*Journal of Clinical and Experimental Neuropsychology*, 2014
http://dx.doi.org/10.1080/13803395.2014.926863

Routledge
Taylor & Francis Group

# A Bayesian latent group analysis for detecting poor effort in a sample of cognitively impaired patients

**Alonso Ortega[1,2], Martina Piefke[1,3,4], and Hans J. Markowitsch[1,3]**

[1]Physiological Psychology, Bielefeld University, Bielefeld, Germany
[2]Escuela de Psicología, Facultad de Medicina, Universidad de Valparaíso, Valparaíso, Chile
[3]Center of Excellence "Cognitive Interaction Technology" (CITEC), Bielefeld University, Bielefeld, Germany
[4]Neurobiology and Genetics of Behavior, Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

(*Received 19 August 2013; accepted 16 May 2014*)

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

Using a Bayesian latent group analysis in a simulation design, we recently showed a high diagnostic accuracy when assessing effort in the context of malingered memory deficits. We here further evaluate our Bayesian model in a sample of cognitively impaired patients. The main analysis showed both high sensitivity and specificity, thus corroborating a high diagnostic accuracy of the model. Additional analysis showed variations on effort estimates after changes in malingering base rates. Variations affected sensitivity, but not specificity, which is in line with typical findings in malingering research. These data suggest that Bayesian analyses may complement and improve existing effort measures.

***Keywords:*** Malingering; Poor effort; Bayesian analysis; Base rates; Cognitive impairment.

According to Barrash, Suhr, and Manzel (2004), there is a clear need for reliable methods to detect exaggeration of cognitive impairment in neuropsychological assessment. In the last two decades, some interesting statistical techniques have been introduced to evaluate examinees' effort in neuropsychological practice (see Larrabee, 2008; Larrabee, Millis, & Meyers, 2008; Mossman, 2000; Mossman & Hart, 1996; Mossman, Wygant, & Gervais, 2012). For instance, Mossman and Hart (1996) first proposed the use of Bayesian analysis for effort assessment. Analyzing data from previously published malingering studies, these authors obtained precise probabilistic estimates, which allowed detection of feigned cognitive impairment (Mossman & Hart, 1996). Later, other researchers (Millis & Volinsky, 2001; Wolfe et al., 2010) used the Bayesian modeling averaging technique to determine which combination of variables from the California Verbal Learning Test (CVLT) best predicts the presence of response bias during effort assessment. More recently, Mossman et al. (2012) implemented a latent class modeling analysis in a Bayesian framework to estimate symptom validity testing (SVT) classification accuracy in the absence of a gold standard.

Along this line, we proposed the use of a Bayesian latent group analysis in effort assessment.

We thank our colleagues in the Physiological Psychology Department and the Center of Excellence "Cognitive Interaction Technology" (CITEC) Research Groups at Bielefeld University for their technical support and helpful advice. We would like to thank Simone Horstmann, patients, and staff of the Evangelic Hospital of Bielefeld (Evangelisches Krankenhaus Bielefeld) for their help and assistance when carrying out our clinical study. We would also like give special thanks to David T. R. Berry for kindly assisting us while preparing our manuscript.

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) [grant number EC 277 to M.P., H.J.M.); Becas Chile Scholarship, CONICYT (A.O.).

Address correspondence to: Alonso Ortega, Physiological Psychology, Bielefeld University, Bielefeld 33615, Germany (E-mail: alonso.ortega@uv.cl).

© 2014 Taylor & Francis

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

In a preliminary study (Ortega, Wagenmakers, Lee, Markowitsch, & Piefke, 2012), our Bayesian model successfully discriminated between experimental malingerers (i.e., healthy participants) and neurological patients with histories of traumatic brain injury (TBI) or stroke (i.e., clinical sample). In a second study, we corroborated our previous findings (Ortega, Labrenz, Markowitsch, & Piefke, 2013), which also revealed excellent diagnostic accuracy for the Bayesian latent group model, using raw scores of two well-established SVTs (i.e., Test of Memory Malingering, TOMM, Tombaugh, 1996; Word Memory Test, WMT, Green, 2005). However, moving forward, the challenge in malingering detection is to improve existing techniques and develop new methods.

In the present study, we further evaluated the accuracy of our Bayesian model for detecting poor effort in two samples of cognitively impaired patients. We performed two different analyses. The main analysis assumed a single prior malingering base rate of 8% as reported by Mittenberg, Patton, Canyock, and Condit (2002) for medical nonlitigant patients (i.e., medical, neurological, or psychiatric cases). Although Bianchini, Greve, and Love (2003) reported that malingering does occur in patients with well-documented evidence of brain damage, base rates tend to be low in this kind of sample (Mittenberg et al., 2002; Sullivan, Lange, & Dawes, 2005), and patients rarely score lower than chance level (i.e., <50% correct answers) in effort measures (Loring, Larrabee, Lee, & Meador, 2007).

