Civil Liability for Clinical Misdiagnosis of Suicidal Intention: Procedure and Guidelines to Minimize Fatal Diagnostic Error

Raúl Quevedo-Blasco\textsuperscript{a}, María José Pérez\textsuperscript{b}, Alejandro Guillén-Riquelme\textsuperscript{a,c}, and Tatiana Hess\textsuperscript{a,c}

\textsuperscript{a}Centro de Investigación Mente, Cerebro y Comportamiento (CIMCYC), Universidad de Granada, Spain; \textsuperscript{b}Universidad Internacional de Valencia, Spain; \textsuperscript{c}Comenius University in Bratislava, Slovakia

**ABSTRACT**

Background/Objectives: A fatal diagnostic error of suicidal intention, i.e., an error implying death or serious injuries (i.e., incapacitating, chronic injury) to the patient, may have civil liability (punishable error) for the clinician. The Scale for Suicidal Ideation (SSI) is the reference psychometric instrument used to mediate suicidal intention. A meta-analytical review was designed with the aim of estimating the true reliability of the SSI in general and in different settings (moderators) with the aim of correcting unreliability raw scores. Method: A total of 90 primary studies reporting SSI's reliability (internal consistency) was found in the literature, yielding a total of 92 effect sizes. Bare-bones meta-analysis of correlation coefficients correcting effect by sampling error were run. Results: The results showed an overall mean true internal consistency of .8904, 95% CI [.8878, .8930], meaning that 42.6% of the population standard deviation is error and 18.11% of an individual's measure is error. Additional estimations (moderators) of SSI's reliability for gender (men: .8873, women: .8808) adaptation version (English: .9212, Korean: .9052, Chinese: .8402, Italian: .9163, Persian: .8612), and population (subclinical: .8769, general: .9230, mental illness: .9040) were obtained. All mean true estimations were under the desirable standard for applied settings where critical decisions are made, .95. Furthermore, for populations with risk of suicide, such as prison inmates and militaries, mean true reliability could not be computed as \( k \) was insufficient. Conclusion: Implications of true reliabilities obtained for the estimation of individuals' true scores and population standard deviations are discussed. Examples of computation of true scores to minimize fatal diagnosis errors were performed for both known reliability and unknown reliability settings (e.g., risk populations).

**RESUMEN**

La responsabilidad civil por el diagnóstico clínico erróneo de la intención suicida: procedimiento y pautas para minimizar el error diagnóstico mortal

**Palabras clave:**

Fiabilidad
Metaanálisis
Error punible
True score

**Antecedentes/Objetivo:** Un error diagnóstico mortal de intención suicida, es decir, un error que implique la muerte o lesiones graves (es decir, incapacitantes, crónicas) para el paciente, puede tener responsabilidad civil (error punible) para el clínico. La Escala de Ideación Suicida (SSI) es el instrumento psicométrico de referencia utilizado para medir la intención suicida. En este estudio se diseñó una revisión metaanalítica para medir la fiabilidad real de la SSI en general y en diferentes ámbitos (moderadores), con el fin de corregir las puntuaciones brutas por la falta de fiabilidad de la medida. Método: En la búsqueda de estudios se encontró un total de 90 estudios primarios que trataban de la fiabilidad (consistencia interna) del SSI, lo que arrojó un total de 92 tamaños del efecto. Se realizaron meta-análisis del tipo bare-bone corrigiendo el efecto por el error de muestreo. Resultados: Se obtuvo una consistencia interna verdadera media global de .8904, IC 95% [ .8878, .8930], es decir, el 42.6% de la desviación estándar de la población es error y el 18.11% de la medida de un sujeto es error. Se obtuvieron estimaciones adicionales (moderadores) de la fiabilidad del SSI para el género (hombres: .8873, mujeres: .8808), versión de adaptación (inglés: .9212, coreano: .9052, chino: .8402, italiano: .9163, persa: .8612) y población (subclínica: .8769, general: .9230, enfermedad mental: .9040). Todas las estimaciones verdaderas medias estaban por debajo del estándar deseable para entornos aplicados donde se toman decisiones críticas, .95. Además, para las poblaciones de riesgo de suicidio, como reclusos y militares, no se pudo calcular la fiabilidad media verdadera, ya que \( k \) era insuficiente. Conclusiones: Se discuten las implicaciones de la fiabilidad verdadera obtenida para la estimación de las puntuaciones verdaderas de los individuos y las desviaciones típicas de las poblaciones. Se ejemplifica el cálculo de puntuaciones verdaderas para minimizar los errores fatales en el diagnóstico, tanto para entornos de fiabilidad conocida, como desconocida (por ejemplo, poblaciones de riesgo).
“Globally, 703,000 people die by suicide every year” (World Health Organization [WHO, 2021b, p.1]). Nowadays, suicide is a major public health issue. Suicide is defined as a ‘continuum which includes ideation, plans, attempts and, consummated suicide behaviours’ (WHO, 2014). According to WHO (2021b), most deaths by suicide are by men and around half of the total completed suicides occurred before the age of fifty. In relation to the country’s income level, this information affirmed that 77% of deaths by suicide occurred in low- and middle-income countries.

