

Inter-observer variation in the assessment of clinical signs in sick Tanzanian children

E. Kahigwa¹, D. Schellenberg^{1,2}, J. Armstrong Schellenberg^{1,3}, J. J. Aponte², P. L. Alonso² and C. Menendez² ¹Ifakara Health Research and Development Centre, P.O. Box 53, Ifakara, Tanzania; ²Unidad de Epidemiología y Bioestadística, Hospital Clinic Villarroel 170, 08036 Barcelona, Spain; ³Swiss Tropical Institute, Socinstrasse 57, CH 4051, Basel, Switzerland

Abstract

We assessed the inter-observer agreement in identification of a range of 24 clinical signs associated with disease presentation in 327 children aged <5 years admitted to hospital in January–June 1999 in Ifakara, southern Tanzania. Children with diagnoses of malaria, pneumonia, diarrhoea, anaemia and malnutrition were examined independently by 2 clinical officers. Findings were recorded on a standard proforma. The Kappa-statistic was used to assess inter-observer agreement for each sign. Physical signs were more likely to be agreed upon by clinicians if they involved inspection than if they involved auscultation. The signs included in the Integrated Management of Childhood Illness (IMCI) algorithm were found to be largely appropriate (Kappa-scores >0.41) although there was only fair agreement (Kappa-score 0.21–0.40) in the detection of neck stiffness and chest indrawing and slight agreement in the detection of dehydration (Kappa-score 0.199). All objective neurological signs were less reliably assessed in infants than in older children. The difficulties surrounding the diagnosis of impaired consciousness in young children should increase vigilance in the diagnosis and management of neurological complications of illnesses in infancy.

Keywords: clinical signs, children, assessment, inter-observer variation, Tanzania

Introduction

The usefulness of a clinical sign is dependent on the ease with which it can be assessed, and in some cases also on its prognostic value. WHO promotes the use of case management guidelines to facilitate the care of sick children. Many such guidelines, for example the Integrated Management of Childhood Illness (IMCI), are intended for use by health workers in peripheral health facilities in developing countries (GOVE, 1997). In Tanzania this role is most frequently played by clinical officers (COs) or lower grades who receive only 2- or 3-years training. The IMCI guidelines rely on the recognition of clinical signs to distinguish mild and moderate illnesses, which may be appropriately managed at peripheral health facilities, from severe illnesses, which need referral for more intensive treatment.

Assessments of the ability of staff with different levels of training to recognize and interpret certain clinical signs have demonstrated that training increases the ability to recognize these signs (GADOMSKI *et al.*, 1993; HANSEN *et al.*, 1994; PERKINS *et al.*, 1997; ANONYMOUS, 1998). Such assessments should be considered when drawing up guidelines and when planning the training of health personnel.

Several studies have documented inter-observer variation in the assessment of clinical signs (HANSEN *et al.*, 1994; SINGHI *et al.*, 1994; ENGLISH *et al.*, 1995; NEWTON *et al.*, 1997). Most of these studies apply the Kappa-statistic (κ) which suggests agreement beyond that expected by chance alone if it is greater than zero (LANDIS & KOCH, 1977). A simple scale has been proposed to facilitate the interpretation of Kappa-values (Table 1). An assessment of respiratory signs (nasal flaring, chest indrawing, deep breathing) using video materials produced overall scores between 0.50 and 0.70. However, there was marked inter-individual variation with scores as low as 0.1 recorded (ENGLISH *et al.*, 1995). There was more variation in the assessment of the Blantyre Coma Score (BCS) for children with severe falciparum malaria ($\kappa = 0.27$), although overall agreement was better than for its individual components, e.g., verbal response ($\kappa = 0.02$) (NEWTON *et al.*, 1997).

Table 1. Classification of Kappa-scores (LANDIS & KOCH, 1977)

κ -Score	Classification
<0	Less agreement than expected by chance alone
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

We report the inter-observer variation in the assessment of clinical signs amongst COs examining young children admitted to hospital in Tanzania.

