Using the Center
for Epidemiologic Studies Depression Scale (CES-D)
to Classify Participants of a Questionnaire Survey
with Regard to Their Individual Inclination to Respond
at Random

Maciej Górkiewicz[1]

[1] Faculty of Health Sciences, Jagiellonian University Medical College, Krakow, Poland

Abstract. In this study, the well-known CES-D depression scale was considered as a prototypical tool for assessing an individual's inclination to respond at random. It was postulated that the set of the responses obtained from the participants of a questionnaire survey could be divided into three classes: the honest responses class; the pure random class, characterized by the same probability of each admissible score; the smart random class, characterized by the same distribution of the probability of scores as in the honest responses class. A two-step classifying procedure was recommended. According to this procedure, the first step should be to separate the pure random responses from the others. As needed, at the second step the smart random responses should be separated from the honest responses. The CES-D scale consists of sixteen direct questions and four reverse questions, each question with four admissible score values, 0, 1, 2, or 3. The material for the analyses includes all of the 4 294 967 296 possible arrangements of the four scores for the sixteen direct questions, and all of the 256 arrangements of the four scores for the four reverse questions, and all of the possible $49 \cdot 13 = 637$ pairs of the possible values of the mean scores from the sixteen direct questions and from the four reverse questions, respectively. It was assumed, without loss of generality, that the honest responses to the sixteen direct questions lead to the resulting distribution: $P = (8/16; 5/16; 2/16; 1/16)$, for a score equal to 0, 1, 2, or 3, respectively. The original iterative procedure was used to compute the exact distributions of the mean scores in the pure random class, and in the smart random class. The discriminant ability to distinguish between the two classes was characterized with the Receiver Operating Characteristic (ROC) curve. In result, the excellent discriminant properties of the CES-D were confirmed. The area under the ROC curve was estimated equal to AUC = 0.98 with standard error SE = 0.004 for the screening of the pure random responses from all of the others, and equal to AUC = 0.87 with standard error SE = 0.006 for the purpose of distinguishing the smart random responses from the honest ones.

Maciej Górkiewicz

Introduction

This study aimed to develop some practical recommendations regarding how to protect a questionnaire survey against unintentional liars. The proposed methodology and the obtained results can be useful for all who use questionnaires and are interested in confirming the credibility of the obtained data (Chmiel et al., 2012; Górkiewicz, 2014), but can be particularly useful for questionnaire developers who have an interest in methodological challenges resulting from the use of survey methods (Bereziewicz et al., 2012; Collins, 2003; Dojka et al., 2003; Garb et al, 2011; Górkiewicz, 2010; Hambleton et al., 1999; Peña-Sánchez et al., 2011; Sijtsma, 2009).

The need for efficient assessment methods has been growing in recent years in all domains of the health sciences. Questionnaire surveys are commonly applied to various medical investigations, most commonly to epidemiological studies, but also to clinical and therapeutic practice. Moreover, besides the standard clinical measures of health, focused on morphology and organ function, in recent years patient-reported opinions regarding courses of therapy are increasingly used in clinical settings, and then, after hospitalization (Chmiel et al., 2010; Gniadek et al., 2010; Wiklund, 2004). For that reason, unrevealed lies and simple mistakes in the data should be considered as a worse kind of disturbance than any undisguised lack of needed data. Unfortunately, the answers of the liars, at first glance, usually look quite similar to the honest ones, so they can be hard to uncover. Therefore, the problem of how far patients' opinions may be trusted has its special importance, and well-grounded methods for estimating the probability of this – that some patients' reports can result from random answering rather than from honest conviction – are required urgently.

It is commonly accepted that incorrect statements and untrue suppositions are very frequent events that may be caused by various intentions and psychological mechanisms. For instance, in the experiments on false memory, performed with experimental groups drawn from the general public, a false recall was reported, usually by about 40–55% of participants for the words suggested to them indirectly, and by about 75–80% of participants for the other words, which were not in fact presented to them (Górkiewicz et al., 2007; Roediger et al., 1995). Similar proportions were found in the other investigations on false memory and the analogous phenomena (Chmiel et al., 2011; Gerrie et al., 2006; West et al., 2014).

In general, the unintentional liars do not pay attention to whether their answers will provide a profitable result for the examination, but they

generate their answers by chance, under some unknown scale of preferences. Thus, with respect to the aims of the questionnaire survey, unintentional lies should be considered as random noise, with unknown statistical properties.