Some of the previous variables have already been handled in the last investigations. For example, a systematic review by Cano-Montalbán and Quevedo-Blasco (2018) aimed to analyse suicidal behaviour worldwide, analysing the research before 2015. About the conclusions, it is worth highlighting the increased prevalence of suicide committed in men (55.8%) and a higher percentage of ideation, planning, and attempts in the case of women. Likewise, they found a higher percentage of completed suicide in adults than in minors. Finally, considering some sociodemographic variables, they concluded that the highest suicide rate was found among marginalized populations with low education levels and unemployed people (75%).

In addition to these sociodemographic variables, other studies were focused on another type of population that might have a higher tendency to suffer from suicidal ideation and other related behaviours. For example, Bentley et al. (2016) conducted a meta-analysis, whose objective was to analyse the role of mental health disorders as risk factors for suicidal ideation. They found a strong link of anxiety disorder and post-traumatic stress disorder with suicidal thoughts. Similarly, other studies relate suicidal behaviour with professional sectors that have high-risk jobs (e.g., police, firefighters, military). Di Nota et al. (2020) brought out this problem after conducting a study on the Canadian police, that showed a higher risk of developing some kinds of mental illness, including suicidal behaviour. Similar results were obtained in the study conducted in Australia (Kyon et al., 2020) with a sample wider than 14,688 subjects and several professions (such as police officers, firefighters, and rescue and ambulance personnel).

It is not surprising that many professional and social sectors have demanded that governments and institutions pay attention to this problem. Internationally, the WHO (2021a) drafted a suicide prevention guide, named ‘Live Life’. A good approach to the phenomenon, the guide starts with raising the awareness that this is a public health problem, one which concerns all of us and requires a joint response.

Returning to the definition of suicide previously provided (WHO, 2014), we must emphasize that suicide is a continuum that encompasses completed suicide as well as ideation and previous plans. It is understandable that suicidal behaviour has received the most attention on account of its salience, but we cannot lose sight of the other factors. In fact, authors like Jobes and Joiner (2019) have drawn attention to ideation, which though may go unnoticed for being more ephemeral and less concrete, may be the basis of suicide attempts.

One of the most frequently used scales measuring aspects of suicidal behaviour is Beck et al.’s (1979) Scale for Suicidal Ideation (SSI). The precursor of the SSI was the Suicide Intent Scale (SIS; Beck et al., 1974), which also served to develop the SSI. It should be noted that the pilot study, carried out for the construction of the SSI, was conducted in a clinical context to predict the risk of suicide; therefore, the original validation of the scale was by means of clinical sampling.