Materials and Methods

Study population

The study was carried out in the paediatric department of the St Francis Designated District Hospital (SFDDH) Ifakara, in south-eastern Tanzania. The department has 70 beds and about 6000 admissions per year. The top 5 primary clinical diagnoses are malaria, anaemia, pneumonia, diarrhoea and malnutrition. SFDDH collaborates with the Ifakara Health Research and Development Centre (IHRDC) in clinical trials, and since 1994 a round-the-clock surveillance system of all admissions to the paediatric ward (0–12 years of age) has been operating (SCHELLENBERG *et al.*, 1999). COs complete a standardized proforma for each child admitted, documenting the history and clinical signs, (Table 2), from which the BCS was derived (MOLYNEUX *et al.*, 1989). The COs who maintain the clinical surveillance are trained to assess clinical signs and complete the surveillance forms in a standardized way, as described in a training manual. Daily supervision by one of us (E.K.) was supplemented by a month of increased-frequency bedside teaching before this study commenced.

Study design

Children were selected for this study if they were aged between 4 months and 6 years, and admitted to the paediatric ward with malaria, pneumonia, diar-

Address for correspondence: Dr Clara Menendez, Unidad de Epidemiología y Bioestadística, Hospital Clinic, Villarroel 170, 08036 Barcelona, Spain; phone +34 93 2275706, fax +34 93 4515272, e-mail cmenende@medicina.ub.es

Table 2. Clinical signs assessed in children aged <5 years admitted to hospital in Ifakara, southern Tanzania

System and sign	Procedure
General examination	
Pulse rate	• Auscultating heart beats
Tachypnoea ^a	• >50 breaths counted per min in infant (>40 in child 1–4 years) by observation of chest wall
Moderate/severe dehydration ^{a,b}	• Assess general condition (restless/lethargic) and thirst (thirsty/unable to drink) and look for reduced skin turgor. Look for sunken eyes, lack of tear production when crying and dry mucous membranes
Pallor ^a	• Observe the colour of conjunctiva, palms and nail beds
Oral candidiasis	• Observation of oral cavity and tongue
Oedema ^a	• Observe for pitting on pressing the dorsal tibia
'Flaky-paint' skin	• Inspection of skin on lower extremities for hypopigmentation and peeling
Hair changes	• Look for sparse distribution of orange-coloured hair
Abdominal examination	
Splenomegaly	• Spleen palpable below costal margin
Hepatomegaly	• Liver palpable 3 cm below costal margin
Cardio-respiratory systems	
Chest indrawing ^a , deep breathing and nasal flaring	• Look for intercostal/subcostal indrawing, chest and nasal movements on inspiration
Crackles, wheezes and gallop rhythm	• Auscultation of the lungs and heart
Central nervous systems	
Conscious level ^a	• Normal or impaired on observation
Position	• Inspect for decorticate/decerebrate/normal posture
Bulging fontanelle	• Palpate to assess if bulging (i.e. palpable bump). Not applicable if the fontanelle is already closed on palpation
Neck stiffness ^a	• Resistance to passive or active flexion of the neck
Inability to sit unsupported	• Observe ability to sit unsupported
Normal eye movement	• Observe if the child is able to follow moving objects
Normal verbal response	• Listen for the sound made in response to calling or sternal rub
Localizing painful stimulus	• Look for movement of hand towards source of pain in response to sternal rub

^aIncluded in the Integrated Management of Childhood Illness algorithm.

^bLevel of dehydration classified according to WHO.

rhoea, anaemia or malnutrition between January and June 1999. Verbal consent was given by the mother before study activities commenced. Clinical examinations were performed independently by 2 COs within 1 h of each other. A form with 24 clinical signs (Table 2) was completed and a record made of whether or not the child was crying at the time of the examination. Only one CO was present during each examination and the findings were not discussed between the COs. Four patients were assessed per session and the COs had no access to the patient histories and none had admitted the child. Seven COs were involved in the clinical assessments (Table 3).

Statistical analysis

Data were double entered at IHRDC using FoxPro for Windows (v 2.6 Microsoft, Seattle, USA), then summarized using Stata (Stata Corp, Texas, USA) and

Table 3. Number of paediatric patients seen by each clinical officer in the hospital study in Ifakara, southern Tanzania

Clinical officer	Number of children examined
A	102
B	80
C	103
D	92
E	60
F	108
G	109

Microsoft Excel (Microsoft). Agreement amongst the COs was assessed using Cohen's Kappa-score as estimated by STATA.

Results

Observations were made on 327 children with a mean age of 20.5 months (interquartile range 8.3–31.2); 134 (41.0%) children were aged <1 year.