The Center for Epidemiologic Studies Depression Scale (CES-D) was created originally with the aim to measure depressive symptomatology in the general population, across a broad range of subject ages (Radloff, 1977, 1991). Nevertheless, for many years it has also been used extensively for groups with various health conditions (Hann et al., 1999; Helmes et al., 1998; Miller et al., 2008; Park et al., 2014).

The CES-D depression scale is simple and quick to administer. It includes only twenty simple questions, consisting of sixteen direct questions, such as question 6: I felt depressed, or question 14: I felt lonely, and four reverse questions, such as question 12: I was happy, or question 16: I enjoyed life. A respondent is asked to choose and indicate only one of four admissible answers to each question. Responses are based on the frequency of depressive feelings and behaviours during the week prior to questionnaire administration. The frequency "Rarely or none of the time (less than 1 day)" is scored with 0 points; the frequency "Some or a little of the time (1–2 days)" is scored with 1 point; "Occasionally or a moderate amount of the time (3–4 days)" is scored with 2 points; and "Most or all of the time (5–7 days)" is scored with 3 points. Therefore, the assessment usually takes no more than about 5–10 minutes. The resulting score is computed as the sum of all twenty scores, but after converting the reverse scores into the direct shape: $0 \rightarrow 3$; $1 \rightarrow 2$; $2 \rightarrow 1$; and $3 \rightarrow 0$.

For therapeutic purposes, two diagnostic thresholds are recommended: a resulting score less than 15 from the twenty questions (that is the mean score equal to $15/20 = 0.75$) for the nonappearance of depression; a resulting score between 15 and 21 (that is the mean score between 0.75 and 1.05) for mild to moderate depression, and a resulting score over 21 (that is the mean score over 1.05) indicating the possibility of major depression (Counselling Resource Research Staff, 2012). However, it should be noted that about 20% of those who achieve a high CES-D score, in fact, do not meet the full psychiatric criteria for major or clinical depression (Counselling Resource Research Staff, 2012). For epidemiological studies, a single threshold is usually applied: a resulting score less than 16 (that is the mean score less than 0.8) for nonappearance of depression; and a resulting score over 16 (that is the mean score over 0.8) indicating the possibility of depression (Dojka et al., 2003; Fava, 1983; Radloff, 1977, 1991). The CES-D has been standardized to many non – English

– speaking populations, and also to the Polish population (Dojka et al., 2003; Fava, 1983; Orme et al., 1988). The English version of the CES-D with proper instructions is easily accessed on-line (Counselling Resource Research Staff, 2012).

The Receiver Operating Characteristic (ROC) is recommended for comparing classifiers, as it does not merely summarize performance at a single arbitrarily selected decision threshold, but across all possible decision thresholds (Górkiewicz, 2009; Zweig et al., 1993). The ROC plots the sensitivity versus one minus the specificity, with the diagnostic variable as the parameter on the ROC curve. In other words, the ROC plots the true positive ratio, denoted TP, versus the false positive ratio, denoted FP. The overall performance of a classifier is usually evaluated with the value of the area under the ROC curve (AUC). An ideal classifier should achieve the area under the curve AUC = 1.0, but a random classifier should achieve AUC = 0.5 (Hanley et al., 1982, 1983; Obuchowski, 1998). With regard to the two-step classifying procedure under examination in this study, at the first step of this procedure sensitivity can be defined as probability that a response created with a pure random generator will be recognized correctly as a pure random response, rather than a smart random or honest response. Then, at the second step, sensitivity can be defined as the probability that an honest response will be recognized correctly as an honest response, but not as a smart random response. Analogously, at the first step, specificity can be defined as the probability that a smart random or honest response will not be recognized wrongly as a pure random response; and at the second step as the probability that an honest response will not be recognized wrongly as a smart random response (Fawcett et al., 2005; Hand et al., 2001).

The random answering of a questionnaire can be considered as a sequence of random choices. Let us imagine that in a single experiment the series of the K independent drawings of a single number X_m from the sequence $\{X_m = X_0 + (m-1) \cdot D\}$ is finalized, at each k-th drawing, with known probabilities $\{P_{k.m}\}$, where: $k = 1, 2, \ldots, K$; $m = 1, 2, \ldots, M$. It is assumed that for each fixed k: $k = 1, 2, \ldots, K$; the condition (1) is fulfilled.

$$\Sigma(P_{km}|\text{fixed } k) = 1; \text{for each fixed } k, \tag{1}$$

where: $k = 1, 2, \ldots, K$; $m = 1, 2, \ldots, M$.

As a result of a single experiment, a sequence of K drawn number Xs arises. The resulting distribution of the sum of the drawn Xs across all the possible results of a single experiment ought to be calculated.