As regards the character of the SSI, it consists of 19 items, with three numerical response alternatives from 0 to 2, in accordance with the degree of intensity. The total sum of the items ranges from 0 to 38 points, where the higher the number the greater the severity. The target population is people 17 years and over. The content of the items ranges from one’s attitude towards oneself, to plans and wishes to die. The seminal version of the SSI (Beck et al., 1979) showed an adequate reliability, $\alpha = .89$, for experimental studies (mean comparisons) but under the desirable standard of .95 for applied settings implying critical decisions (e.g., clinical diagnosis; Nunnally, 1978).

Any assessment must be constructed following an accurate procedure to ensure the reliability criterion (Muñiz & Fonseca-Pedrero, 2019; Podsakoff et al., 2003). However, reliability does not refer to a stable and unchanged concept, but to a concept that will change in accordance with the population to which it is applied. Since the reported reliability varies among studies, the reported is an estimation. Thus, the true reliability should be computed (Schmidt & Hunter, 2015).

As health professionals, psychologists play an important role in suicidal ideation. A correct approach to suicide behaviour requires education and knowledge about the phenomenon as well as a correct assessment of it. A fatal diagnostic error, i.e., an error implying death or serious injuries to the patient (i.e., incapacitating, chronic injury), may have civil liability (punishable error) for the therapist (Arce, 2017). Thus, misdiagnosis of suicidal intention may imply a fatal outcome or serious damage for the patient. To cope with the misdiagnosis of the severity of suicidal intention (non-appreciation of severity) by therapists, we count on meta-analyses of the reliability of the worldwide reference psychometric instrument measure of suicidal intention, the SSI (Beck et al., 1979), in order to provide therapists with scientific criteria and guidelines to correct measurement error for unreliability.

**Method**

**Design and Procedure**

In the present paper, the first step was to follow the instructions of the REGEA protocol (Sánchez-Meca et al., 2021), which was specially designed for conducting meta-analytical studies or revisions. In addition, the guidelines of the original PRISMA protocol by Page et al. (2021) have been followed, including the recommendations by Rubio-Aparicio et al. (2018). To ensure that there were no previous systematic reviews or meta-analyses on this topic, we searched for scientific papers covering the topic under discussion in order to preclude any similar or duplicated work. Our search stemmed from PROSPERO (registration no: CRD42022345575), and the databases that are particularly stated in the methodology. As a result, in Google Scholar search engine (which indexes scholarly literature) we encountered the abovementioned study published by King et al. (2014). However, after a detailed analysis, we concluded that their search strategy does not match neither our criteria nor PRISMA’s protocol. As for this, we opted for further and unique research.

The research process began on February 25 2022 and lasted till March 6 2022. As regards the search strategy, it was different from the traditional one (equation with keywords, truncators, and Booleans). In this study, collected data were obtained via an atypical strategy. First, Beck et al’s (1979) Assessment of suicidal intention: The Scale for Suicide Ideation original article was found in each of the bibliographic databases. It should be noted that it became essential to adjust the title of the original article on account of its variations of diacritic symbols, such as hyphens or double dots. After that, all cited articles were downloaded to our program database.

The databases used were Web of Science (Core Collection), Scopus, PubMed, and the ProQuest platform (including Psychology Database, Criminology Collection, PsycArticles, PsycExtra, PsycInfo, PsycTest, Sociology Collection, Criminal Justice Database, ProQuest Dissertations, and Theses Global, International Bibliography of the Social Sciences, Health and Medical Collection, and Nursing and Allied Health Database). Time range spans the period from 1979 (the year the SSI was created) to December 31 2021. Studies selection criteria are given below.
Inclusion and Exclusion Criteria

The studies that we had gathered had to meet the following criteria: a) samples of adults were exclusively used (over 18 years of age), b) all studies had to provide Cronbach's alpha index of the SSI, c) the metadata of the publications had to be either in English or Spanish (prevailing languages), d) the design of the studies had to be experimental, quasi-experimental, descriptive, correlational, or instrumental, e) the type of publication, i.e., the focus was on articles published between 1979 and 2021, and f) the gender and health status of participants. It follows that the samples might include people suffering from physical and mental health problems.

