Comparison between the abuse assessment screen and the revised conflict tactics scales for measuring physical violence during pregnancy

M E Reichenheim, C L Moraes

Study objective: Because of the promise of its ability to quickly identify cases of violence against women during pregnancy, the abuse assessment screen (AAS) should be the focus of numerous psychometric evaluations. This paper assesses its measurement accuracy compared with the revised conflict tactics scales (CTS2) used as standard.

Design: Cross sectional study. Besides several ancillary questions, the AAS consists of three anchor questions about violence against pregnant women. These are inclusive, respectively covering lifetime, preceding 12 months, and pregnancy periods. These questions are the main focus of this article. The CTS2 physical aggression scale consists of 12 items divided into minor and severe subscales. A positive event is defined as having at least one positive item in the respective subscale. The 12 item score is also used.

Setting and participants: The instruments were applied to 748 women, 24 to 72 hours after delivery in three major public sector maternity wards of Rio de Janeiro from March to September 2000.

Main results: According to the CTS2, prevalences of minor and severe physical violence perpetrated against a pregnant woman are 18.4% (95% CI 15.7 to 21.4) and 7.6% (95% CI 5.8 to 9.8), respectively. Taking these subscales as standards, sensitivities are 31.9% (95% CI 24.9 to 40.3) and 61.4% (95% CI 47.6 to 74.0), respectively. Specificities are above 97%.

Conclusion: These findings are somewhat worrying because the number of victims who are not identified and offered assistance is considerable. On a practical note, it would be sensible not to use the AAS as a stand alone screening tool until more evidence is gathered.

Since its proposal, the AAS has been used in clinical or community settings and several epidemiological studies. However, to the best of the authors’ knowledge, the instrument’s proponents have so far only carried out formal psychometric evaluations. These are restricted to one reliability test-retest assessment where an 83% agreement score was found between two sittings and construct validity appraisals based, among others, on comparisons with other widely used evaluation tools—the index spouse abuse, the danger assessment screen, and the conflict tactics scales (CTS). These assessments showed that women who were identified as abused on the three question AAS also scored significantly higher on the others. Proper as this may be, the ability to correctly identify positive and negative cases has never been investigated.

Because of the promise of its ability to quickly and easily identify cases of violence against women during pregnancy, the AAS should be the focus of numerous psychometric evaluations. This paper assesses the measurement accuracy of the AAS question specifically addressing the pregnancy period with the revised conflict tactics scales (CTS2) used as standard.

METHODS

Design, study population, and field procedures

This study is subsidiary to a hospital based case-control study exploring the relation between violence within families of pregnant women and premature childbirth. Data collection took place from March to September 2000 in three major public sector maternity wards of Rio de Janeiro. Cases
comprised all premature newborn infants accrued during the six month period. To comply with the pre-specified sample size, about 14% of the 3800 eligible live births with gestational age above 36 weeks were randomly selected as controls. Subjects were drawn from a list of live births occurring in the previous 24 hours. As two women refused finishing interviews, one had missing data, and 23 situations did not refer to steady relationships involving current or former partners, the effective sample amounts to 748 couples (233 cases and 515 controls). Women with diabetes mellitus, systemic arterial hypertension, or who gave birth to neonates with severe congenital malformations, infections associated with prematurity, or twins were excluded.

Five extensively trained and closely supervised interviewers visited the hospitals on a daily basis. Interviews were conducted during the first 48 hours postpartum, before the woman’s discharge from hospital, in a reserved area and without the presence of the husband or partner. A multidimensional structured questionnaire was used, which included modules to assess violence. The AAS module was applied before the CTS2 on the same sitting during the data collection of the main study. The study was formally approved by the research ethics committee of the Rio de Janeiro Municipal Health Department in conformance to the principles embodied in the Declaration of Helsinki. Participation in the study followed a free and informed consent. Confidentiality of information was guaranteed. All the women received information on public facilities in Rio de Janeiro for managing families suffering from violence and were encouraged to seek help if they felt it was necessary.

The CTS2 physical violence scale used as standard
Given the encouraging evaluations and successful use in at least 20 countries of the first CTS, the CTS2 was subsequently developed for exclusively detecting partner violence. The complete CTS2 consists of 78 items encompassing five scales, namely, negotiation, psychological aggression, physical violence, sexual coercion, and injury. Items relate to the respondent and the respective partner. The reliability and validity of the instrument has been studied since its release in 1996 by Straus et al. Initially, the authors showed that all five subscales were internally consistent and were duly related as laid down by the underlying theory. Later, Newton et al. used a confirmatory factor analytic procedure to corroborate the original five factor model. Similarly, on testing the instrument on incarcerated female subjects, Luente et al. confirmed the five dimensions, especially with regards to the respondent as perpetrator. In another study on incarcerated population, Jones et al. also identified a similar dimensional pattern, although their factor analysis failed to clearly separate the psychological from the physical scale. The authors, none the less, found that all subscales of the CTS2 were positively and significantly related with another 19 item checklist to assess abusive behaviour. As to internal consistency, all three recent studies also showed reasonably high \( \alpha \) coefficients, especially in relation to the physical aggression scales.