Since our Bayesian latent group model incorporates relevant prior information about malingering base rates, we also performed an additional analysis. The latter aimed at observing potential variations on patient's Bayesian effort estimates after changes in the prior information about malingering base rates. This additional analysis included malingering base rates obtained from: (a) a study of Sullivan et al. (2005) in nonlitigant medical or psychiatric patients; (b) a study of adults who claimed cognitive disabilities to obtain social security benefits (Chafetz, 2008); (c) a study of suspected Alzheimer's disease in geriatric samples (Duff et al., 2011); and (d) a sample of litigating mild head injury patients (Mittenberg et al., 2002). Rosenfeld, Sands, and Van Gorp (2000) emphasized the importance of considering base rates in effort assessment. Whereas some classification accuracy indices do not vary according to the base rate of a condition (e.g., sensitivity, specificity, odds and likelihood ratios; see Streiner, 2003), there are other indices that may vary considerably depending on the condition's base rate (e.g., positive and negative predictive value (NPV), incremental predictive value, quality predictive value, Cohen's kappa coefficient, pre- and posttest odds; see Streiner, 2003). However, many researchers (Grove, 2005; Mossman, 2000; Rosenfeld et al., 2000) suggest that the accuracy of any predictive method (e.g., Bayesian model) may vary remarkably according to the base rate of the condition being predicted (e.g., effort) in any given setting.

In the main analysis we expected to observe high specificity for the Bayesian model when discriminating between honest response and symptom exaggeration patients. In the additional analysis we expected to observe some variations in the classification accuracy of the Bayesian model as long as priors about malingering base rate varied. Both main and additional analyses may underline the importance of considering malingering base rates when evaluating the presence of poor effort. Finally, the present study may also provide relevant information on the accuracy of the proposed Bayesian model to detect poor effort in clinical settings.

## METHOD

### Participants

Our sample consisted of 40 cognitively impaired patients recruited at the Evangelic Hospital of Bielefeld (Evangelisches Krankenhaus Bielefeld). These patients showed different levels of cognitive impairment after stroke ($N$ = 27; 67.5%), traumatic brain injury ($N$ = 4; 10%), and multiple sclerosis ($N$ = 3, 7.5%). Patients presenting other medical or neurological conditions constituted 15% of the total sample ($N$ = 6). These included carbon monoxide poisoning ($N$ = 1), borreliosis ($N$ = 1), epilepsy ($N$ = 1), Parkinson disease ($N$ = 1), herpes viral encephalitis ($N$ = 1), and meningitis ($N$ = 1). The mean age of the patients was 54.33 years ($SD$ = 14.27 years; mode = 60 years; minimum = 18; maximum = 85); 62.5% of patients were male and 37.5% female. Average time since injury was 4.78 months ($SD$ = 3.93 months).

Exclusion criteria included severe anterograde amnesia, global aphasia, severe visual and/or auditory deficits (e.g., visual–spatial hemineglect syndrome), severe sensorimotor deficits (e.g., apraxia), and large bilateral cerebral damage. We also excluded patients who were involved in any litigation process with health insurance companies. This decision was made to preclude the inclusion of potential "true malingerers" that might contaminate our findings.

Subsequently, patients were randomly assigned to the honest response [$N$ = 20; 13 males (65%),

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

7 females (35%); mean age = 60.05 ± 12.93 years] or the symptom exaggeration (SE) group [*N* = 20; 12 males (60%), 8 females (40%); mean age = 48.60 ± 13.48 years]. Within the honest response group the average time since injury was 5.85 months (*SD* = 4.93 months), whereas the average time since injury for the symptom exaggeration group was 3.70 months (*SD* = 2.59 months). All patients were native speakers of German.

We obtained written informed consent from all patients prior to their participation in the study. All experimental procedures were in conformity with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and also had approval from the ethics committee of the German Society of Psychology (Deutsche Gesellschaft für Psychologie; DGPs).

## Procedure

Prior to effort assessment, all patients were administered a complete neuropsychological battery. The aim of this preliminary evaluation was (a) to obtain a neuropsychological profile of each patient in different cognitive domains (see Measures section), (b) to estimate the degree of patients' cognitive impairment, and (c) to exclude extremely cognitive impaired participants. None of the participants had to be excluded from the study due to extreme cognitive deficits. In the Results section we provide a summary of the patients' neuropsychological test results.

We implemented a simulation research design, but included only cognitively impaired participants. Using a clinical sample may allow for a better extrapolation of our findings to real-life settings. Patients were randomly assigned to the honest response group and to the symptom exaggeration group. Patients in the former group were instructed to give their best during effort testing, whereas patients in the latter group were asked to exaggerate their cognitive complaints during effort assessment. All patients were administered a qualitative posttest interview, including questions to assess compliance with the assigned instructions (e.g., did you give your best while taking the tests? Did you exaggerate your cognitive problems while taking the tests? What strategies did you use in order to exaggerate your complaints? Do you think these strategies were successful? Despite the instructions you had, did you know the right answers?). According to the patients' reports, all of them followed the given instructions. Thus none of the patients had to be excluded from the study.