The needed data includes: the number K of drawings; an arithmetic sequence $\{X_m = X_0 + (m-1) \cdot D\}$; a matrix of probabilities $\{P_{k.m}\}$, where: $K > 0$; $D \neq 0$; $k = 1, 2, \ldots, K$; $m = 1, 2, \ldots, M$.

Any arithmetic sequence $\{X_m = X_0 + (m-1) \cdot D\}$, where: $D \neq 0$; $m = 1, 2, \ldots, M$, may be easy transformed with formula (2) into a sequence of M successive natural numbers from 0 to $M - 1$. This transformation is reversible, and consequently, any sum Y_K of the K natural numbers can be transformed into a suitable sum of K numbers from sequence $\{X_m\}$ with formula (3). Then, for any assumed number of K drawings from the sequence of M successive natural numbers from 0 to $M - 1$, the set of all possible sums includes all integers from $Y_K = 0$ to $Y_K = \max(Y_K)$, where $\max(Y_K)$ can be calculated with formula (4).

$$S_m = (X_m - X_0)/D, \tag{2}$$

where: $D \neq 0$; $m = 1, 2, \ldots, M$

$$\Sigma_K(X_m) = K \cdot X_0 + D \cdot Y_K, \tag{3}$$

where: X_0, D – parameters of a sequence $\{X_m\}$, $\Sigma_K(X_m)$ – sum of the K numbers freely chosen from sequence $\{X_m\}$, Y_K – a random variable, equal to sum of the freely chosen K natural numbers less then $M - 1$.

$$\max(Y_K) = (M - 1) \cdot K \tag{4}$$

With regard to formulas (2), (3) and (4), the initial task may be formulated as follows: given numbers K, M, the matrix of probabilities $\{P_{km}\}$, and the sequence $\{X_m\}$, first calculate the exact distribution of probability for the set of all possible sums Y_K, from $Y_K = 0$ to $Y_K = \max(Y_K)$, and then transform the values of Y_K with formula (3).

All of the needed calculations are rather straightforward. For the $K = 1$ the required distribution $\{P(Y_1)\} = (P(Y_1 = 0), P(Y_1 = 1), \ldots, P(Y_1 = M))$ follows directly from the known data, (5).

$$\{P(Y_1)\} = \{P_{k=1.m}\} = (P_{1.1}, P_{1.2}, \ldots, P_{1.M}), \tag{5}$$

where: $m = 1, 2, \ldots, M.$; random variable Y_1 is defined on sequence $0, 1, \ldots, M - 1$.

Then, for any assumed $K > 1$ the required distribution $\{P(Y_K)\}$ may be calculated with formulas (6) and (7) on the base of a previously calculated distribution $\{P(Y_{K-1})\}$.

Maciej Górkiewicz

$$P(Y_K = i) = \Sigma g(m) \cdot (P(Y_{K-1} = i - m + 1) \cdot P(S_m = m - 1)), \quad (6)$$

where $m = 1, 2, \ldots, M$

$$\{P(Y_K)\} = (P(Y_K = 0), P(Y_K = 1), \ldots, P(Y_K = (M-1) \cdot K)), \quad (7)$$

where: $i = 0, 1, \ldots, \max(Y_K)$; $g(m) = 1$ if $0 \leq i - m + 1 \leq (M-1) \cdot (K-1)$, and $g(m) = 0$ otherwise.

In other words, the required distribution $\{P(Y_K)\}$ may be calculated in $K - 1$ steps, starting from the known distribution $\{P(Y_1)\}$, by $\{P(Y_2)\}$, $\{P(Y_3)\}, \ldots$, step by step up to $\{P(Y_K)\}$. The proposed iterative procedure (5), (6) and (7) is illustrated with use of some freely assumed distributions $\{P(Y_2)\}$ and $\{P(_{k=3.m})\}$, in Table 1.

What about the number of possible results from a single experiment of the K independent drawings of a single number S from sequence $S = 0, 1, \ldots, M-1$? It is easy to notice that at least, in the case of each probability $P_{km} = 0$ or 1, the number K of the possible arrangements of K numbers can occur; and at most, in the case of each $P_{km} > 0$, the number M^K of the possible results can occur; where: $k = 1, 2, \ldots, K$; $m = 1, 2, \ldots, M$. For instance, for $M = 4$ and $K = 16$ the number of possible results can arrive at $M^K = 4\,294\,967\,296$ of the possible arrangements. The proposed iterative procedure creates the opportunity to calculate the required distribution $\{P(Y_{16})\}$ in $K - 1 = 15$ steps, with use of a table such as Table 1.

