaggeration criteria with list of references.

(DOCX)

S5 Table. Included and excluded studies at full text screening with reasons.

(DOCX)

S6 Table. Data extracted from included studies.

(DOCX)

S1 Fig. Meta-regression for proportion of females among 42 studies ($p=0.16$).

(DOCX)

S2 Fig. a- Funnel plots of overall prevalence (Egger's test $p=0.13$) and b- prevalence in subgroup of studies with female proportion <40% (Egger's test $p=0.16$).

(DOCX)

S3 Fig. Subgroup analysis for type of conditions (test of interaction $p=0.95$).

(DOCX)

S4 Fig. Subgroup analysis for confidence in reference standard (test of interaction $p=0.84$).

(DOCX)

S5 Fig. Subgroup analysis for similar age and/or education between groups (test of interaction p=0.47).
(DOCX)

S6 Fig. Meta-regression for average age among 46 cohorts (p=0.18).
(DOCX)

S7 Fig. Meta-regression for average education level among 45 cohorts (p=0.65).
(DOCX)

S1 Checklist. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist.
(DOCX)

S2 Checklist. Meta-analysis of Observational Studies in Epidemiology (MOOSE) checklist.
(DOCX)

Acknowledgments

We would like to thank Michael Bagby from the Departments of Psychology and Psychiatry at University of Toronto for his contributions to the initial discussions around conceptualization and design of this study. No financial compensation was provided to any of these individuals.

Author contributions

Conceptualization: Andrea J. Darzi, Rachel J. Couban, Regina Kunz, Gordon H. Guyatt, Jason W. Busse.

Data curation: Andrea J. Darzi, Li Wang, John J. Riva, Rami Z. Morsi, Rana Charide, Rachel J. Couban, Samer G. Karam, Kian Torabiardakani, Annie Lok, Shamil Ebrahim, Jason W. Busse.

Formal analysis: Andrea J. Darzi, Li Wang, Jason W. Busse.

Investigation: Andrea J. Darzi, Jason W. Busse.

Methodology: Andrea J. Darzi, John J. Riva, Shamil Ebrahim, Regina Kunz, Jason W. Busse.

Project administration: Andrea J. Darzi, Jason W. Busse.

Resources: Andrea J. Darzi, Jason W. Busse.

Software: Andrea J. Darzi, Jason W. Busse.

Supervision: Andrea J. Darzi, Jason W. Busse.

Validation: Andrea J. Darzi, John J. Riva, Sheena Bance, Regina Kunz, Gordon H. Guyatt, Jason W. Busse.

Visualization: Andrea J. Darzi, Jason W. Busse.

Writing – original draft: Andrea J. Darzi, Jason W. Busse.

Writing – review & editing: Andrea J. Darzi, Li Wang, John J. Riva, Rami Z. Morsi, Rana Charide, Rachel J. Couban, Samer G. Karam, Kian Torabiardakani, Annie Lok, Shamil Ebrahim, Sheena Bance, Regina Kunz, Gordon H. Guyatt, Jason W. Busse.

References

1. Canada S. Canadian Survey on Disability, 2017 to 2022; 2023. Available from: <https://www150.statcan.gc.ca/n1/daily-quotidien/231201/dq231201b-eng.htm>
2. Disability and Health Data System (DHDS) [Internet]; 2020 [cited 2023 Jan 16]. Available from: <https://dhds.cdc.gov/SP?LocationId=59&Category-Id=DISEST&ShowFootnotes=true&showMode=&IndicatorIds=STATYPE,AGEIND,SEXIND,RACEIND,VETIND&pnl0=Chart,fal...>