## Measures

### *Neuropsychological battery*

We applied measures of different cognitive domains, such as: (a) visuospatial abilities (Rey–Osterrieth Complex Figure; Osterrieth, 1944; Rey, 1941); (b) attention (d2-test; Brickenkamp & Zillmer, 1998); (c) analytic thinking (Leistungsprüfsystem, LPS; Horn, 1983); and (d) word recognition and fluency (Leistungsprüfsystem, LPS; Horn, 1983). In addition, all patients were screened for depression (Beck Depression Inventory, BDI; Beck, Steer, & Brown, 2006) and anxiety (Beck Anxiety Scale, BAS; Margraf & Ehlers, 2007). Neuropsychological data allowed characterization of the average cognitive performance of each group and precluded preexistent "between-group" significant differences concerning cognitive performance that might have affected results.

### *Visual recognition forced-choice task*

We used a visual recognition forced-choice task, which was previously developed and validated in our research group. Its content was evaluated by 12 experts who rated all items in a 3-point scale as: (a) "essential," (b) "useful but not essential," and (c) "not necessary." We calculated the content validity ratio (CVR) to select the best 50 stimuli and the best 50 distractors (see Lawshe, 1975; Wilson, Pan, & Schumsky, 2012). The classification accuracy of the visual recognition forced-choice task was validated against that of the TOMM and the WMT using the same Bayesian latent group model (Ortega et al., 2013). The visual recognition forced-choice task has both 95% sensitivity and specificity (confidence interval, CI [75, 99]) and an overall classification accuracy of 95% (area under the curve (AUC) = .95; CI [87-1]).

The visual recognition forced-choice task has one learning phase and one recognition phase. In the learning phase, a sequence of 50 simple colored drawings is exposed. The stimulus exposure time (SET) is set to 3 s per trial. During the recognition phase, every stimulus presented in the learning phase (i.e., target) is paired with a novel stimulus (i.e., distractor). The examinee's task is to recognize and select each target in a sequence of 50 "target–distractor" pairs. Examinees make their choices by pressing a keyboard computer key (i.e., A or B). The "target–distractor" pairs remain on the screen until the examinee provides an answer (i.e., target exposure time; TET). Following Hiscock and Hiscock's (1989) recommendation, a trial-by-trial feedback is provided to

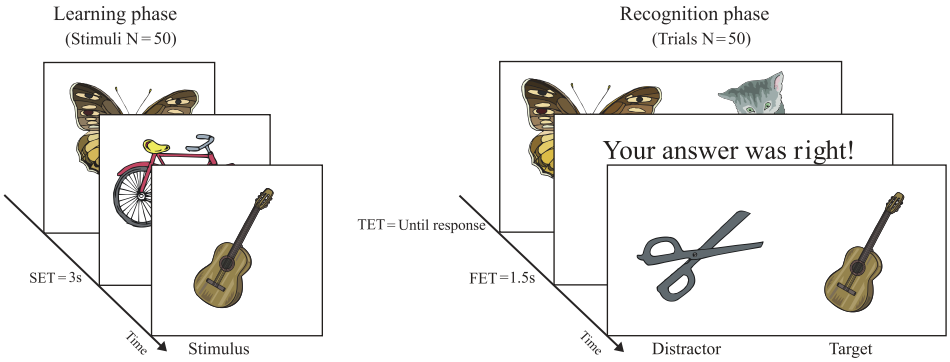


**Figure 1.** Visual recognition forced-choice task. SET = stimulus exposure time; TET = target exposure time; FET = feedback exposure time. To view a color version of this figure, please see the online issue of the Journal.

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

the examinee after each recognition trial. The feedback exposure time (i.e., FET) is 1.5 s (Figure 1). Examinees obtain one point for each successfully recognized target. Therefore, a maximum score of 50 can be obtained. The visual recognition forced-choice task was implemented and applied using the DirectRT™ software (Jarvis, 2008). Raw scores of the visual recognition forced-choice task served as input to estimate participants' probabilities of displaying poor effort using the Bayesian model.

## Data analysis

### *Descriptive and inferential statistics*

Descriptive statistics analyses were conducted to summarize the patient's performance on the different neuropsychological tests. We obtained descriptive statistics for both the honest response and the symptom exaggeration group. Before conducting inferential analyses, univariate normality was evaluated by performing a one-sample Kolmogorov–Smirnov test. Likewise, a Levene's test was applied to evaluate the homogeneity of variances. Both assumptions for parametric hypothesis testing were met for all neuropsychological variables. Finally, we conducted a $t$ test for independent samples to observe whether cognitive performance results were comparable between both groups or not, across all cognitive domains. All inferential analyses assumed a significance level $\alpha = .05$ as a criterion to reject the null hypothesis.

### *Bayesian latent group analysis*

The present Bayesian analysis allows identifying patients who are displaying poor effort during testing. A special advantage of Bayesian analysis is that prior information of a condition's base rate (e.g., malingering) can be incorporated into the model. This feature of Bayesian analyses may be especially well suited for effort assessment. Every Bayesian analysis combines prior information about the parameters (e.g., base rates) with the collected data (e.g., SVT's results). As a product, both sources of information are reflected in the so-called "posterior" distribution.