It was postulated that the set of responses obtained from the participants of a questionnaire survey can be divided into three classes: the honest responses class, pure random class, and the smart random class. The CES-D scale, considered in this study, consists of $K = 16$ direct questions and $L = 4$ reverse questions, each question with $M = 4$ admissible scores, equal to 0, 1, 2, or 3. It was assumed that in the honest class, the probabilities that underly the responses to the $K = 16$ direct questions are in relation (8) with the probabilities that underly the responses to the $L = 4$ reverse questions. On the contrary to this assumption, it was assumed that in both random classes the score of the any CES-D question was drawn with the same distribution of probability, without any distinction between the direct scores and the reverse scores. In the pure random class, the probabilities are the same for each admissible score, but in the smart random class these probabilities can differ from one another (Table 2).

$$P_m|(\text{honest answer to direct question}) =$$
$$= P_{4-m}|(\text{honest answer to reverse question}), \quad (8)$$

where: $m = 1, 2, 3$, or 4; $\Sigma P_m = 1$; $\Sigma P_{4-m} = 1$; P_m denotes the probability $P(\text{score} = m - 1)$.

Table 1. Computing distribution $\{P(Y_3)\}$ on the base of the known distribution $\{P(Y_2)\}$

	S	$S = 0$	$S = 1$	$S = 2$	$S = 3$
$Y = 0$	0.250	$Y + S = 0$ $P = 0.1250$	$Y + S = 1$ $P = 0.0781$	$Y + S = 2$ $P = 0.0313$	$Y + S = 3$ $P = 0.0156$
$Y = 1$	0.313	$Y + S = 1$ $P = 0.1563$	$Y + S = 2$ $P = 0.0977$	$Y + S = 3$ $P = 0.0391$	$Y + S = 4$ $P = 0.0195$
$Y = 2$	0.223	$Y + S = 2$ $P = 0.1113$	$Y + S = 3$ $P = 0.0696$	$Y + S = 4$ $P = 0.0278$	$Y + S = 5$ $P = 0.0139$
$Y = 3$	0.141	$Y + S = 3$ $P = 0.0703$	$Y + S = 4$ $P = 0.0439$	$Y + S = 5$ $P = 0.0176$	$Y + S = 6$ $P = 0.0088$
$Y = 4$	0.055	$Y + S = 4$ $P = 0.0273$	$Y + S = 5$ $P = 0.0171$	$Y + S = 6$ $P = 0.0068$	$Y + S = 7$ $P = 0.0034$
$Y = 5$	0.016	$Y + S = 5$ $P = 0.0078$	$Y + S = 6$ $P = 0.0049$	$Y + S = 7$ $P = 0.0020$	$Y + S = 8$ $P = 0.0010$
$Y = 6$	0.004	$Y + S = 6$ $P = 0.0020$	$Y + S = 7$ $P = 0.0012$	$Y + S = 8$ $P = 0.0005$	$Y + S = 9$ $P = 0.0002$
$Y\|K = 2$	$P(Y)$	$P(S) = 0.5$	$P(S) = 0.3125$	$P(S) = 0.125$	$P(S) = 0.0625$
Total.P	1.000	0.5	0.3125	0.125	0.0625

The values of Y_2 are shown in the first column from the left, named $Y|K = 2$, and the values of $P(Y_2)$ in the second column, named $P(Y)$;

The $M = 4$ successive natural numbers, from $S = 0$ to $S = M - 1 = 3$, are shown in the first row from the top, named S, and the values of $\{P_{(k=3.m)}\}$, denoted $P(S)$, in the second row;

In the remaining cells, the sum $Y + S$ indicates the component of the $Y_3 = Y + S$, respectively, and the P denotes the probability of this component; the value of this probability was computed as a product $P(Y) \cdot P(S)$ associated with the cell under consideration;

The resulting probability $P(Y_3 = i)$ may be calculated as the sum of the probabilities P shown at all cells with index $i = Y + S$; for instance, the probability $P(Y_3 = 1)$ may be calculated as sum $P(Y_3 = 1) = 0.1563 + 0.0781 = 0.2344$ (not shown in the table).