3. Martin DW. *Independent medical evaluation: a practical guide*. Springer; 2018.
4. Ebrahim S, Sava H, Kunz R, Busse JW. Ethics and legalities associated with independent medical evaluations. *CMAJ*. 2014;186(4):248–9. <https://doi.org/10.1503/cmaj.131509> PMID: 24491474
5. Gill D, Green P, Flaro L, Pucci T. The role of effort testing in independent medical examinations. *Med Leg J*. 2007;75(Pt 2):64–71. <https://doi.org/10.1258/rsmmlj.75.2.64> PMID: 17822166
6. Mæland S, Holmås TH, Øyeflaten I, Husabø E, Werner EL, Monstad K. What is the effect of independent medical evaluation on days on sickness benefits for long-term sick listed employees in Norway? A pragmatic randomised controlled trial, the NIME-trial. *BMC Public Health*. 2022;22(1):400. <https://doi.org/10.1186/s12889-022-12800-1> PMID: 35216560
7. Barth J, de Boer WE, Busse JW, Hoving JL, Kedzia S, Couban R. Inter-rater agreement in evaluation of disability: systematic review of reproducibility studies. *BMJ*. 2017;356.
8. Kunz R, von Allmen DY, Marelli R, Hoffmann-Richter U, Jeger J, Mager R, et al. The reproducibility of psychiatric evaluations of work disability: two reliability and agreement studies. *BMC Psychiatry*. 2019;19(1):205. <https://doi.org/10.1186/s12888-019-2171-y> PMID: 31266488
9. Bachmann M, de Boer W, Schandlmaier S, Leibold A, Marelli R, Jeger J, et al. Use of a structured functional evaluation process for independent medical evaluations of claimants presenting with disabling mental illness: rationale and design for a multi-center reliability study. *BMC Psychiatry*. 2016;16:271. <https://doi.org/10.1186/s12888-016-0967-6> PMID: 27474008
10. Boskovic I, Gallardo CT, Vrij A, Hope L, Merckelbach H. Verifiability on the run: an experimental study on the verifiability approach to malingered symptoms. *Psychiatr Psychol Law*. 2018;26(1):65–76. <https://doi.org/10.1080/13218719.2018.1483272> PMID: 31984064
11. Rumschik SM, Appel JM. Malingering in the psychiatric emergency department: prevalence, predictors, and outcomes. *Psychiatr Serv*. 2019;70(2):115–22. <https://doi.org/10.1176/appi.ps.201800140> PMID: 30526343
12. Greve KW, Heinly MT, Bianchini KJ, Love JM. Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *Clin Neuropsychol*. 2009;23(2):343–62. <https://doi.org/10.1080/13854040802054169> PMID: 18609328
13. Costa D. Psychiatric detection of exaggeration in reports of memory impairment. *J Nerv Ment Dis*. 1999;187(7):446–8. <https://doi.org/10.1097/00005053-199907000-00010> PMID: 10426467
14. Walczyk JJ, Sewell N, DiBenedetto MB. A review of approaches to detecting malingering in forensic contexts and promising cognitive load-inducing lie detection techniques. *Front Psychiatry*. 2018;9:700. <https://doi.org/10.3389/fpsyg.2018.00700> PMID: 30622488
15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1. <https://doi.org/10.1186/2046-4053-4-1> PMID: 25554246
16. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group*. *JAMA*. 2000;283(15):2008–12. <https://doi.org/10.1001/jama.283.15.2008> PMID: 10789670
17. BSCI I. Open Science Framework.
18. Distiller S. *Data management software*. Ottawa (ON): Evidence Partners; 2011.
19. Rogers RE. *Clinical assessment of malingering and deception*. Guilford Press; 2008.
20. Rogers R, Kropp PR, Bagby RM, Dickens SE. Faking specific disorders: a study of the Structured Interview of Reported Symptoms (SIRS). *J Clin Psychol*. 1992;48(5):643–8. [https://doi.org/10.1002/1097-4679\(199209\)48:5<643::aid-jclp2270480511>3.0.co;2-2](https://doi.org/10.1002/1097-4679(199209)48:5<643::aid-jclp2270480511>3.0.co;2-2) PMID: 1401150
21. Heilbronner RL, Sweet JJ, Morgan JE, Larabee GJ, Millis SR, 1 CP. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol*. 2009;23(7):1093–129.
22. Nies KJ, Sweet JJ. Neuropsychological assessment and malingering: a critical review of past and present strategies. *Arch Clin Neuropsychol*. 1994;9(6):501–52. <https://doi.org/10.1093/arclin/9.6.501> PMID: 14590999
23. Slick DJ, Sherman EM, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol*. 1999;13(4):545–61. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545) PMID: 10806468
24. Bianchini KJ, Greve KW, Glynn G. On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine J*. 2005;5(4):404–17. <https://doi.org/10.1016/j.spinee.2004.11.016> PMID: 15996610
25. Aguerrevere LE, Greve KW, Bianchini KJ, Ord JS. Classification accuracy of the Millon Clinical Multiaxial Inventory-III modifier indices in the detection of malingering in traumatic brain injury. *J Clin Exp Neuropsychol*. 2011;33(5):497–504. <https://doi.org/10.1080/13803395.2010.535503> PMID: 21424973
26. Cook RJ, Farewell VT. Conditional inference for subject-specific and marginal agreement: two families of agreement measures. *Can J Stat*. 1995;23(4):333–44. <https://doi.org/10.2307/3315378>
27. Rogers R. *Clinical assessment of malingering and deception. (No Title)*. 2009.
28. Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Statist*. 1950;21(4):607–11. <https://doi.org/10.1214/aoms/117729756>
29. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Arch Public Health*. 2014;72:1–10.

30. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2) PMID: 3802833
31. Miller JJ. The inverse of the freeman-tukey double arcsine transformation. *Am Stat*. 1978;32(4):138. <https://doi.org/10.2307/2682942>
32. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383–94. <https://doi.org/10.1016/j.jclinepi.2010.04.026> PMID: 21195583
33. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–34. <https://doi.org/10.1136/bmj.315.7109.629> PMID: 9310563
34. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 2008;8:79. <https://doi.org/10.1186/1471-2288-8-79> PMID: 19036172
35. Schandlmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192(32):E901–6. <https://doi.org/10.1503/cmaj.200077> PMID: 32778601
36. StataCorp L. Stata statistical software: release 16. College Station (TX): StataCorp; 2019.
37. Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, et al. When and how to update systematic reviews: consensus and checklist. *BMJ*. 2016;354.
38. Greve KW, Springer S, Bianchini KJ, Black FW, Heinly MT, Love JM. Malingering in toxic exposure: classification accuracy of Reliable Digit Span and WAIS-III Digit Span scaled scores. *Assessment*. 2007;14(1):12–21.
39. Lees-Haley PR, English LT, Glenn WJ. A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychol Rep*. 1991;68(1):203–10. <https://doi.org/10.2466/pr0.1991.68.1.203> PMID: 2034762
40. Greve KW, Bianchini KJ, Black FW, Heinly MT, Love JM, Swift DA, et al. The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *Neurotoxicology*. 2006;27(6):940–50. <https://doi.org/10.1016/j.neuro.2006.06.009> PMID: 16904749
41. Greiffenstein MF, Gola T, Baker WJ. MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *Clin Neuropsychol*. 1995;9(3):230–40. <https://doi.org/10.1080/13854049508400485>
42. Lees-Haley PR. Provisional normative data for a credibility scale for assessing personal injury claimants. *PR*. 1990;66(3):1355. <https://doi.org/10.2466/pr0.66.3.1355-1360>
43. Suhr J, Tranel D, Wefel J, Barrash J. Memory performance after head injury: contributions of malingering, litigation status, psychological factors, and medication use. *J Clin Exp Neuropsychol*. 1997;19(4):500–14. <https://doi.org/10.1080/01688639708403740> PMID: 9342686
44. van Gorp WG, Humphrey LA, Kalechstein AL, Brumm VL, McMullen WJ, Stoddard MA, et al. How well do standard clinical neuropsychological tests identify malingering? A preliminary analysis. *J Clin Exp Neuropsychol*. 1999 Apr;21(2):245–50. <https://doi.org/10.1076/jcen.21.2.245.933> PMID: 10425521
45. Sweet JJ, Wolfe P, Sattlberger E, Numan B, Rosenfeld JP, Clingerman S, et al. Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Arch Clin Neuropsychol*. 2000;15(2):105–13. <https://doi.org/10.1093/arclin/15.2.105> PMID: 14590555
46. Greve KW, Bianchini KJ, Mathias CW, Houston RJ, Crouch JA. Detecting malingered performance on the Wechsler Adult Intelligence Scale. Validation of Mittenberg's approach in traumatic brain injury. *Arch Clin Neuropsychol*. 2003;18(3):245–60. <https://doi.org/10.1093/arclin/18.3.245> PMID: 14591458
47. Lu PH, Boone KB, Cozolino L, Mitchell C. Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers recognition trial in the detection of suspect effort. *Clin Neuropsychol*. 2003;17(3):426–40. <https://doi.org/10.1076/clin.17.3.426.18083> PMID: 14704893
48. Barrash J, Suhr J, Manzel K. Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTx): validation with clinical samples. *J Clin Exp Neuropsychol*. 2004;26(1):125–40. <https://doi.org/10.1076/jcen.26.1.125.23928> PMID: 14972700
49. Heinly MT, Greve KW, Bianchini KJ, Love JM, Brennan A. WAIS digit span-based indicators of malingered neurocognitive dysfunction: classification accuracy in traumatic brain injury. *Assessment*. 2005;12(4):429–44. <https://doi.org/10.1177/1073191105281099> PMID: 16244123
50. Curtis KL, Greve KW, Bianchini KJ, Brennan A. California verbal learning test indicators of Malingered Neurocognitive Dysfunction: sensitivity and specificity in traumatic brain injury. *Assessment*. 2006;13(1):46–61. <https://doi.org/10.1177/1073191105285210> PMID: 16443718
51. Etherton JL, Bianchini KJ, Ciota MA, Heinly MT, Greve KW. Pain, malingering and the WAIS-III Working Memory Index. *Spine J*. 2006;6(1):61–71. <https://doi.org/10.1016/j.spinee.2005.05.382> PMID: 16413450
52. Greve KW, Bianchini KJ, Love JM, Brennan A, Heinly MT. Sensitivity and specificity of MMPI-2 validity scales and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *Clin Neuropsychol*. 2006;20(3):491–512. <https://doi.org/10.1080/13854040590967144> PMID: 16895861
53. Greve KW, Bianchini KJ, Doane BM. Classification accuracy of the test of memory malingering in traumatic brain injury: results of a known-groups analysis. *J Clin Exp Neuropsychol*. 2006;28(7):1176–90. <https://doi.org/10.1080/13803390500263550> PMID: 16840243
54. Greve KW, Bianchini KJ. Classification accuracy of the Portland Digit Recognition Test in traumatic brain injury: results of a known-groups analysis. *Clin Neuropsychol*. 2006;20(4):816–30. <https://doi.org/10.1080/13854040500346610> PMID: 16980264
55. Ardolf BR, Denney RL, Houston CM. Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clin Neuropsychol*. 2007;21(6):899–916. <https://doi.org/10.1080/13825580600966391> PMID: 17886149