The Bayesian latent group model assumes the existence of two latent groups: the honest response and symptom exaggeration group. The model also assumes that patients from the honest response group will have a higher success rate than symptom exaggeration patients. Since the unique possible outcome per trial is "success" or "failure," we assume that our data are binomially distributed. We used a standard "beta-binomial" hierarchical model (see Lee & Wagenmakers, 2013), in which group membership will determine or—at least—exert some influence on each patient's success rate. In a beta-binomial distribution there are two parameters, "α" and "β," which are counts for "successes" and "failures," respectively. For instance, the expression beta(7, 3) represents 7 successes and 3 failures from a total of 10 trials that are "beta-binomially" distributed. As a result of combining prior information about malingering base rates and a patient's outcomes, we obtain Bayesian posterior individual classification estimates or $p(z_i|\mathrm{D})$. These $p(z_i|\mathrm{D})$ estimates represent the level of effort displayed by an examinee during effort testing. Since $p(z_i|\mathrm{D})$ estimates are expressed in probabilistic terms, results can be interpreted in a very intuitive way. Values closer to 0 represent lower probabilities that an examinee gave poor effort during testing. Values closer to 1 represent higher probabilities that an examinee had poor effort during testing (see Ortega et al., 2012).

**TABLE 1**
Descriptive and inferential statistics on neuropsychological performance for the honest response and symptom exaggeration groups

| *Variable* | *Group* | *Mean* | *SD* | *t* | *p* | *Sig.* |
|---|---|---|---|---|---|---|
| Time since injury (TSI)/months | HR | 5.85 | 4.74 | 1.779 | .083 | *ns* |
| | SE | 3.70 | 2.59 | | | |
| Visuospatial abilities/ROCF test | HR | 52.60 | 8.71 | –1.378 | .176 | *ns* |
| | SE | 56.00 | 6.78 | | | |
| Attention/d2 test | HR | 90.80 | 13.15 | –1.451 | .155 | *ns* |
| | SE | 96.45 | 11.41 | | | |
| Analytic thinking/LPS test | HR | 49.00 | 10.21 | –1.086 | .284 | *ns* |
| | SE | 52.00 | 6.96 | | | |
| Word recognition and fluency/LPS test | HR | 49.55 | 6.27 | –1.279 | .209 | *ns* |
| | SE | 52.17 | 6.67 | | | |

*Note.* HR = honest response group; SE = symptom exaggeration group; ROCF = Rey–Osterrieth Complex Figure; LPS = Leistungsprüfsystem; sig. = significance; *ns* = non significance.

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

The model was implemented using the Markov chain Monte Carlo (MCMC) sampling method (e.g., Gamerman & Lopes, 2006; Gilks, Richardson, & Spiegelhalter, 1996), which provides accurate estimates of posterior probability distributions. All Bayesian analyses were conducted using the WinBUGS software program (Lunn, Spiegelhalter, Thomas, & Best, 2009). Additional information about Bayesian inference can be found in O'Hagan and Forster (2004).

## RESULTS

### Descriptive and inferential statistics

The honest response group (mean = 48.55 ± 1.64) exceeded the symptom exaggeration group (mean = 34.65 ± 4.67) in the visual recognition forced-choice task ($t = 12.559$; $p < .05$).

The symptom exaggeration group was younger than the honest response group (t = 2.741; $p < .05$). However, no statistically significant differences were observed between both groups across all cognitive domains: (a) visuospatial abilities, (b) attention, (c) analytic thinking, and (d) word recognition and fluency.

The average time since injury between both groups was equivalent ($t = 1.779$, $p = .083$, *ns*) as was also observed for anxiety levels ($t = 1.343$, $p = .187$, *ns*). In contrast, significant between-group differences ($t = 2.463$, $p < .05$) were observed on the measure of depression (Table 1).

### Bayesian latent group model: Main analysis

The main analysis allowed for an accurate differentiation between patients in the honest response and symptom exaggeration groups (Table 2). The sensitivity of the Bayesian model was estimated to be 90% (95% CI [69, 97]), whereas the specificity was estimated to be 100% (95% CI [83.9, 100]).