Table 2. The assumed relations between distributions of scoring in the honest and random classes of responses to the direct versus reverse questions of the CES-D scale

class of response	Scoring of the direct questions				Scoring of the reverse questions			
	$S = 0$	$S = 1$	$S = 2$	$S = 3$	$S = 0$	$S = 1$	$S = 2$	$S = 3$
pure random	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
smart random	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4
honest					P_4	P_3	P_2	P_1

In the first row from the top, all four possible scores to the CES-D questions are shown: $S = 0$, $S = 1$, $S = 2$, $S = 3$; The probabilities that a particular score will be chosen under a considered strategy of scoring (pure random, smart random, honest scoring) are shown in the remaining cells of the table.

Maciej Górkiewicz

In this study, a two-step classifying procedure was recommended. At the first step of the procedure, the smart random response class and the honest response class should be considered jointly with respect to the set of $K = 16$ direct questions, because of the indistinguishable (from assumption) distributions of scoring. Thus, at the first step, the pure random responses should be separated from the two remaining classes. Here, the mean score of the $K = 16$ direct questions is recommended as the discriminant variable. As needed, at the second step the smart random responses should be separated from the honest responses. Here, the difference between the mean score at the $K = 16$ direct questions versus the mean score at the $L = 4$ reverse questions is recommended as the discriminant variable.

The problem of choosing the cut-off values according to some real-world circumstances were not included to the scope of this study. However, it should be mentioned that with respect to a small set of data ($K = 16$, $L = 4$), and the skewed distributions of the honest and smart random responses, the problem of choosing the proper statistical test should be considered here with particular caution (Górkiewicz et al., 2012).

Material and Methods

The material for the analyses includes the three data sets well-defined only by the structure of the CES-D questionnaire: the $K = 16$ direct questions and $L = 4$ reverse questions, each question with $M = 4$ admissible scores, equal to 0, 1, 2, or 3. The first data set includes the number $M^K = 16^4 = 4\,294\,967\,296$ of arithmetic mean values of the scores occurring in each possible arrangement of the $M = 4$ admissible scores at the $K = 16$ direct questions. The second data set includes the number $M^L = 4^4 = 256$ of arithmetic mean values of the scores occurring in each possible arrangement of the $M = 4$ admissible scores at the $L = 4$ reverse questions. The third data set includes the numbers $(M \cdot K + 1) \cdot (M \cdot L + 1) = 49 \cdot 13 = 637$ pairs of the possible values of the mean scores from the sixteen direct questions and from four reverse questions, respectively.

It was assumed, without loss of generality, that the honest responses to the sixteen direct questions lead to the resulting distribution: $\{P_i\} = (8/16; 5/16; 2/16; 1/16)$, for a score equal to 0, 1, 2, or 3, respectively. In result, Table 2 took the form of Table 3.

The original iterative procedure proposed in this study (5), (6) and (7) was used to compute the exact distributions of the sum of scores in the

Table 3. The assumed distributions of probability of scoring at the honest and random classes of responses on the direct versus reverse questions of the CES-D scale

class of response	Scoring of the direct questions				Scoring of the reverse questions			
	$S = 0$	$S = 1$	$S = 2$	$S = 3$	$S = 0$	$S = 1$	$S = 2$	$S = 3$
pure random	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
smart random	0.5	0.3125	0.125	0.0625	0.5	0.3125	0.125	0.0625
honest					0.0625	0.125	0.3125	0.5

In the first row from the top, all four possible scores to the CES-D questions are shown: $S = 0$, $S = 1$, $S = 2$, $S = 3$; The probabilities that a particular score will be chosen under a considered strategy of scoring (pure random, smart random, honest scoring) are shown in the remaining cells of the table.

pure random class, and in the smart random class. Then, for a number of questions equal to 4 or 16, respectively, the sum of scores were considered as discrete random variables.

The discriminant ability to distinguish between the two classes was characterized with the Receiver Operating Characteristic (ROC) curve. All analyses were made with well-known procedures, among other details described by Górkiewicz (2009), on the base of works published by Hanley et al. (1983) and Obuchowski (1998). All of the computations are quite straightforward, so they may be made with any spreadsheet software, e.g. with Excel for Windows (Górkiewicz et al., 2001). If needed, useful Java programs for calculating ROC curves with proper instructions are easily available on-line (Eng, n.d.). There the needed data may be typed or be entered from programs such as Microsoft Excel or Word (Eng, n.d.).

Results

The excellent discriminant properties of the CES-D were observed with regard to both steps of the two-step classifying procedure under examination in this study. For the first step, the area under the ROC curve was estimated equal to AUC = 0.98 with standard error SE = 0.004 and 95%CI confidence interval from 0.97 to 0.99 (Table 4). For the second step, the area under the ROC curve was estimated equal to AUC = 0.87 with standard error SE = 0.006, and 95%CI confidence interval from 0.75 to 0.99 (Table 5).