The prevalence of each clinical sign and associated Kappa-score is presented in Table 4. There was at least fair agreement in the assessment of most signs. The Figure shows that crying during one of the examinations tended to reduce the reliability with which signs were elicited. The Kappa-scores for some signs appeared to vary with age. Table 5 shows signs for which the Kappa-score in children aged >1 year fell into a different agreement category than the assessment for children aged <1 year. Signs detected on inspection tended to have higher Kappa-scores than those elicited by palpation or auscultation.

Neurological signs were amongst the most reliably detected although the Kappa-scores were uniformly worse for objectively assessed signs in younger children. This age effect was also reflected in the summary BCS. No such effect was noted for the relatively subjective assessment of impaired consciousness [<1 year, $\kappa = 0.57$ (0.40, 0.73); ≥ 1 year, $\kappa = 0.57$ (0.43, 0.71)], although the level of agreement for this sign was not as great as the more objective signs in older children. Agreement was at least moderate for 4 of the 7 IMCI signs evaluated: oedema, general level of consciousness, tachypnoea and pallor.

Table 4. Prevalence and level of inter-observer agreement (Kappa-scores) of clinical signs in children admitted to hospital in Ifakara, southern Tanzania

Sign	Prevalence <i>n/N</i> (%)	Kappa-score (95% CI)	Agreement
Fontanelle (bulging)	8/158 (5.1)	0.930 (0.774, 1.000)	Almost perfect
Flaky paint skin	34/327 (10.4)	0.813 (0.705, 0.921)	
Unrousable coma (BCS<4)	66/327 (20.2)	0.674 (0.565, 0.782)	Substantial
Oedema	53/325 (16.3)	0.673 (0.565, 0.781)	
Ability to sit	140/326 (42.9)	0.644 (0.535, 0.752)	
Verbal response	60/325 (18.5)	0.630 (0.522, 0.739)	
BCS, any impairment	115/327 (35.2)	0.630 (0.522, 0.739)	
Orange hair	64/326 (19.6)	0.625 (0.516, 0.733)	
General LoC	75/327 (22.9)	0.572 (0.464, 0.679)	Moderate
Tachypnoea	167/296 (56.4)	0.565 (0.451, 0.679)	
Splenomegaly	94/327 (28.7)	0.563 (0.455, 0.671)	
Motor response	88/327 (26.9)	0.562 (0.453, 0.670)	
Pallor	77/327 (23.5)	0.507 (0.398, 0.615)	
Oral <i>Candida</i>	42/325 (12.9)	0.499 (0.390, 0.608)	
Crackles	112/327 (34.3)	0.482 (0.373, 0.590)	
Eye movements	52/321 (16.2)	0.432 (0.322, 0.541)	
Position	12/327 (3.7)	0.421 (0.313, 0.528)	
Nasal flaring	53/327 (16.2)	0.369 (0.262, 0.476)	Fair
Wheeze	34/327 (10.4)	0.345 (0.237, 0.453)	
Chest indrawing	89/327 (27.2)	0.338 (0.229, 0.446)	
Pulse >120 beats/min	269/325 (82.8)	0.335 (0.226, 0.443)	
Fontanelle (sunken)	31/158 (19.6)	0.327 (0.172, 0.482)	
Hepatomegaly	60/327 (18.3)	0.305 (0.198, 0.412)	
Deep breathing	26/327 (8.0)	0.272 (0.164, 0.380)	
Neck stiffness	15/326 (4.6)	0.266 (0.169, 0.364)	
Dehydration (moderate/severe)	12/327 (3.7)	0.199 (0.091, 0.307)	Slight
Gallop rhythm	10/157 (6.4)	-0.034 (-0.055, -0.013)	Poor

BCS, Blantyre Coma Score; LoC, level of consciousness; 95% CI, 95% confidence interval.