56. Greve KW, Bianchini KJ, Roberson T. The Booklet Category Test and malingering in traumatic brain injury: classification accuracy in known groups. *Clin Neuropsychol*. 2007;21(2):318–37. <https://doi.org/10.1080/13854040500488552> PMID: 17455021
57. Henry GK, Enders C. Probable malingering and performance on the Continuous Visual Memory Test. *Appl Neuropsychol*. 2007;14(4):267–74. <https://doi.org/10.1080/09084280701719245> PMID: 18067423
58. O'Bryant SE, Engel LR, Kleiner JS, Vasterling JJ, Black FW. Test of memory malingering (TOMM) trial 1 as a screening measure for insufficient effort. *Clin Neuropsychol*. 2007;21(3):511–21. <https://doi.org/10.1080/13854040600611368> PMID: 17455034
59. Aguerrevere LE, Greve KW, Bianchini KJ, Meyers JE. Detecting malingering in traumatic brain injury and chronic pain with an abbreviated version of the Meyers Index for the MMPI-2. *Arch Clin Neuropsychol*. 2008;23(7–8):831–8. <https://doi.org/10.1016/j.acn.2008.06.008> PMID: 18715751
60. Curtis KL, Thompson LK, Greve KW, Bianchini KJ. Verbal fluency indicators of malingering in traumatic brain injury: classification accuracy in known groups. *Clin Neuropsychol*. 2008;22(5):930–45. <https://doi.org/10.1080/13854040701563591> PMID: 18756393
61. Greve KW, Lotz KL, Bianchini KJ. Observed versus estimated IQ as an index of malingering in traumatic brain injury: classification accuracy in known groups. *Appl Neuropsychol*. 2008;15(3):161–9. <https://doi.org/10.1080/09084280802324085> PMID: 18726736
62. Ord JS, Greve KW, Bianchini KJ. Using the Wechsler Memory Scale-III to detect malingering in mild traumatic brain injury. *Clin Neuropsychol*. 2008;22(4):689–704. <https://doi.org/10.1080/13854040701425437> PMID: 17853130
63. Greve KW, Ord J, Curtis KL, Bianchini KJ, Brennan A. Detecting malingering in traumatic brain injury and chronic pain: a comparison of three forced-choice symptom validity tests. *Clin Neuropsychol*. 2008;22(5):896–918. <https://doi.org/10.1080/13854040701565208> PMID: 18756391
64. Henry GK, Heilbronner RL, Mittenberg W, Enders C, Domboski K. Comparison of the MMPI-2 restructured Demoralization Scale, Depression Scale, and Malingered Mood Disorder Scale in identifying non-credible symptom reporting in personal injury litigants and disability claimants. *Clin Neuropsychol*. 2009;23(1):153–66. <https://doi.org/10.1080/13854040801969524> PMID: 18609325
65. Greve KW, Bianchini KJ, Etherton JL, Ord JS, Curtis KL. Detecting malingered pain-related disability: classification accuracy of the Portland Digit Recognition Test. *Clin Neuropsychol*. 2009;23(5):850–69. <https://doi.org/10.1080/13854040802585055> PMID: 19255913
66. Greve KW, Curtis KL, Bianchini KJ, Ord JS. Are the original and second edition of the California Verbal Learning Test equally accurate in detecting malingering? *Assessment*. 2009;16(3):237–48. <https://doi.org/10.1177/1073191108326227> PMID: 19098280
67. Greve KW, Etherton JL, Ord J, Bianchini KJ, Curtis KL. Detecting malingered pain-related disability: classification accuracy of the test of memory malingering. *Clin Neuropsychol*. 2009;23(7):1250–71. <https://doi.org/10.1080/13854040902828272> PMID: 19728222