The CTS2 Portuguese version used here as standard is the result of a formal adaptation process. After an earlier evaluation of concept, item, and semantic equivalences, a second study presented a wide range of psychometric properties of the proposed version. Besides showing acceptable reliabilities for each subscale, the factor analysis once again identified a pattern with recognisable correspondence to the underlying dimensions. The study also evaluated construct validity focusing, as in Straus et al., on the pattern of association regarding the domains covered by the five scales. Relations between those and other theory related dimensions were scrutinised as well. It could be shown that, in tandem with literature, victims were preferentially adolescents, poorly educated, of low socioeconomic status, in relationships involving alcohol and illicit drug use more often, and less in contact with health services. This paper involves exclusively the physical violence scale and subscales as perpetrated by the woman’s partner or ex-partner during pregnancy. The minor and severe subscales consist of five and seven items, respectively. The content of each item is outlined in table 1 presented in the results section. Complete wordings can be found in Straus et al.

Variables and data analysis
The criterion validity analysis of the AAS compared with CTS2 relates to women who responded to both violence modules. The recall frame strictly relates to pregnancy. For that reason, only one of the three AAS anchor questions is of interest here and reads as follows: “Since you have been pregnant, have you been hit, slapped, kicked, or otherwise physically hurt by someone?” A “true” positive event is defined as having at least one positive item in each CTS2 subscale. The score presented in figure 2 is based on the summation of all 12 dichotomised items.

Quality control of data entry, processing and analysis are conducted in Stata release 8.0 (College Station, TX). Sensitivity (Se) and specificity (Sp) estimates are carried out through the software’s dtag routine, which provides exact binomial confidence intervals. Similarly, false positive and false negative estimates are obtained via ci (manual: [R] ci) using the binomial option. The same routine is used for estimating descriptive proportions, whereas continuous variables are analysed by summarise (manual: [R] summarise). Point-biserial correlations are obtained via pbis. For the purpose of presentation, the curves in figure 1 are slightly smoothed through the lowess program using a bandwidth of 0.6 (manual: [R] lowess).

RESULTS
The study population comprises mostly young women/mothers with mean age of 23.9 (SD: 6.5); poorly educated with 57.4% (95% CI: 53.8 to 61.0) having attended less than eight years of schooling; with an average of 6 (SD: 2.5) prenatal visits; and coming from low income families with a median monthly income per capita of US$ 96.7 (c5%: 26.5, c95%: 346.4). Seventy five per cent (95% CI: 71.8 to 78.1) were either married or living with a partner at the time. According to the CTS2, prevalences of minor and severe physical violence perpetrated by a partner against a pregnant woman are 18.4% (95% CI: 15.7 to 21.4) and 7.6% (95% CI: 5.8 to 9.8), respectively. Noting that most severe acts come about when minor events are also committed, the figure for the total scale is 18.9% (95% CI: 16.2 to 22.0). As for the AAS question, 6.7% (95% CI: 5.0 to 8.7) of women stated having been batted during pregnancy. The point-biserial correlation with the full 12 item CTS2 physical violence scale is 0.68.

Figure 1 presents the sensitivity and specificity of the AAS anchor question on physical abuse during pregnancy as contrasted with the CTS2 physical violence scale. Sp estimates are all above 97%, irrespective of the severity of the events according to the CTS2. In contrast, Se estimates are much lower, especially with regards to minor acts. No important difference could be found on testing accuracy according to women’s age (adolescence), educational status, and marital status (whether women lives with partner).

Table 1 provides an insight to the pattern of physical abuse missed out by the AAS question covering the same period as the CTS2 (pregnancy). For minor events, except for the item referring to having been “slapped”, above 40% of women who had been victim according to the other items of the CTS2 were not detected on the AAS. Also note that although most
Comparison between the AAS and the CTS2

The amount of potential victims of violence failing to be third of minor and one third of severe episodes are missed. how people are classified shows another picture. Almost two to reinforce such a conclusion. However, a closer look into fairly high point-biserial correlation between the AAS CTS2 (7.6%). If taken as a construct validity criterion, 47 the nancy (6.7%) is similar to the severe cases recognised by the fraction of cases detected by the instrument during preg-

DISCUSSION

The ability to accurately identify subjects, whether for screening purposes at the community level or in research contexts, should be an essential quality of any instrument. In the field of intimate violence, confidently dismissing suspected cases or, conversely, detecting and following up “true” events is all too important to achieve successful results in terms of offering women help and resources, as well as to improve the validity of research results.