On the other hand, all studies meeting the following exclusion criteria were excluded: a) different types of documents (institutional reporting, conference proceedings, etc.), b) population age (under 18 years), and c) people with intellectual disabilities, were eliminated.

Codification of Variables

In the first phase, the primary tasks were to identify variables for setting out selection criteria. This procedure was carried out independently by two reviewers after the screening process and the extraction of final articles. Coded variables were of two types—the first for the identification of the samples in each article to describe the final global sample, the next effect size variables, or those necessary for this, together with moderators. For the specific extraction data, it was necessary to draft a codebook with the following variables: (a) year of publication, (b) country of study, (c) sample size of participants, (d) participant gender, (e) population type (general population or clinical sample), (f) mean, and (g) standard deviation (SD) of age, (h) maximum and minimum age, and (i) SSI Cronbach's alpha.

Agreement of Codification (Codification Fidelity)

To keep quality, reliability, and objectivity of the work and coding process by two reviewers, a program was applied to calculate Cohen's kappa coefficient, in the case of categorical variables. Thus, values between .69 and 1.0, with a mean of .84 (SD = .12), were considered a good and excellent interpretation. On the other hand, agreement on continuous variables was estimated with the intraclass correlation, also with high agreement, between .65 and 1.0 (M = .93, SD = .15).

Data Analysis

The procedure consisted of a bare-bones correlational meta-analysis correcting effect by sampling error (Schmidt & Hunter, 2015). Effect sizes were taken directly from primary studies. As for this, sample size weighted mean effect size (\( \bar{R} \)); sample size weighted observed variance of correlations (\( S^2_{\text{obs}} \)); sample size weighted observed standard deviation of correlations (\( SD_{\text{obs}} \)); variance attributed to sampling error variance (\( S^2_{\text{err}} \)); percent of observed variance accounted by sampling error variance (% var); variance of correlations after removing sampling error variance (\( S^2_{\text{res}} \)); standard deviation of correlations after removing sampling error variance (\( SD_{\text{res}} \)); 95% CIr: 95% confidence interval for \( r \); and 80% credibility interval (80% CI) were estimated. The 95% CI of \( r \) was computed for heterogeneous studies (% var < 60). The limits of the credibility interval establish the minimum (lower limit) expected effect for 90% of the studies, and the maximum effect (upper limit) expected for 90% of studies. If artefactual errors (% var) explains the bulk of the variance (60%, rule for correction by sampling error only: Hunter et al., 1982), then the non-explained variance is not systematic, describing the mean true correlation. Counterwise, if the variance explained by artifacts is lesser than 60%, then the studies are heterogeneous, i.e., the variance is explained by moderators than sampling error.

The comparison of meta-analytic mean true effects between two moderators or a moderator with a test value (criterion) was performed with the difference between two correlations (\( q \)), and then zeta for the difference with the associated probability (\( \alpha \)) (Arias et al., 2020).

Results

Study Selection

A total of 5,784 documents was obtained from the database search mentioned above. After removing duplicate documents (3,375), other 341 documents were eliminated during the second screening phase as they did not meet the inclusion criteria of this work. Consequently, the following step consisted of elimination by title/abstract (752 articles).

In the phase of full-text review, 101 studies were included; the remaining studies were removed for not fulfilling the purpose of this meta-analysis. Finally, 90 articles (with 92 effect sizes) met selection criteria (see Figure 1).

Descriptive Characteristics of the Studies

A total of 90 studies was selected from which 92 effect sizes were obtained. The total sample of all studies was 24,119 participants. Most of the studies (86.95%) comprised both genders, men and women, whereas only 9.78% consisted only of women. Age range was 18 to 98, with the average age of 39.66 years (SD = 9.03).