Table 4. Task of screening pure random response to the CES-D scale

$m16$	$P\|$(pure random)	$P\|$(honest or smart random)	TP	FP	AUC
$m16 \leq 0.5$	0.000	0.167	0.0002	0.1671	0.00007
$0.5 < m16 \leq 1.0$	0.046	0.723	0.047	0.890	0.0596
$1 < m16 \leq 1.5$	0.498	0.109	0.544	0.999	0.5507
$1.5 < m16 \leq 2$	0.428	0.001	0.972	1.000	0.3535
$m16 > 2$	0.028	0.000	0.999	1.000	0.0161
Total P, AUC	1.00	1.00	–	–	$0.980 < 1$

$m16$ – mean score of the 16 direct questions of the CES-D scale;

$P\|$(pure random) – probability of occurrence of a specified value of $m16$ under pure random scoring;

$P\|$(honest or smart random) – probability of occurrence of a specified value of $m16$ under honest or smart random scoring;

TP – true positive rate: probability that a pure random response will obtain $m16$ below a cut-off equal to a specified value of $m16$;

FP – false positive rate: probability that an honest response or a smart random response will obtain $m16$ below a cut-off equal to a specified value of $m16$;

AUC – area under the ROC curve that is below a curve $FP = f(TP)$

Table 5. Task of screening smart random response to the CES-D scale

$m16$–$m4$	$P\|$honest	$P\|$smart random	TP	FP	AUC
> 0.25	0.0011	0.34	0	0	0.00001
0.25	0.0005	0.05	0.000	0.048	0.0002
0.125	0.0026	0.15	0.002	0.391	0.001
0	0.0015	0.05	0.003	0.438	0.001
-0.125	0.0075	0.13	0.006	0.584	0.002
-0.25	0.0039	0.04	0.010	0.621	0.005
< -0.25	0.98	0.25	0.017	0.750	0.860
Total P, AUC	1.00	1.00	–	–	$0.870 < 1$

$m16$–$m4$ – the difference between a mean score of the 16 direct questions and a mean score of the 4 reverse questions of the CES-D scale;

$P\|$(honest) – probability of occurrence of a specified value of a difference $m16$–$m4$ under honest scoring or smart random scoring;

$P\|$(smart random) – probability of occurrence of a specified value of a difference $m16$–$m4$ under smart random scoring;

TP – true positive rate: probability that an honest response will obtain a difference of $m16$–$m4$ below a cut-off equal to the specified value of $m16$–$m4$;

FP – false positive rate: probability that a smart random response will obtain a difference of $m16$–$m4$ below a cut-off equal to the specified value of $m16$–$m4$;

AUC – area under ROC curve, that is below a curve $FP = f(TP)$

Discussion

In this paper it has been shown that the CES-D scale can be used to confirm patients' credibility, owing to the proper proportion of opposite questions among the twenty items of this questionnaire, and owing to the shifted population norm of the measured phenomenon (inclination to depression). Thus, it was proved that it may be considered as a prototypical tool for creating a new questionnaire with the built-in ability to assess an individual's inclination to respond at random among the participants of a questionnaire survey.

This paper can be considered as a direct continuation of the study performed by Górkiewicz (2014), with some reference to several further studies, namely those carried out by Chmiel et al. (2011, 2012, 2014), Gniadek et al. (2010), Górkiewicz (2010), Górkiewicz et al. (2005).

In practice, the standard way of confirming individuals' credibility consists of a repeated examination method, in which individuals are administered the same set of questions on two different occasions (Górkiewicz et al., 2005). In spite of the popularity of this approach, in using a repeated examination method, the investigator needs to take into account the possibility of ordinary disturbances. First, this method is appropriate only if the phenomenon that the scale measures is known to be stable over the time interval between examinations. However, it also must be remembered that if the time interval between examinations is too short, then the individuals under examination can simply remember their previous answers. Then, the motivation to give honest responses to the repeated questions may be distinctly lower than at the first occasion. For these reasons, some other related techniques were developed (Chmiel et al., 2011, 2012, 2014).

The iterative procedure proposed in this study can be considered as the generalization of the procedure proposed by Górkiewicz (2014) for the issue of varying distributions in a particular series when drawing a single number from a given arithmetic sequence.

In practice, the distribution of the mean score from CES-D examinations is usually approximated with a log-normal distribution of probability. The use of the proposed iterative procedure creates an opportunity to avoid extensive computations, without any raw simplifications.