Focusing solely on prevalence of severe events, on the surface, the AAS seems to be performing rather well. The fraction of cases detected by the instrument during pregnancy (6.7%) is similar to the severe cases recognised by the CTS2 (7.6%). If taken as a construct validity criterion, the fairly high point-biserial correlation between the AAS question encompassing pregnancy and the CTS2 scale tends to reinforce such a conclusion. However, a closer look into how people are classified shows another picture. Almost two thirds of minor and one third of severe episodes are missed. The amount of potential victims of violence failing to be further explored and managed is somewhat worrying, particularly because the AAS was conceived as a screening tool in prenatal contacts.

As conveyed by table 1, the pattern of CTS2 items missed out by the AAS shows how much wording of questions is important for detection. As the number of explicit examples of acts in the AAS is fewer than CTS2, it is not surprising that the percentage of people to show positive responses in the first may not be as high as that in the second. But there is a difference between minor and major events. The perception of a severe violence act as an emergency situation and the search for help may redouble the respondent’s concentration, leading any of the two instruments to sense cases more accurately. In contrast, minor violent acts may be overlooked if not overtly asked.13 In four of the five CTS2 minor types, nearly half of subjects were missed by the AAS. This calls for a change in wording of AAS to cover minor violent episode appropriately.

In effect, the authors of the AAS have recently proposed a modified edition, wherein the same key questions on perpetrated events were slightly expanded.47 Essentially, two violent acts—pushed and shoved—were added to the wordings of the questions covering the same recall periods. It is plausible that this enhancement may have lead to an amelioration of the performance of the instrument because, as conveyed by table 1, these two acts shared the responsibility of a high percentage of non-coverage with “grabbed”. Perhaps, adding this to the question would also make a difference. Clearly, a formal testing of the new format should be encouraged, although it remains to be seen if some structural problems can be overcome by a carefully improved wording. It should be recalled that, as in the original format, there is still reliance on stand alone anchor questions to

of the overlooked events are minor, for some major items (for example, having been ‘punched or hit with something that could hurt’ or ‘kicked’), as much as one quarter failed to be uncovered.

The increasing dependency of Se on severity of the event is further explored in figure 2. Using the CTS2 scale comprising all 12 events as standard, a 20% false negative level is reached around a score of 4. On a more stringent note, a score of 8 has uncovered.

**Table 1**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>CTS2 item</th>
<th>N*</th>
<th>n†</th>
<th>%‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minor</strong></td>
<td>Had something thrown that could hurt</td>
<td>29</td>
<td>15</td>
<td>51.7 (32.5 to 70.6)</td>
</tr>
<tr>
<td></td>
<td>Had arm or hair twisted</td>
<td>37</td>
<td>15</td>
<td>40.5 (24.8 to 57.9)</td>
</tr>
<tr>
<td></td>
<td>Was pushed or shoved</td>
<td>78</td>
<td>39</td>
<td>50.0 (38.5 to 61.5)</td>
</tr>
<tr>
<td></td>
<td>Was grabbed</td>
<td>93</td>
<td>56</td>
<td>60.2 (49.5 to 70.2)</td>
</tr>
<tr>
<td></td>
<td>Was slapped</td>
<td>51</td>
<td>12</td>
<td>23.5 (12.8 to 37.5)</td>
</tr>
<tr>
<td><strong>Severe</strong></td>
<td>Had a knife or gun used against her</td>
<td>10</td>
<td>2</td>
<td>20.0 (2.5 to 55.6)</td>
</tr>
<tr>
<td></td>
<td>Was punched or hit with something that could hurt</td>
<td>35</td>
<td>9</td>
<td>25.7 (12.5 to 43.3)</td>
</tr>
<tr>
<td></td>
<td>Was choked</td>
<td>7</td>
<td>1</td>
<td>14.3 (0.4 to 57.9)</td>
</tr>
<tr>
<td></td>
<td>Was slammed against the wall</td>
<td>31</td>
<td>6</td>
<td>19.4 (7.4 to 37.5)</td>
</tr>
<tr>
<td></td>
<td>Was beaten up</td>
<td>25</td>
<td>2</td>
<td>8.0 (1.0 to 26.0)</td>
</tr>
<tr>
<td></td>
<td>Was burned or scalded on purpose</td>
<td>2</td>
<td>0</td>
<td>0 (-)</td>
</tr>
<tr>
<td></td>
<td>Was kicked</td>
<td>35</td>
<td>9</td>
<td>25.7 (12.5 to 43.3)</td>
</tr>
</tbody>
</table>

*Number of item positives according to CTS2; †number of subjects missed out by the AAS; ‡percentages of subjects missed out by the AAS with 95% exact binomial confidence intervals in parentheses.