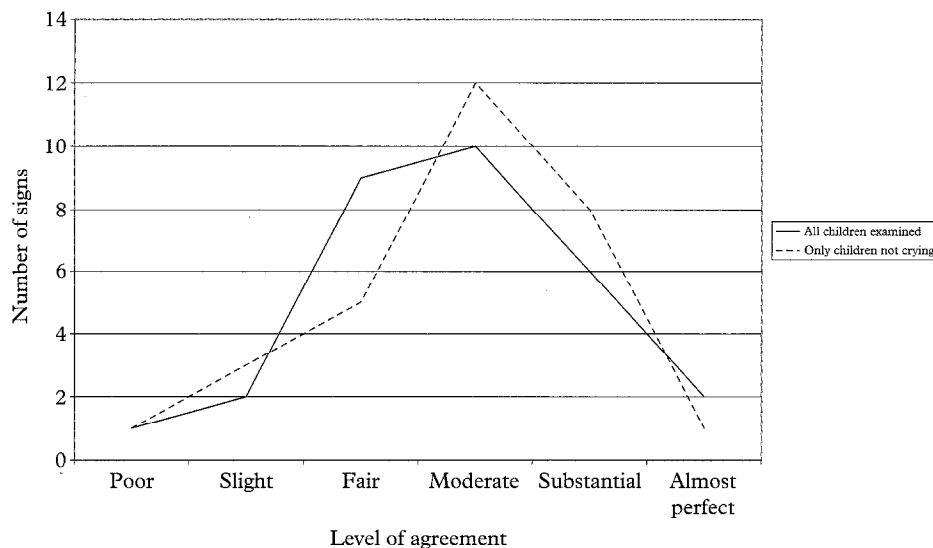


Figure. The effect of crying on inter-observer variation in assessments of clinical signs in sick Tanzanian children.

Discussion

Caution is needed in the interpretation of differences in Kappa-scores in this study as confidence intervals tended to overlap, rendering many apparent differences not statistically significant. Nevertheless several interesting patterns were discernible. For instance, the predictable observation that crying tends to reduce the reliability with which signs were elicited should be borne in mind both in routine clinical practice and in the design of studies such as this.

There was considerable variation in the level of

agreement of different signs, ranging from worse than expected by chance alone (gallop rhythm) to almost perfect agreement (flaky paint skin and bulging fontanelle). Signs detected by inspection tended to have higher levels of agreement than signs elicited by palpation or auscultation, increasing the potential value of visual aids such as pictures or video when training clinicians to recognize the former. In contrast, signs involving auscultation, such as gallop rhythm, wheezes and crackles, were less reliably identified, in keeping with earlier findings (MARGOLIS & GADOMSKI, 1998).

Table 5. Effect of child's age on inter-observer variation in the assessment of selected clinical signs (Ifakara, Tanzania)

Sign	Children aged <1 year		Children aged ≥1 year	
	Prevalence n/N (%)	κ (95% CI)	Prevalence n/N (%)	κ (95% CI)
Assessed less reliably in younger children				
Oedema	7/133 (5.3)	0.174 (0.019, 0.329)	46/192 (24.0)	0.718 (0.577, 0.859)
Orange hair	13/134 (9.7)	0.283 (0.114, 0.451)	51/192 (26.6)	0.714 (0.572, 0.855)
Unrousable coma (BCS<4)	31/134 (23.1)	0.501 (0.332, 0.669)	35/193 (18.1)	0.810 (0.669, 0.951)
Eye movements	20/132 (15.2)	0.248 (0.077, 0.418)	32/189 (16.9)	0.551 (0.408, 0.693)
Motor response	62/134 (46.3)	0.384 (0.215, 0.553)	26/193 (13.5)	0.666 (0.525, 0.807)
Verbal response	21/133 (15.8)	0.448 (0.280, 0.617)	39/192 (20.3)	0.724 (0.582, 0.865)
Flaky paint skin	3/134 (2.2)	0.560 (0.393, 0.728)	31/193 (16.1)	0.834 (0.693, 0.974)
Wheeze	15/134 (11.2)	0.206 (0.038, 0.374)	19/193 (9.8)	0.458 (0.317, 0.599)
BCS, any impairment	68/134 (50.7)	0.478 (0.308, 0.647)	47/193 (24.4)	0.719 (0.578, 0.860)
Pulse >120 beats/min	120/133 (90.2)	0.176 (0.007, 0.345)	149/192 (77.6)	0.353 (0.211, 0.494)
Ability to sit	82/133 (61.7)	0.537 (0.367, 0.707)	58/193 (30.1)	0.655 (0.514, 0.796)
More reliable in younger children				
Position	4/134 (3.0)	0.479 (0.320, 0.638)	8/193 (4.1)	0.376 (0.235, 0.517)
Dehydration (moderate/severe)	8/134 (6.0)	0.247 (0.080, 0.415)	4/193 (2.1)	0.143 (0.010, 0.275)
Oral <i>Candida</i>	19/134 (14.2)	0.610 (0.441, 0.780)	23/191 (12.0)	0.407 (0.265, 0.549)
Deep breathing	18/134 (13.4)	0.363 (0.195, 0.530)	8/193 (4.1)	0.077 (-0.064, 0.218)
Chest indrawing	45/134 (33.6)	0.479 (0.309, 0.648)	44/193 (22.8)	0.187 (0.047, 0.328)
Neck stiffness	10/134 (7.5)	0.338 (0.175, 0.501)	5/192 (2.6)	0.000 (0.0)