Regarding 'sample type' category, it was divided into four variables (general population, subclinical, mental psychological illness, and physical illness). Most of the sample comprised the subjects with mental illness (46.73%), followed by the subclinical category (27.17%), then the general population (21.73%), and finally, a population with physical illness (10.86%). It should be noted that these categories are not exclusive.

In relation to the country of study, most studies (35.86%) were from the United States, South Korea (22.82%), and China (6.52%). Other studies were carried out in Europe (10.11%), mainly Italy (5.43%) and Scotland (3.26%). The remaining 14.13% studies were conducted in other countries.

Analysis of Atypical Values

The data were explored in search of extreme values (± 3 * IQR), outliers (± 1.5 * IQR), and abnormal with the application of Chauvenet's criterion (± 1.96 * SD). The results showed (\( \bar{M} = .8841 \) CI [.8721, .8962], SD = .0583, min. = .74, max. = .98) no extreme, outliers, or abnormal effect sizes were observed. Thus, all primary data are normal. Additionally, data have a normal distribution, \( \bar{Z} (N = 92) = 1.34, ns \), with a slight negative, and non-significant asymmetry, -0.611 (SE = 0.251) and kurtosis, -0.470 (SE = 0.498).

Overall Meta-analysis

The results (see Table 1) show that, in general, the SSI has a mean true internal consistency of .8904, not reaching the desirable reliability criterion for a measure in applied contexts involving decisions (.95; Nunnally, 1978), \( q (N = 24,119) = -.408, Z = -44.81, p < .001 \), nor the minimum criterion (.90; Nunnally, 1978), \( q (N = 90) = .583, \bar{Z} (N = 90) = 1.34, p = .182 \), nor the minimum criterion (.95; Nunnally, 1978), \( q (N = 95) = .712, Z = -32.19, p < .001 \).
24,119) = -.048, \( Z = -5.27, \ p < .001 \). Consequently, it is necessary to correct the direct score for the unreliability of the measure in order not to make errors in the assessment of suicide risk that imply that risk is not appreciated when it is present (fatal diagnostic error in a clinical setting, and punishable error subject to civil liability in judicial setting). However, this result is not generalisable to all studies as there may be studies (measurement settings) with satisfactory internal consistency for assessments in applied domains (upper limit of the credibility interval > .95). Conversely, there are also measurement contexts in which the internal consistency of the measure is very unsatisfactory for the applied assessment (CI lower bound .8189, i.e., the 42.6% of the population standard deviation is error and the 18.11% of the measure of an individual is error). Likewise, the percentage of variance explained by sampling error (% var) accounts for less than 60% of the variance, i.e., the studies are heterogeneous and subject to moderators.

**Moderator Study**

The ISS is equally reliable in females and males, \( q(N = 776) = -.030, \ Z = -.59, \ p = .278 \), so no correction for gender-adjusted unreliability is needed.

When analysing the adaptations, the Chinese adaptation is significantly less reliable than the English original, \( q(N = 2,964) = - .375, \ Z = - 14.43, \ p < .001 \), as is the Persian, \( q(N = 1,781) = -.299, \ Z = - 8.92, \ p < .001 \), and the Korean, \( q(N = 4,713) = - .375, \ Z = - 4.71, \ p < .001 \), but not so for the Italian, \( q(N = 1,107) = - .031, \ Z = - 0.73, \ p = .233 \). Thus, Chinese, Persian, and Korean versions must be corrected.
Table 1. Meta-analyses of the Internal Consistency of the Scale for Suicidal Ideation