Maciej Górkiewicz

Conclusions

In this paper, it was demonstrated that a single patient's response to all twenty questions of the Center for Epidemiologic Studies Depression Scale (CES-D) can be used not only with the aim of evaluating the patient's inclination to depression, but also with the aim of assessing his or her inclination to answer at random. Moreover, it was proved that a single patient's response to all twenty questions of the Center for Epidemiologic Studies Depression Scale (CES-D) can be used not only to differentiate a pure random strategy of scoring from an honest scoring strategy, but also to distinguish an honest scoring from a smart random scoring (which attempts to imitate an honest scoring).

R E F E R E N C E S

Bereziewicz, W., & Górkiewicz, M. (2012). Jak dużo a priori w a posteriori: poznanie naukowe z zastosowaniem metod statystyki. *Cogitatum*, 2, 1–9, Retrieved from http://filozof.uni.lodz.pl/knf/cogitatum/numer2/cog2bereziewicz.pdf.

Chmiel, I., & Górkiewicz, M. (2010). The scope of acceptance by patients motherly and friendly style of nurse's supporting behaviour in palliative care. *Problemy Pielęgniarstwa*, 18(4), 399–405.

Chmiel, I., & Górkiewicz, M. (2011). Method for evaluating an individual inclination to guess Yes-or-No answers in case of a diversity of opinion at group of trustworthy responders. In L. Bobrowski, T. Burzykowski, J. Doroszewski, & C. Enachescu (Eds.). *114-th ICB Seminar – VIII-th International Seminar: Statistics and Clinical Practice*, 8–11 June 2011 (pp. 59–61). Warsaw, Poland: IBIB PAN.

Chmiel, I., & Górkiewicz, M. (2012). How much credible are the responses obtained from an individual respondent in a non-repeated questionnaire survey: looking for practical methods with a statistical support. *Studies in Logic Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 29(42), 55–73. Retrieved from http://logika.uwb.edu.pl/studies/vol42.html.

Chmiel, I., & Górkiewicz, M. (2014). Measuring physical ability among convalescents on the base of patient-reported quality of life: statistical support for credibility of an individual patient's report. In L. Bobrowski, U. Mansmann, & C. Enachescu (Eds.). *Statistics and Clinical Practice. Lecture Notes of the ICB Seminar., ICB Polish Academy of Sciences*, June 2014 (pp. 84). Warsaw, Poland: IBIB PAN.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 229–238.

Counselling Resource Research Staff. (2012). Welcome to the Center for Epidemiologic Studies Depression Scale (CES-D), a Screening Test for Depression. Retrieved from http://counsellingresource.com/lib/quizzes/depression-testing/cesd/.

Dojka, E., Górkiewicz, M., & Pająk, A. (2003). Psychometric value of CES-D scale for the assessment of depression in Polish population. *Psychiatria Polska*, 37(2), 281–292.

Eng, J. (n.d.). ROC analysis: web-based calculator for ROC curves. Retrieved from http://www.jrocfit.org.

Fava, G. A. (1983). Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. *J Clin Psychol.*, 39(2), 249–251.

Fawcett, T, & Flash, P. A. (2005). A Response to Webb and Ting's On the Application of ROC Analysis to Predict Classification Performance Under varying Class Distributions. *Machine Learning*, 58(1), 33–38.

Garb, H. N, Wood, J. M., & Fiedler, E. R. (2011). A Comparison of Three Strategies for Scale Construction to Predict a Specific Behavioral Outcome. *Assessment*, 18(4), 399–411.

Gerrie, M. P., Belcher, & L. E., Garry, M. (2006). Mind the gap: false memories for missing aspects of an event, *Applied Cognitive Psychology*, 20(5), 689–696.

Gniadek, A., Kozicka, M., & Górkiewicz, M. (2010). Nieoczekiwane, pośrednie i pozorne powiązania czynników jakości życia osób starszych. In Z. Sienkiewicz, W. Fidecki, & G. Wójcik (Eds.). *Wybrane Determinanty Pielęgniarstwa*, Część II (pp. 118–127). Warsaw, Poland: Oficyna Wydawnicza Warszawskiego Uniwersytetu Medycznego.

Górkiewicz, M. (2009). Using Propensity Score with Receiver Operating Characteristics (ROC) and Bootstrap to Evaluate Effect Size in Observational Studies. *Biocybernetics and Biomedical Engineering*, 29(4), 41–61.