BCS, Blantyre Coma Score; 95% CI, 95% confidence interval.

That some signs were so unreliably assessed is of concern as the COs involved in these assessments were better trained and supervised than many working in hospitals and health facilities in the periphery. Peripheral facilities are usually staffed by COs and guidelines intended for use in such settings should concentrate on signs that can be reliably recognized. It was encouraging therefore that agreement tended to be better for signs included in the IMCI algorithm.

Many of the children included in this study were aged <1 year and the study suggested that some signs are more or less reliably assessed in infants than in older children. Signs such as chest indrawing and deep breathing, for which the Kappa-scores were higher in infants, may become more subtle as the chest cage grows, so that they are more easily missed in older children. Of particular note, however, was the increased difficulty in detecting objective neurological signs in infants (eye movements, motor response to a painful stimulus, verbal response, BCS and ability to sit). This corroborates earlier observations from which it was concluded that the ability to locate a painful stimulus is unreliable before the age of 9 months (NEWTON *et al.*, 1997). This finding has particular implications in areas of intense malaria transmission where the burden of disease falls on infants (SCHELLENBERG *et al.*, 1999). The difficulties of documenting impaired consciousness in such children may have contributed to the myth that neurological impairment is uncommon in malaria in infancy (SNOW *et al.*, 1994). However, in the current study conscious level was judged to be either normal or abnormal and this subjective assessment was equally reliable in children of different ages. Furthermore impairment of consciousness assessed on this basis in malaria cases aged <9 months has been shown to be common and an independent risk factor for death (SCHELLENBERG *et al.*, 1999). Increased recognition of impaired consciousness in young malaria cases may result in improved case management of this vulnerable group.

In conclusion, this study has documented the level of inter-observer agreement in the assessment of a number of common clinical indicators. Physical signs may not be reliably elicited from a child examined while crying and are more likely to be agreed upon by clinicians if they involve inspection than if they involve auscultation. The signs included in the IMCI algorithm were found to be largely appropriate but there is a need to

pay special attention during training to signs such as dehydration, neck stiffness and chest indrawing, for which agreement was not optimal. The difficulties surrounding the diagnosis of impaired consciousness in young children should be recognized and lead to increased vigilance in the diagnosis and management of neurological complications of illnesses in infancy.

Acknowledgements

We thank the parents and guardians of all children involved in the study; Dr P. Kibatala (Medical Director) and the staff of St Francis Designated District Hospital, without whose cooperation this study would not have been possible; the clinical officers of Ifakara Health Research and Development Centre for performing the clinical examinations; and the District Medical Officer, Dr F. Lwilla and his District Health Management Team for facilitating the acceptance of the research team in the local community. K. Senkoro assisted with data management. This study was funded by the Spanish Agency for International Cooperation (AECI) and the UNDP/World Bank/WHO Special Programme for Research Training in Tropical Diseases. Ifakara Health Research and Development Centre receives funds from the Swiss Agency for Development and Cooperation.