**TABLE 2**
Individual posterior classification probabilities for the honest response and symptom exaggeration groups

| | *Group* | | | | | | |
|---|---|---|---|---|---|---|---|
| | *Honest response* | | | | *Symptom exaggeration* | | |
| | | $p(z_i\|D)$ | | | | $p(z_i\|D)$ | |
| $Patient_i$ | *Raw score* | *Mean* | *SD* | $Patient_i$ | *Raw score* | *Mean* | *SD* |
| $P_1$ | 50 | .000 | .000 | $P_{21}$ | 35 | .999 | .037 |
| $P_2$ | 50 | .000 | .000 | $P_{22}$ | 31 | 1 | .000 |
| $P_3$ | 49 | .000 | .006 | $P_{23}$ | 40 | .836 | .371 |
| $P_4$ | 49 | .000 | .000 | $P_{24}$ | 21 | 1 | .000 |
| $P_5$ | 49 | .000 | .000 | $P_{25}$ | 39 | .932 | .251 |
| $P_6$ | 50 | .000 | .009 | $P_{26}$ | 37 | .990 | .098 |
| $P_7$ | 46 | .003 | .056 | $P_{27}$ | 33 | 1 | .016 |
| $P_8$ | 50 | .000 | .000 | $P_{28}$ | 36 | .995 | .069 |
| $P_9$ | 49 | .000 | .006 | $P_{29}$ | 38 | .975 | .156 |
| $P_{10}$ | 49 | .000 | .006 | $P_{30}$ | 36 | .996 | .066 |
| $P_{11}$ | 50 | .000 | .000 | $P_{31}$ | 35 | .999 | .029 |
| $P_{12}$ | 48 | .000 | .015 | $P_{32}$ | 30 | 1 | .000 |
| $P_{13}$ | 46 | .003 | .056 | $P_{33}$ | 36 | .996 | .066 |
| $P_{14}$ | 50 | .000 | .000 | $P_{34}$ | 29 | 1 | .000 |
| $P_{15}$ | 50 | .000 | .000 | $P_{35}$ | 39 | .933 | .250 |
| $P_{16}$ | 46 | .003 | .054 | $P_{36}$ | 33 | 1 | .014 |
| $P_{17}$ | 49 | .000 | .000 | $P_{37}$ | 30 | 1 | .000 |
| $P_{18}$ | 45 | .014 | .120 | $P_{38}$ | 38 | .971 | .167 |
| $P_{19}$ | 49 | .000 | .006 | $P_{39}$ | 40 | .825 | .380 |
| $P_{20}$ | 47 | .001 | .027 | $P_{40}$ | 37 | .991 | .096 |

*Notes.* Assuming a prior on malingering base rate of 8% (Mittenberg et al., 2002). $p(z_i|D)$ = individual posterior classification probability.

The positive predictive value (PPV) was .99 (95% CI [.82, 1]), and the NPV was .91 (95% CI [.72, .98]). The area under the receiver operating characteristic curve (AUC) was .95 (95% CI [.87, 1]). According to Hosmer and Lemeshow's

(2000) criteria, the observed AUC value represents outstanding overall diagnostic accuracy.

### Bayesian latent group model: Additional analysis

Additional analysis incorporated different priors about malingering base rates reported by four published studies, ranging from 4% to 41%.

Within the symptom exaggeration group we observed major variations in some patient's $p(z_i|D)$ mean values when priors about malingering base rates varied (e.g., $P_{23}$ and $P_{39}$; see Table 3). When assuming a malingering base rate of 4% (i.e., Sullivan et al., 2005) sensitivity was equal to 70% (95% CI [45.7, 88]), whereas specificity was equal to 100% (95% CI [84, 100]). Of note, when priors about malingering base rate increased to 12% (Chafetz, 2008), 33% (Duff et al., 2011), and 41% (i.e., Mittenberg et al., 2002), both sensitivity and specificity increased to 100% (95% CI [83.9, 100]).

Within the honest response group, individual $p(z_i|D)$ mean values showed only minor variations. Consequently, the specificity of the Bayesian model can be estimated to be 100% (95% CI [84, 100]) regardless of variations on malingering base rates (see Table 3).

## DISCUSSION

The present study aimed at further evaluating the accuracy of a Bayesian latent group model for effort assessment. Our previous studies (Ortega et al., 2012, 2013) showed high classification accuracy for the Bayesian latent group model using simulation research designs. In the present experiment no healthy participants were included, which —in our view—increased its external validity. Regardless of the age differences observed between both groups, no differences were found in any of the cognitive domains evaluated. Therefore, results allow attributing patients' performance mainly to the role instructions and subsequently to the level of effort displayed during testing.

As described above, this study included two analyses. First, we performed a main analysis assuming a prior about malingering base rate of 8% (Mittenberg et al., 2002). Second, we conducted an additional analysis assuming different priors on malingering base rates (Chafetz, 2008; Duff et al., 2011; Mittenberg et al., 2002; Sullivan et al., 2005). As hypothesized, the main analysis allowed ruling in the presence of poor effort within the symptom exaggeration group, showing high sensitivity levels (i.e., 90%). Within the honest response group, specificity levels were excellent (i.e., 100%)—a result consistent with our latest findings (Ortega et al., 2013). Results of the main analysis support the use of Bayesian models for effort assessment in clinical settings.

To suggest the presence of poor effort, we assumed $p(z_i|D)$ mean values higher than 90%. However, this value might vary depending on the prior information about malingering base rates. Here, our main suggestion is to assume a subjective Bayesian interpretative approach (Hájek, 2003) in which the closer to 1 a probability is, the higher is our certainty about the presence of poor effort. However, we would also like to provide clinicians some additional guidelines for Bayesian posterior individual estimates interpretation.

First, clinicians can assume that any diagnostic method will be best used to rule out a condition when priors about base rates of a condition are low (Streiner, 2003). Additional analysis showed a significant decrement in sensitivity (i.e., 70%) when priors on malingering base rates were low (i.e., 4%), whereas specificity was not affected. In that case, we recommend interpreting $p(z_i|D)$ mean values higher than 90% as indicators of poor effort. Values in the range between 85% and 90% may be considered only as "suspicious" of poor effort, and those lower than 85% should be ruled out. Thus, in the case of rather low priors on malingering base rates (e.g., 1% to 10%), our recommendation is to be more conservative when interpreting the Bayesian effort estimates in order to avoid false positives. This guideline is in good agreement with actual standards in malingering research (see Iverson, 2007).