Figure 1 Sensitivity (Se) and specificity (Sp) of the abuse assessment screen (AAS) as compared with the revised conflict tactics scale (CTS2) on physical violence. *Positive event defined as at least one positive item in the (sub)scale; †Se = a/(a + c); ‡Sp = d/(b + d); 5 item scale; *7 item scale; *12 item scale.
classify people; questions contain wordings covering several issues at once; and questions are posed without much of a preamble. The first feature tends to restrict reliability and content validity,\textsuperscript{26} the second may confuse the respondent,\textsuperscript{26} whereas the third may lead to false denials if a respondent is already unwilling or afraid of talking about contentious issues.

Three methodological issues are worth considering when appreciating our findings. It may be contended that the results arise in a different context and culture than that of the instrument’s development, which, in principle, could hamper or even preclude generalisations. It must be emphasised, however, that both versions used in this study underwent extensive and careful trans-cultural adaptations.\textsuperscript{34, 35, 49} Strict guidelines were followed, including formal evaluations of concept, item, semantic, and measurement equivalences between the Portuguese versions and the original instruments in English.\textsuperscript{30, 31}

A second feature that needs consideration is that the CTS2, being a structured closed question schedule, may not be held as the optimal tool for the detection of violence cases. Ideally, as in any criterion validity study, the standard should be based on a reference diagnostic procedure\textsuperscript{31} and in the absence of such a standard, to engage in an evaluation based on correlations or other agreement estimates.\textsuperscript{33} Yet, a “construct validity” type of approach such as this would simply tend to corroborate previous findings showing similar directions of detection between the AAS and other allegedly “non-gold” instruments.\textsuperscript{16, 17, 24} The main gist of the debate would then be missed, which is whether and how much the AAS succeeds/fails in identifying violence during pregnancy. There is undoubtedly a demarcation problem in terms of what is to be considered a “true” case, given that the CTS2 has never been officially appointed in the literature as the reference tool. However, it should be emphasised that, together with its precursor—the CTS1\textsuperscript{16}—the instrument is by far the most widely used tool of its class and, as shown in the methods section, has an extensive and solid track of psychometric evaluations behind it. For this reason, the CTS2 may be provisionally proposed as standard, although the assessment of the results needs to acknowledge that an “alloyed” standard is being used.

A third point to bear is that both instruments were applied in one sitting during the data collection of the main case-control study. As the application of the CTS2 followed the AAS questions, you cannot rule out interviewers tending towards positive events as a result of preceding positive AAS items or, conversely, disregarding any CTS2 item after negative answers on the anchor question. Although the authors believe that these are not probable occurrences given that the CTS2 is a structured schedule just covering factual events and used after a rigorous training programme, it should be pointed out that, if bias indeed happened, these results would tend to be conservative. At any rate, it would be an improvement in future investigations to use the CTS2 first in part of the sample to evaluated the effect of the application order, a procedure that was not implemented in this research as the data collection strategy was not specifically conceived with any particular psychometric study in mind.

An interesting development would be to invest in a screening instrument someway between the AAS and the CTS2. The CTS2 is quite large, even when the scope of committed actions is restricted to the pregnant woman as victims (39 items). Although beneficial as a research device or as part of an in depth family evaluation schedule, it is not as desirable as a quick screening tool. Thus, an abridged edition of the CTS2, possibly coupled with some descriptive items concerning more severe events found in the AAS, would be a compromise to look for. Further studies comparing possible offshoots to the original instruments or even others sources of information with regards to timing and performance would be of great interest. It would also be worth discussing whether the AAS can be used as an interview guide, rather than a structured form. Perhaps, the anchor questions may serve well as starting points for more exhaustive qualitative appraisals.

While those suggestions deserve attention and extension, new rounds of psychometric evaluations of the AAS should be considered before investing in new alternatives. Similar assessments as the one presented in this paper, carried out on the original AAS in English and in other languages, could be fruitful, either to strengthen the instrument’s position or, conversely, to confidently clear the way for new developments. On a practical note, it would be sensible not to use the AAS as a stand alone screening tool until more evidence is gathered.

**Authors’ affiliations**

M.E. Reichenheim, C L Moraes, Departamento de Epidemiologiologia, Instituto de Medicina Social, Núcleo de Pesquisa das Violências (NUPEVI), Universidade do Estado do Rio de Janeiro, Brazil

Funding: this project was funded in part by a grant from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Brazilian National Research Council), grant 663073/9987 (PRONEX Project) and in part by FAPERJ (Rio de Janeiro State Research Foundation),
grants E-26/171.223/98 and E-26/150.893/99. MER is partially funded by CNPq, grant 300234/94-5.

Conflicts of interest: none declared.

REFERENCES