68. Greve KW, Ord JS, Bianchini KJ, Curtis KL. Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Arch Phys Med Rehabil*. 2009;90(7):1117–26. <https://doi.org/10.1016/j.apmr.2009.01.018> PMID: 19577024
69. Bortnik KE, Boone KB, Marion SD, Amano S, Ziegler E, Cottingham ME, et al. Examination of various WMS-III logical memory scores in the assessment of response bias. *Clin Neuropsychol*. 2010;24(2):344–57. <https://doi.org/10.1080/13854040903307268> PMID: 19921593
70. Curtis KL, Greve KW, Brasseux R, Bianchini KJ. Criterion groups validation of the Seashore Rhythm Test and Speech Sounds Perception Test for the detection of malingering in traumatic brain injury. *Clin Neuropsychol*. 2010;24(5):882–97. <https://doi.org/10.1080/13854041003762113> PMID: 20486016
71. Greve KW, Bianchini KJ, Etherton JL, Meyers JE, Curtis KL, Ord JS. The Reliable Digit Span test in chronic pain: classification accuracy in detecting malingered pain-related disability. *Clin Neuropsychol*. 2010;24(1):137–52. <https://doi.org/10.1080/13854040902927546> PMID: 19816837
72. Ord JS, Boettcher AC, Greve KW, Bianchini KJ. Detection of malingering in mild traumatic brain injury with the Conners' Continuous Performance Test-II. *J Clin Exp Neuropsychol*. 2010;32(4):380–7. <https://doi.org/10.1080/13803390903066881> PMID: 19739010
73. Roberson CJ, Boone KB, Goldberg H, Miora D, Cottingham M, Victor T, et al. Cross validation of the b Test in a large known groups sample. *Clin Neuropsychol*. 2013;27(3):495–508. <https://doi.org/10.1080/13854046.2012.737027> PMID: 23157695
74. Bianchini KJ, Aguerrevere LE, Guise BJ, Ord JS, Etherton JL, Meyers JE, et al. Accuracy of the Modified Somatic Perception Questionnaire and Pain Disability Index in the detection of malingered pain-related disability in chronic pain. *Clin Neuropsychol*. 2014;28(8):1376–94. <https://doi.org/10.1080/13854046.2014.986199> PMID: 25517267
75. Guise BJ, Thompson MD, Greve KW, Bianchini KJ, West L. Assessment of performance validity in the Stroop Color and Word Test in mild traumatic brain injury patients: a criterion-groups validation design. *J Neuropsychol*. 2014;8(1):20–33. <https://doi.org/10.1111/jnp.12002> PMID: 23253228
76. Patrick RE, Horner MD. Psychological characteristics of individuals who put forth inadequate cognitive effort in a secondary gain context. *Arch Clin Neuropsychol*. 2014;29(8):754–66. <https://doi.org/10.1093/arclin/acu054> PMID: 25318597
77. Aguerrevere LE, Calamia MR, Greve KW, Bianchini KJ, Curtis KL, Ramirez V. Clusters of financially incentivized chronic pain patients using the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF). *Psychol Assess*. 2018;30(5):634–44. <https://doi.org/10.1037/pas0000509> PMID: 28627924
78. Bianchini KJ, Aguerrevere LE, Curtis KL, Roebuck-Spencer TM, Frey FC, Greve KW, et al. Classification accuracy of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2)-Restructured form validity scales in detecting malingered pain-related disability. *Psychol Assess*. 2018;30(7):857–69. <https://doi.org/10.1037/pas0000532> PMID: 29072481
79. Curtis KL, Aguerrevere LE, Bianchini KJ, Greve KW, Nicks RC. Detecting malingered pain-related disability with the pain catastrophizing scale: a criterion groups validation study. *Clin Neuropsychol*. 2019;33(8):1485–500. <https://doi.org/10.1080/13854046.2019.1575470> PMID: 30957700