Second, when base rates are high any diagnostic method will be best used to rule in a condition (Streiner, 2003). In that case, almost any result over 90% can be interpreted as definite indicator of poor effort. Additional analysis demonstrated that even moderate values of prior on malingering base rates (e.g., 33% to 41%) led to both 100% sensitivity and specificity. The higher the prior on malingering base rate, the higher our confidence in the classification accuracy.

A third guideline is incorporating raw scores of one—or more—additional SVT (e.g., TOMM) into the Bayesian model, obtaining the $p(z_i|D)$ mean value and then comparing the posterior probability achieved using each single test (see Ortega et al., 2012). A last suggestion is that combining traditional decision rules (i.e., cutoff scores) with probabilistic estimates might help clinicians to diminish the risk of misclassification that is present when decisions rely solely on a

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

**TABLE 3**
Individual classification probabilities for the honest response and symptom exaggeration groups assuming different priors about malingering base rates

| | | *Prior on malingering base rates* | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sullivan et al. (2005)*[a] | | *Chafetz (2008)*[a] | | *Duff et al. (2011)*[a] | | *Mittenberg et al. (2002)*[a] | |
| | | *Nonlitigant medical*[b] | | *Social security claimants*[b] | | *Suspected Alzheimer*[b] | | *Litigant mild head injury*[b] | |
| | | *4%*[c] | | *12%*[c] | | *33%*[c] | | *41%*[c] | |
| | | $p(z_i\|D)$ | | $p(z_i\|D)$ | | $p(z_i\|D)$ | | $p(z_i\|D)$ | |
| *Group* | *Participant*$_i$ | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* | *Mean* | *SD* |
| Honest response | $P_1$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .006 |
| | $P_2$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .006 |
| | $P_3$ | .000 | .000 | .000 | .000 | .000 | .014 | .000 | .014 |
| | $P_4$ | .000 | .000 | .000 | .000 | .000 | .012 | .000 | .017 |
| | $P_5$ | .000 | .000 | .000 | .000 | .000 | .014 | .000 | .012 |
| | $P_6$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| | $P_7$ | .001 | .037 | .006 | .077 | .025 | .156 | .036 | .187 |
| | $P_8$ | .000 | .000 | .000 | .000 | .000 | .006 | .000 | .009 |
| | $P_9$ | .000 | .000 | .000 | .011 | .000 | .006 | .000 | .011 |
| | $P_{10}$ | .000 | .000 | .000 | .006 | .000 | .015 | .000 | .019 |
| | $P_{11}$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| | $P_{12}$ | .000 | .011 | .000 | .016 | .001 | .032 | .001 | .035 |
| | $P_{13}$ | .001 | .033 | .005 | .073 | .022 | .147 | .036 | .187 |
| | $P_{14}$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .009 |
| | $P_{15}$ | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| | $P_{16}$ | .001 | .030 | .006 | .077 | .023 | .151 | .036 | .186 |
| | $P_{17}$ | .000 | .006 | .000 | .000 | .000 | .014 | .000 | .014 |
| | $P_{18}$ | .005 | .071 | .025 | .155 | .088 | .284 | .131 | .337 |
| | $P_{19}$ | .000 | .000 | .000 | .000 | .000 | .011 | .000 | .012 |
| | $P_{20}$ | .000 | .018 | .001 | .036 | .005 | .069 | .007 | .084 |
| Symptom exaggeration | $P_{21}$ | .990 | .100 | .999 | .024 | 1 | .011 | 1 | .006 |
| | $P_{22}$ | 1 | .017 | 1 | .012 | 1 | .000 | 1 | .000 |
| | $P_{23}$ | .608 | .488 | .912 | .284 | .975 | .156 | .983 | .129 |
| | $P_{24}$ | 1 | .000 | 1 | .000 | 1 | .000 | 1 | .000 |
| | $P_{25}$ | .783 | .412 | .964 | .187 | .992 | .092 | .995 | .074 |
| | $P_{26}$ | .950 | .219 | .996 | .065 | .999 | .029 | 1 | .023 |
| | $P_{27}$ | .998 | .049 | 1 | .009 | 1 | .000 | 1 | .009 |
| | $P_{28}$ | .978 | .147 | .998 | .045 | 1 | .012 | 1 | .017 |
| | $P_{29}$ | .894 | .308 | .987 | .112 | .997 | .051 | .999 | .034 |
| | $P_{30}$ | .976 | .154 | .998 | .040 | 1 | .023 | 1 | .011 |
| | $P_{31}$ | .990 | .097 | .999 | .027 | 1 | .016 | 1 | .009 |
| | $P_{32}$ | 1 | .014 | 1 | .006 | 1 | .000 | 1 | .000 |
| | $P_{33}$ | .979 | .144 | .998 | .040 | 1 | .023 | 1 | .014 |
| | $P_{34}$ | 1 | .006 | 1 | .000 | 1 | .000 | 1 | .000 |
| | $P_{35}$ | .785 | .411 | .966 | .183 | .991 | .092 | .994 | .077 |
| | $P_{36}$ | .997 | .052 | 1 | .006 | 1 | .000 | 1 | .000 |
| | $P_{37}$ | 1 | .016 | 1 | .000 | 1 | .000 | 1 | .000 |
| | $P_{38}$ | .889 | .315 | .988 | .110 | .997 | .057 | .998 | .042 |
| | $P_{39}$ | .594 | .491 | .909 | .288 | .971 | .168 | .984 | .125 |
| | $P_{40}$ | .953 | .213 | .997 | .059 | .999 | .034 | .999 | .026 |