Górkiewicz, M. (2010). Kognitywistyczne spojrzenie na skale psychometryczne: kiedy trudne jest łatwiejsze od łatwego... In A. Pohl, M. Goc, M. Kittel, T. Konik, & M. Siedlecka (Eds.). *Rocznik Kognitywistyczny, IV* (pp. 81–86). Krakow, Poland: Wydawnictwo Uniwersytetu Jagiellońskiego.

Górkiewicz, M. (2014). Detecting patient's random answers to questionnaire with opposite items and the shifted population norm. In L. Bobrowski, U. Mansmann, & C. Enachescu (Eds.). *Statistics and Clinical Practice. Lecture Notes of the ICB Seminar., ICB Polish Academy of Sciences*, June 2014 (pp. 127–130). Warsaw, Poland: IBIB PAN.

Górkiewicz, M., Ciszek, E., & Szczygieł, A. (2005). Ocenianie kompetencji eksperta na podstawie procedury powtórnego rozmieszczania. In Z. J. Brzeziński (Ed.). *Horyzonty Metodyczne w Epidemiologii* (pp. 125–129). Warsaw, Poland: Komisja Epidemiologii i Zdrowia Publicznego PAN.

Górkiewicz, M., & Gniadek, A. (2012). The practicality of any nonparametric statistical procedure should be confirmed thoroughly with regard to the data distribution under study. *Studies in Logic Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 29(42), 27–42. Retrieved from http://logika.uwb.edu.pl/studies/vol42.html.

Górkiewicz, M., & Kołacz, J. (2001). *Statystyka Medyczna. Podejście praktyczne przy zastosowaniu programu MS Excel.* Krakow, Poland: Wydawnictwo Uniwersytetu Jagiellońskiego.

Górkiewicz, M., & Kreiner, D. S. (2007). Gender Differences in Creating False Memory under the DRM Paradigm. *European Epi-Marker*, 11(2), 6–12.

Hambleton, R. K., & Patsula, L. (1999). Increasing the Validity of Adapted Test: Myths to be Avoided and Guidelines for Improving Test Adaptation Practices. *J Appl Testing Technology*, 1(1), 1–30.

Hand, D. J., & Till, R. J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, 45(2), 171–186.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristics (ROC) curve. *Radiology*, 143, 29–36.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristics curves derived from the same cases. *Radiology*, 148, 839–843.

Hann, D., Winter, K., & Jacobsen, P. (1999). Measurement of depressive symptoms in cancer patients: Evaluation of the Center for Epidemiological Studies Depression Scale (CES-D). *Journal of Psychosomatic Research*, 46, 437–443.

Helmes, E., & Nielson, W. R. (1998). An examination of the internal structure of the Center for Epidemiological Studies-Depression Scale in two medical samples. *Personality and Individual Differences*, 25, 735–743.

Miller, W. C., Anton, H. A., & Townson, A. F. (2008). Measurement properties of the CESD scale among individuals with spinal cord injury. *Spinal Cord.*, 46(4), 287–292.

Obuchowski, N. A. (1998). Sample size calculations in studies of test accuracy, *Stat. Meth. Med. Res.*, 7, 371–392.

Orme, J. G., Reis, J., & Herz, E. (1988). Factorial and discriminant of the Center for Epidemiologic Studies Depression (CES-D) Scale. *Med. Care*, 26, 775–789.

Park, N. S., Jang, Z., Lee, B. S., Ko, J. E., & Chiriboga, D. A. (2014). The Impact of Social Resources on Depressive Symptoms in Racially and Ethnically Diverse Older Adults: Variations by Groups With Differing Health Risks. *Research on Aging*, 36, 322–342.

Peña-Sánchez, J. N., Domagala, A., Górkiewicz, M., Targowska, M., & Oleszczyk, M. (2011). Adapting a tool in Poland for the measurement of the physicians' career satisfaction. *Journal of Family Medicine Topics*, 12(1), 58–65. Retrieved from http://pmr.org.pl/.

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl Psych Measurement*, 3, 385–401.

Radloff, L. S. (1991). The Use of the Center for Epidemiological Studies of Depression Scale in Adolescents and Young Adults. *J Youth Adoles*, 20(2), 149–166.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814.

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha, *Psychometrika*, 74(1), 107–120.

West, R. L., & Stone, K. R. (2014). Age Differences in Eyewitness Memory for a Realistic Event. *J Gerontol B Psychol Sci Soc Sci*, 69(3), 338–347.

Wiklund, I. (2004). Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life, *Fundamental & Clinical Pharmacology*, 18(3), 351–363.

Zweig, M. H., & Campbell, G. (1993). Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, 39(4), 561–577.