80. Mittenberg W, Patton C, Canyock EM, Condit DC. Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol*. 2002;24(8):1094–102. <https://doi.org/10.1076/jcen.24.8.1094.8379> PMID: 12650234
81. Puente-López E, Pina D, López-López R, Ordi HG, Bošković I, Merten T. Prevalence estimates of symptom feigning and malingering in Spain. *Psychol Inj Law*. 2023;16(1):1–17. <https://doi.org/10.1007/s12207-022-09458-w> PMID: 35911787
82. Merten T, Dandachi-FitzGerald B, Hall V, Bodner T, Giromini L, Lehrner J, et al. Symptom and performance validity assessment in European countries: an update. *Psychol Inj Law*. 2022;15(2):116–27. <https://doi.org/10.1007/s12207-021-09436-8> PMID: 34849185
83. Plohmann AM, Hurter M. Prevalence of poor effort and malingered neurocognitive dysfunction in litigating patients in Switzerland. *Z Neuropsychol*. 2017.
84. Barsky AJ, Peekna HM, Borus JF. Somatic symptom reporting in women and men. *J Gen Intern Med*. 2001;16(4):266–75. <https://doi.org/10.1046/j.1525-1497.2001.00229.x> PMID: 11318929
85. Lövestad S, Krantz G. Men's and women's exposure and perpetration of partner violence: an epidemiological study from Sweden. *BMC Public Health*. 2012;12:945. <https://doi.org/10.1186/1471-2458-12-945> PMID: 23116238
86. Umubyeyi A, Mogren I, Ntaganira J, Krantz G. Women are considerably more exposed to intimate partner violence than men in Rwanda: results from a population-based, cross-sectional study. *BMC Womens Health*. 2014;14:99. <https://doi.org/10.1186/1472-6874-14-99> PMID: 25155576
87. Nicol AL, Sieberg CB, Clauw DJ, Hassett AL, Moser SE, Brummett CM. The association between a history of lifetime traumatic events and pain severity, physical function, and affective distress in patients with chronic Pain. *J Pain*. 2016;17(12):1334–48. <https://doi.org/10.1016/j.jpain.2016.09.003> PMID: 27641311
88. Bush SS, Heilbronner RL, Ruff RM. Psychological assessment of symptom and performance validity, response bias, and malingering: official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychol Inj Law*. 2014;7(3):197–205. <https://doi.org/10.1007/s12207-014-9198-7>
89. Sweet JJ, Heilbronner RL, Morgan JE, Larrabee GJ, Rohling ML, Boone KB, et al. American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol*. 2021;35(6):1053–106. <https://doi.org/10.1080/13854046.2021.1896036> PMID: 33823750
90. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117. <https://doi.org/10.1136/bmj.c117> PMID: 20354011
91. Häuser W, Fitzcharles MA. Facts and myths pertaining to fibromyalgia. *Dialogues Clin Neurosci*. 2022.
92. Koesling D, Bozzaro C. Chronic pain as a blind spot in the diagnosis of a depressed society: on the implications of the connection between depression and chronic pain for interpretations of contemporary society. *Med Health Care Philos*. 2022;1–10.
93. Burke MJ, Silverberg ND. New framework for the continuum of concussion and functional neurological disorder. *BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine*; 2024.
94. Weiss KJ, Van Dell L. Liability for diagnosing malingering. *J Am Acad Psychiatry Law*. 2022;45:339–47.