[a]Study. [b]Sample. [c]Base rate.

cutoff score. Of course, malingering determination must always be complemented with information provided by other sources (see Boone, 2007; Larrabee, Greiffenstein, Greve, & Bianchini, 2007; Slick, Sherman, & Iverson, 1999). Furthermore, we do not recommend the use of Bayesian posterior individual estimates as a single decision rule to determine poor effort. It should be assumed that each evaluation is unique.

In any case, we believe that reporting a patient's effort levels in probabilistic terms is an advantage

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

of using Bayesian approaches for neuropsychological assessment. The previous statement is consistent with the idea of considering the nature of effort as a continuous variable that varies in magnitude rather than a discrete variable that can be arbitrarily categorized using cutoff scores (Bigler, 2012; Iverson, 2010; Walters et al., 2008; Walters, Berry, Lanyon, & Murphy, 2009). Along this line, Bigler (2012) warned about the risk of misclassification when determining the presence of any condition (i.e., poor effort) using a single cutoff score. In Bigler's (2012) view, the misclassification risk is particularly sensitive in the case of examinees that score in the range between "above chance" (i.e., >50% correct answers) and "below the cutoff score" (i.e., near-pass scores; see Bigler, 2012). For example, patients $P_{23}$ and $P_{39}$ (Table 2) obtained "near-pass" scores. Nonetheless, they also have verified cognitive impairment, which complicates the decision-making process.

Our findings raise two relevant issues: (a) the importance of considering malingering base rates when assessing effort, and (b) the risk of misclassification when determining poor effort based only on cutoff scores. Test results should not be interpreted independently from the base rate of a condition in a given population. A decade ago, Rosenfeld et al. (2000) suggested including information about malingering base rates in effort assessment. From our perspective, these considerations have not received enough attention.

The present study did not include patients with suspected motivations to underperforming during effort testing or involved in litigation processes with health insurance companies. Even though this was a methodological decision, including "at-risk" patients (i.e., potential malingerers) should be considered in further studies. For instance, testing our Bayesian model using a known-groups research design could be an alternative. Another potential limitation is that our main analysis assumed a base rate obtained approximately a decade ago (Mittenberg et al., 2002). It is known that base rates tend to vary after demographic changes or variations on the incidence of a condition in a particular context. Nevertheless, Loring et al. (2007) stated that malingering base rates tend to be low in cognitively impaired patients who are not seeking monetary compensation or any other kind of benefits.

In Ortega et al. (2012) we provided a supplementary file that contains a step-by-step description of how to conduct a Bayesian latent group analysis. Recently, Lee and Wagenmakers (2013) devoted a book chapter about latent mixture models that includes a brief but complete description of a latent group assessment of malingering. This kind of analysis does not require advanced statistical knowledge and, therefore, can be performed by any interested clinician or researcher. Information about malingering base rates is available in the scientific literature (e.g., Ardolf, Denney, & Houston, 2007; Chafetz, 2008; Larrabee, 2007; Mittenberg et al., 2002), which allows clinicians to conduct their own analyses. Therefore, we encourage malingering researchers and practitioners to explore the potential benefits of Bayesian models in effort assessment.

Even though our results are not conclusive support for use of Bayesian analysis as a standard procedure in effort assessment, we remain optimistic. In our view, Bayesian estimates can complement results obtained with traditional effort measures and other additional sources of information and thus improve decision-making processes when assessing effort.

## REFERENCES


Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist*, *21*, 899–916.

Barrash, J., Suhr, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the auditory verbal learning test (AVLTX): Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology*, *26*, 125–140.

Beck, A. T., Steer, R. A., & Brown, G. L. (2006). *Beck Depression Inventory* (2nd ed.). Frankfurt: Harcourt Test Services GmbH.

Bianchini, K. J., Greve, K. W., & Love, J. M. (2003). Definite malingered neurocognitive dysfunction in moderate/severe traumatic brain injury. *Clinical Neuropsychologist*, *17*, 574–580.

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 632–642.

Boone, K. B. (2007). A reconsideration of the Slick et al. criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). New York, NY: Guilford Press.

Brickenkamp, R., & Zillmer, E. (1998). *The d2 test of attention*. Seattle, WA: Hogrefe & Huber.

Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *Clinical Neuropsychologist*, *22*, 529–546.