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>$r_w$</th>
<th>$S^2_{se}$</th>
<th>SD$_{se}$</th>
<th>$S^2_{res}$</th>
<th>% Var</th>
<th>$S^2_{res}$</th>
<th>SD$_{res}$</th>
<th>95% CI</th>
<th>80% CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>General meta-analysis</td>
<td>92</td>
<td>24,119</td>
<td>.8904</td>
<td>.0034</td>
<td>.0580</td>
<td>.0002</td>
<td>4.93</td>
<td>.0032</td>
<td>.0565</td>
<td>.8787,.939</td>
</tr>
<tr>
<td>Women</td>
<td>7</td>
<td>1,100</td>
<td>.8808</td>
<td>.0008</td>
<td>.0292</td>
<td>.0003</td>
<td>38.31</td>
<td>.0005</td>
<td>.0229</td>
<td>.8675,.8941</td>
</tr>
<tr>
<td>Men</td>
<td>3</td>
<td>600</td>
<td>.8873</td>
<td>.0026</td>
<td>.0513</td>
<td>.0002</td>
<td>8.64</td>
<td>.0024</td>
<td>.0491</td>
<td>.8702,.9044</td>
</tr>
<tr>
<td>English</td>
<td>23</td>
<td>7,862</td>
<td>.9212</td>
<td>.0015</td>
<td>.0386</td>
<td>.0001</td>
<td>4.60</td>
<td>.0014</td>
<td>.0377</td>
<td>.9178,.9246</td>
</tr>
<tr>
<td>Korean</td>
<td>14</td>
<td>3,167</td>
<td>.9052</td>
<td>.0013</td>
<td>.0359</td>
<td>.0001</td>
<td>11.29</td>
<td>.0111</td>
<td>.0339</td>
<td>.8989,.9115</td>
</tr>
<tr>
<td>Chinese</td>
<td>6</td>
<td>1,767</td>
<td>.8402</td>
<td>.0006</td>
<td>.0239</td>
<td>.0003</td>
<td>51.58</td>
<td>.0003</td>
<td>.0166</td>
<td>.8265,.8539</td>
</tr>
<tr>
<td>Italian</td>
<td>5</td>
<td>590</td>
<td>.9163</td>
<td>.0004</td>
<td>.0201</td>
<td>.0002</td>
<td>54.47</td>
<td>.0002</td>
<td>.0136</td>
<td>.9033,.9293</td>
</tr>
<tr>
<td>Persian</td>
<td>6</td>
<td>987</td>
<td>.8612</td>
<td>.0084</td>
<td>.0915</td>
<td>.0004</td>
<td>4.89</td>
<td>.0080</td>
<td>.0893</td>
<td>.8450,.8774</td>
</tr>
<tr>
<td>Subclinical population</td>
<td>34</td>
<td>9,118</td>
<td>.8769</td>
<td>.0022</td>
<td>.0468</td>
<td>.0002</td>
<td>9.15</td>
<td>.0020</td>
<td>.0446</td>
<td>.8721,.8817</td>
</tr>
<tr>
<td>General population</td>
<td>8</td>
<td>4,983</td>
<td>.9230</td>
<td>.0034</td>
<td>.0581</td>
<td>.0000</td>
<td>1.05</td>
<td>.0033</td>
<td>.0578</td>
<td>.9189,.9271</td>
</tr>
<tr>
<td>Mental illness population</td>
<td>23</td>
<td>4,051</td>
<td>.9040</td>
<td>.0017</td>
<td>.0413</td>
<td>.0002</td>
<td>11.31</td>
<td>.0015</td>
<td>.0389</td>
<td>.8983,.9096</td>
</tr>
</tbody>
</table>

Note. $K$ = number of effect sizes; $N$ = total sample size; $r_w$ = sample size weighted mean effect size; $S^2_{se}$ = sample size weighted observed variance of correlations; SD$_{se}$ = sample size weighted observed standard deviation of correlations; $S^2_{res}$ = variance attributed to sampling error variance; % Var = percent of observed variance accounted by sampling error variance; $S^2_{res}$ = variance of correlations after removing sampling error variance; SD$_{res}$ = standard deviation of correlations after removing sampling error variance; 95% CI = 95% confidence interval for $Z = 14.02, p < .001$, and with the mentally ill population, $q(N = 4,469) = .247, Z = 5.02, p < .001$. It is also significantly more internally consistent in the clinical population than in the subclinical population, $q(N = 5,609) = .131, Z = 6.94, p < .001$. For physical illness we found only two effect sizes, so a meta-analysis is not possible because $k$ is insufficient ($< 3$). With these two effect sizes the average effect weighted by sampling error was $.8282 [.7913,.8651]$, an insufficient internal consistency index for evaluation in applied fields.