References

- Anonymous (1998). Health worker performance after training in integrated management of childhood illness: Western Province, Kenya, 1996-1997. *Morbidity and Mortality Weekly Report*, **47**, 998-1001.
- English, M., Murphy, S., Mwangi, I., Crawley, J., Peshu, N. & Marsh, K. (1995). Interobserver variation in respiratory signs of severe malaria. *Archives of Disease in Childhood*, **72**, 334-336.
- Gadomski, A. M., Khallaf, N., el Ansary, S. & Black, R. E. (1993). Assessment of respiratory rate and chest indrawing in children with ARI by primary care physicians in Egypt. *Bulletin of the World Health Organization*, **71**, 523-527.
- Gove, S. (1997). Integrated management of childhood illness by outpatient health workers: technical basis and overview. The WHO working group on guidelines for integrated management of the sick child. *Bulletin of the World Health Organization*, **75**, 7-24.
- Hansen, M., Sindrup, S. H., Christensen, P. B., Olsen, N. K., Kristensen, O. & Friis, M. L. (1994). Interobserver variation in the evaluation of neurological signs: observer dependent factors. *Acta Neurologica Scandinavica*, **90**, 145-149.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.
- Margolis, P. & Gadomski, A. (1998). Does this infant have

- pneumonia? *Journal of the American Medical Association*, **280**, 134.
- Molyneux, M. E., Taylor, T. E., Wirima, J. J. & Borgstein, A. (1989). Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. *Quarterly Journal of Medicine*, **71**, 441–459.
- Newton, C. R. J. C., Chokwe, T., Armstrong Schellenberg, J., Winstanley, P. A., Forster, D., Peshu, N., Kirkham, F. J. & Marsh, K. (1997). Coma scales for children with severe falciparum malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **91**, 161–165.
- Perkins, B. A., Zucker, J. R., Otiemo, J., Jafari, H. S., Paxton, L., Redd, S. C., Nahlen, B. L., Schwartz, B., Oloo, A. J., Olango, C., Gove, S. & Campbell, C. C. (1997). Evaluation of an algorithm for integrated management of childhood illness in an area of Kenya with high malaria transmission. *Bulletin of the World Health Organization*, **75**, 33–42.
- Schellenberg, D., Menendez, C., Kahigwa, E., Font, F., Galindo, C., Acosta, C., Armstrong Schellenberg, J., Aponte, J., Kimario, J., Urassa, H., Mshinda, H., Tanner, M. & Alonso, P. (1999). African children with malaria in an area of intense *Plasmodium falciparum* transmission: features on admission to the hospital and risk factors for death. *American Journal of Tropical Medicine and Hygiene*, **61**, 431–438.
- Singhi, S., Dhawan, A., Kataria, S. & Walia, B. N. (1994). Validity of clinical signs for the identification of pneumonia in children. *Annals of Tropical Paediatrics*, **14**, 53–58.
- Snow, R. W., Bastos de Azevedo, I., Lowe, B. S., Kabiru, E. W., Nevill, C. G., Mwankusye, S., Kassiga, G., Marsh, K. & Teuscher, T. (1994). Severe childhood malaria in two areas of markedly different falciparum transmission in East Africa. *Acta Tropica*, **57**, 289–300.

Received 30 January 2001; revised 5 July 2001; accepted for publication 12 July 2001

Announcements

Design of Vaccination Programmes From Seroepidemiology to Cost-Effectiveness A Warwick University Short Course

University of Warwick, Coventry, UK
8–12 July 2002

Further details can be obtained from Dr Stephen Hicks, Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK; phone +44 (0)2476 523540, fax +44 (0)2476 523701, e-mail s.j.hicks@warwick.ac.uk, web site <http://www.bio.warwick.ac.uk/shortcourses>

6th International Meeting on Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases (MEEGID-VI)

Paris, France
24–27 July 2002

For further information contact Michel Tibayrenc, Centre d'Etudes sur le Polymorphisme des Micro-organismes, IRD, Montpellier, France, e-mail michel.tibayrenc@mpl.ird.fr, web site <http://cepm.mpl.ird.fr>

Diagnostic Parasitology Course

Bethesda, Maryland, USA
22 July–2 August 2002

For further information contact Dr John H. Cross, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814-4799, USA; phone +1 301 295 3139, fax +1 301 295 1971.

ISOPS IV 4th International Symposium on Phlebotomine Sand Flies

Salvador, Bahia, Brazil
3–7 August 2002

For more information contact Executive Secretariat, Eventus System, Rua Oito de Dezembro, 547, Graça, 40150.000, Salvador, Bahia, Brazil; phone +55 71 264 3477, fax +55 71 264 0508, e-mail informa@eventussystem.com.br, website www.eventussystem.com.br