Duff, K., Spering, C. C., O'Bryant, S. E., Beglinger, L. J., Moser, D. J., Bayless, J. D., … Scott, J. G. (2011). The RBANS Effort Index: Base rates in geriatric samples. *Applied Neuropsychology*, *18*, 11–17.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd ed.). New York, NY: Chapman & Hall/CRC.

Downloaded by [Alonso Ortega] at 08:57 11 June 2014

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice (interdisciplinary statistics)*. London: Chapman & Hall.

Green, P. (2005). *Word Memory Test for Windows: Test manual* (Rev. ed.). Edmonton: Green's Publishing.

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, *61*, 1233–1243.

Hájek, A. (2003). What conditional probability could not be. *Synthese*, *137*, 273–323.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, *11*, 967–974.

Horn, W. (1983). *Leistungsprüfsystem* [Performance assessment system] (2nd ed.). Göttingen: Hogrefe Verlag.

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley Interscience.

Iverson, G. L. (2007). Identifying exaggeration and malingering. *Pain Practice*, *7*, 94–102.

Iverson, G. L. (2010). Detecting exaggeration, poor effort, and malingering in neuropsychology. In A. Horton, Jr. & L. C. Hartlage (Eds.), *Handbook of forensic neuropsychology* (2nd ed., pp. 91–135). New York, NY: Springer.

Jarvis, B. G. (2008). DirectRT (Version v2008.1.13) [Computer software]. New York, NY: Empirisoft Corporation.

Larrabee, G. J. (2007). Malingering, research designs, and base rates. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 1–13). Oxford: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *Clinical Neuropsychologist*, *22*, 666–679.

Larrabee, G. J., Greiffenstein, M., Greve, K., & Bianchini, K. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–372). Oxford: Oxford University Press.

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2008). Sensitivity to brain dysfunction of the Halstead–Reitan vs an ability-focused neuropsychological battery. *Clinical Neuropsychologist*, *22*, 813–825.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563–575.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: A practical course*. Cambridge: Cambridge University Press.

Loring, D. W., Larrabee, G. J., Lee, G. P. , & Meador, K. J. (2007). Victoria Symptom Validity Test performance in a heterogeneous clinical sample. *Clinical Neuropsychologist*, *21*, 522–531.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Margraf, J., & Ehlers, A. (2007). *Beck Angst-Inventar–BAI* [Beck anxiety inventory–BAI]. Frankfurt: Harcourt Test Services.

Millis, S. R., & Volinsky, C. T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology*, *23*, 809–828.

Mittenberg, W., Patton, C., Canyock, E., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094–1102.

Mossman, D. (2000). The meaning of malingering data: Further applications of Bayes' theorem. *Behavioral Sciences & the Law*, *18*, 761–779.

Mossman, D., & Hart, K. J. (1996). Presenting evidence of malingering to courts: Insights from decision theory. *Behavioral Sciences & the Law*, *14*, 271–291.

Mossman, D., Wygant, D. B., & Gervais, R. O. (2012). Estimating the accuracy of neurocognitive effort measures in the absence of a "gold standard." *Psychological Assessment*, *24*, 815–822.

O'Hagan, A., & Forster, J. J. (2004). *Kendall's advanced theory of statistics: Vol. 2B. Bayesian inference* (2nd ed.). London: Arnold.

Ortega, A., Labrenz, S., Markowitsch, H. J., & Piefke, M. (2013). Diagnostic accuracy of a Bayesian latent group analysis for the detection of malingering-related poor effort. *Clinical Neuropsychologist*, *27*, 1019–1042.

Ortega, A., Wagenmakers, E.-J., Lee, M. D., Markowitsch, H. J., & Piefke, M. (2012). A Bayesian latent group analysis for detecting poor effort in the assessment of malingering. *Archives of Clinical Neuropsychology*, *27*, 453–465.

Osterrieth, P. A. (1944). Le test du copie d'une figure complexe [The test of copying a complex figure]. *Archives des Psychologie* (Paris), *30*, 206–356.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathic traumatique [The psychological examination in cases of traumatic encepholopathy]. *Archives des Psychologie* (Paris), *28*, 286–340.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*, 349–359.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist*, *13*, 545–561.

Streiner, D. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, *81*, 209–219.

Sullivan, K., Lange, R. T., & Dawes, S. (2005). Methods of detecting malingering and estimated symptom exaggeration base rates in Australia. *Journal of Forensic Neuropsychology*, *4*, 49–70.

Tombaugh, T. N. (1996). *Test of Memory Malingering TOMM*. New York, NY: Multi-Health Systems.

Walters, G. D., Berry, D. T. R., Lanyon, R. I., & Murphy, M. P. (2009). Are exaggerated health complaints continuous or categorical? A taxometric analysis of the health problem overstatement scale. *Psychological Assessment*, *21*, 219–226.

Walters, G. D., Rogers, R., Berry, D. T. R., Miller, H. A., Duncan, S. A., McCusker, P. J., … Granacher, R. P. (2008). Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment*, *20*, 238–247.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, *45*, 197–210.

Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test–II (CVLT–II). *Clinical Neuropsychologist*, *24*, 153–168.