Discussion

Given the fact that the world has been affected in recent times by problems like war, climate change, and diseases, mankind and individuals are struggling to surmount psychological difficulties, among other things, and suicides remain the focus of many discussions and present a problematic issue, even with many initiatives and projects supported either by the WHO (2021a) or political interventions in specific countries. As can be seen in some of the aforementioned examples, the problem of suicides does not end (WHO, 2021b, 2021c). Fortunately, a lot of organizations works on protocols for prevention, awareness, and sensitization of the population (WHO, 2021a). But we must be conscious of the fact that it will require exertion and a multicultural and multidisciplinary aspect. Because of this, we believe that one of the many contributions from psychology could be the implementation of studies like this one, which can prove the validity of instruments. The use of reliability instruments to detect evidence of suicide (such as the SSI) is essential in psychological autopsy (see, for example, Cano-Montalbán & Quevedo-Blasco, 2018). This will help clarify equivocal deaths within a judicial process (by differentiating between proved suicides and possible homicides or accidents disguised as suicides), without ignoring other techniques and instruments (Deeb et al., 2022; Vrij, Deeb, et al., 2021; Vrij et al., 2022; Vrij, Mann, et al., 2021; Vrij, Palena, et al., 2021), and efficacy studies in forensic practice (Gancedo, Fariña, et al., 2021; Gancedo, Sammarco, et al., 2021). Also, the dearth of studies about suicides in Europe is undoubtedly relevant to the implementation of new research on the SSI (within psychology applied to the legal context).

Due to the clinical and legal relevance of misdiagnosis of suicidal ideation, there is a need for guidelines and criteria that allow the clinician to know the reliability of the instruments and the considerations to be corrected in specific cases. In addition, due to the high prevalence of suicide among prisoners, it is relevant to have specificities of use in this population. The fact is that studies like systematic reviews, meta-analyses, or evaluations of the quality of published studies may be considered irrelevant. Yet, the need for this research on the quality of studies and their journals is being increasingly emphasized (Quevedo-Blasco et al., 2019). Based on this type of review (Muñiz et al., 2020; Viladrich et al., 2021) and specific studies (Prieto-Macías et al., 2020; Quevedo-Blasco et al., 2019; Quevedo-Blasco & Palomares-Rodríguez, 2023; Quevedo-Blasco et al., 2021), it will be possible to more rigorously select the knowledge and instruments that significantly contribute to clinical and forensic practice. Hence, in this meta-analytic study, a reliability generalization analysis of Beck et al.’s (1979) SSI was conducted. Apart from the paper published by King et al. (2014), this is a unique meta-analysis study that uses this scale with the aim of evaluating its psychometric properties, specifically its reliability by alpha ($\alpha$) value after its application in different contexts with specific selection criteria.

The reliability of the instrument allows the reliability of an individual's observed score to be obtained. Thus, with the above average reliabilities that best fits each case, it is possible to obtain
whether a particular individual’s score falls within the suicide risk criterion despite the direct score being lower. To do this we must calculate the standard error of the difference between the test score and the criterion, applying the $SE_{yp} = SD * \sqrt{r(1-r)}$ formula, where $SD$ is the standard deviation of the normative population (10 if normalised to T scores) and $r$ is the computed in these meta-analyses mean true correlation (however, it is recommended to take the lower limit of the credibility interval as it is between around .05 and .10 of the mean true reliability) of the condition of interest (moderator or overall if no moderators are fitted). With the standard error of the difference the upper limit of the 95% confidence interval is computed to obtain the highest score that the individual could have obtained, upper limit = $x + (1.96 * SE_{yp})$, where $x$ is the score obtained by the individual on the ISS. If the upper limit is higher than the criterion, the individual is at the severity of suicidal intent level set by the criterion. In sum, the observed score is an estimate that could have resulted from this computation. This corrects for measurement error by controlling for the risk of a false negative (false negative of suicide risk) which is undesirable in this assessment context. For example, for the 90 patients hospitalized for suicidal ideation of the study of Beck et al. (1979), the standard deviation was 8.44 and the lower limit of the credibility interval 8735 (correct the mean true correlation .04 i.e., the transient error) for the English version, the upper limit for a raw score of 24 (medium intent of suicide) would be 29.88 (high intent of suicide). Thus, a person with a raw score of 24 could be in high risk of intent of suicide.

Unfortunately, there is no literature for a meta-analytic review with populations with high suicide prevalence, such as prison inmates or military personnel, which, moreover, are usually where the measurement instruments are less reliable. In these cases, the reliability for these subsamples is obtained from the reliability of the normative population through the range restriction. Briefly, the following formula is applied, $R_y = 1-(SD/SD)^2(1-R_y)$, where $R_y$ is the reliability of the instrument in the subsample; $SD$ is the standard deviation of the normative population (if there is no study with normative population, studies with general population can be taken to weight the observed reliability by sampling error so that a mean reliability for the general population is obtained, obtaining the combined $SD$). $SD$ is the standard deviation of the subsample and $R_y$ the reliability of the instrument in the normative population (if this is not available or the $SD$ has been obtained from several general population studies, $R_y$ is obtained by weighting the reliability, $r_y$ in each study by the sampling error). As an example, for a $SD$ of 6.32, a $SD$ of 4.22 and a $R_y$ of .92, the resulting reliability would be of .88. With the computed reliability (transient error should be subtracted to the computed reliability, i.e., between .05 and .10; as false negatives of suicidal intention are fatal errors, the most conservative compute of the reliability is the recommended solution), the upper limit of the confidence interval for the observed raw score may be obtained and consequently a protection against false negatives. With these actions not only the patient is protected but also the clinician against civil liability demands.

In terms of limitations, we would like to emphasize several remarks on the search strategy. First, the search method applied in our paper is deemed to be correct, though there is a possibility for further improvement in the form of complementary search criteria to maximize outcome potential. The second limitation is closely related to age. The SSI is applicable to individuals ranging from 17 years, notwithstanding the fact that one of the selection criteria involves individuals who must be at least 18 years. It would be important for future research to focus on the adolescent population. Furthermore, primary studies have limitations that need to be considered. First, internal consistency (Cronbach's alpha) is a measure that overestimates reliability (Cronbach, 1947). In fact, it does not measure transient error (time-random-variation in responses by respondents). Schmidt et al. (2003) estimated transient error at .11 for negative affectivity, a personal trait related to suicidal intention. Thus, Cronbach's alpha may overestimate around 10% true reliability. Second, the reliability of the measure does not include response bias that can also be introduced by the subject through mood (Podsakoff et al., 2003). In this case, dissimulation (symptom concealment) should not only be suspected but should also be a differential diagnosis in assessment contexts where a misdiagnosis may result in fatalities or injuries for the respondents (Arce et al., 2023). As for this, statistical data must always be treated with caution because of certain statistical errors or symptom concealment, hence the importance of the psychological autopsy is increasing (Cano-Montalbán & Quevedo-Blasco, 2018). The scientific community needs to further develop studies with samples (like inmates with physical problems) to be able to adjust more precisely mean error and thus alpha estimation. This will facilitate diagnosis and minimise fatal consequences.

Conflict of interest
The authors declare no conflicts of interest.

Acknowledgement
The authors acknowledge and thank Professor Ramón Arce for his methodological support of this manuscript.

References
References marked with an asterisk indicate studies included in the meta-analysis.


