

The long-term outcome after severe trauma of children in Flanders (Belgium): a population-based cohort study using the International Classification of Functioning—related outcome score

Patrick Van de Voorde · Marc Sabbe · Roula Tsonaka ·
Dimitris Rizopoulos · Paul Calle · Annick De Jaeger ·
Emmanuel Lesaffre · Dirk Matthys ·
on behalf of the PENTA Study Group

Received: 6 May 2010 / Accepted: 25 July 2010 / Published online: 10 August 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Important long-term health problems have been described after severe paediatric trauma. The International Classification of Functioning (ICF) was developed as a universal framework to describe that health. We evaluated outcome in children after ‘severe’ trauma (defined as: hospitalised >48 h) by means of a questionnaire based on this ICF construct (IROS). Questionnaires were sent to children; one year after this trauma and to ‘control’ children without any previous ‘severe’ trauma. We created propensity score-matched pairs ($n=133$) and evaluated differences

in health perception. IROS characteristics were investigated by means of Item Response Theory models. We then estimated the health state of each individual based on his/her response pattern (factor score z_{01}) and investigated the effect of selected covariates with simple linear regression. Significant odds ratios for differences between matched groups ($p<0.05$) were observed for among others emotional problems, mobility, societal life and family burden, but not for chronic pain. Children in the trauma group showed, e.g. significant more physician (estimated relative risk $R' 1.7$)

Electronic supplementary material The online version of this article (doi:10.1007/s00431-010-1271-y) contains supplementary material, which is available to authorized users.

P. Van de Voorde (✉) · A. De Jaeger · D. Matthys
Paediatric Intensive Care Unit, University Hospital Ghent,
1K12IC, De Pintelaan 185,
9000 Ghent, Belgium
e-mail: patrick.vandevoorde@ugent.be

A. De Jaeger
e-mail: annick.dejaeger@ugent.be

D. Matthys
e-mail: dirk.matthys@ugent.be

M. Sabbe
Emergency Medicine, University Hospital Leuven,
Leuven, Belgium
e-mail: marc.sabbe@uz.kuleuven.ac.be

E. Lesaffre
Interuniversity Institute for Biostatistics,
Leuven, Belgium
e-mail: e.lesaffre@erasmusmc.nl

R. Tsonaka
Department of Medical Statistics and Bioinformatics,
Leiden University Medical Center,
Leiden, the Netherlands
e-mail: S.Tsonaka@lumc.nl

D. Rizopoulos · E. Lesaffre
Department of Biostatistics, Erasmus Medical Center,
Rotterdam, the Netherlands

D. Rizopoulos
e-mail: d.rizopoulos@erasmusmc.nl

P. Van de Voorde · P. Calle
Emergency Medicine, University Hospital Ghent,
Ghent, Belgium

P. Calle
e-mail: Paul.Calle@ugent.be

and psychologist (R' 3.5) visits. IROS primarily provides information from medium to high health burden levels and factor scores ranged from 0.41 (lowest) to 0.967 (highest burden). A significant impact on health burden could only be proven for the 'state at discharge' ($p=0.015$), although there was a tendency towards worse factor scores for children that were older, had a higher Injury Severity Score or after traffic injury. In conclusion, we showed that the burden of health problems for children and families after severe trauma is still high and physical, as well as psychosocial in nature. The health state at discharge seems to predict long-term outcome, which might be of importance in view of, e.g. trajectory assistance. IROS may provide an improved scoring system to evaluate outcome after (paediatric) injury or critical illness.

Keywords Trauma · Outcome · Child · International Classification of Functioning · Item Response Theory

Abbreviations

AIS	Abbreviated Injury Scale
CI	Confidence interval
e.g.	Example given
ICF-CY	ICF (WHO), version for children and youngsters
ICF	International Classification of Functioning (WHO)
i.e.	it est
IROS	ICF-related outcome score
IRT	Item Response Theory
ISS	Injury Severity Score
OR	Odds ratio
PENTA	A paediatric network around trauma (Flanders, Belgium)
POPC	Paediatric overall performance category
R'	Liddell's exact estimated relative risk
SD	Standard deviation
y	Years of age
z01	Factor score, rescaled to [0,1]

Background

Important long-term health problems have been described in adults and children after critical illness or severe trauma. These problems are seen in all dimensions of health and include among others emotional or intellectual functioning, but also locomotion, problems with sexuality, burden on caregivers [1, 18, 20, 26, 32]. Several authors therefore advise the systematic collection of outcome data in research and daily practice [2]. Still, studies describing long-term outcome of severely injured children are relatively sparse.

Moreover, published data are not necessarily representative for the local situation, because of cultural or regional health system differences. Existing studies often suffer from selection bias due to exclusion criteria, drop-out or instrument-related issues [8, 25, 37]. Further, most instruments measure only certain domains of health [14, 21, 22]. The choice of these domains is rather arbitrary and often not related to a strong underlying construct. Instruments are thus under discussion because of shortcomings in their content validity and/or in the justification of the methods used, but also because of problems with certain assumptions made (e.g. equality of interval distance, independence of attributes) [6, 10, 14, 15, 33]. In addition, for most instruments, the evidence of validity in severely injured children is still limited [27, 36].

In 2001, the World Health Organisation endorsed the International Classification of Functioning (ICF) as a conceptual framework and a 'common language' to describe health [38]. As part of the Flemish Paediatric Trauma Registry (PENTA), we evaluated long-term outcome in children and their families after severe trauma by means of a questionnaire based on this ICF construct: the ICF-Related Outcome Score (IROS) [35]. The aim of this report is to describe that outcome and the characteristics of the IROS instrument in doing so.

Methods

Patients and controls

Patients were recruited from the PENTA registry (Flanders Belgium, 2005) [35]. Review board approval and informed consent were obtained. All surviving children with 'severe' trauma, i.e. hospitalised >48 h were eligible ($n=229$). Exclusion criteria included loss of follow-up ($n=15$), language barrier ($n=19$) and child abuse ($n=4$). Follow-up was planned after 12 months. A questionnaire was sent by post, supported by telephonic reminder (up to three times); the response rate was 76.4% ($n=146$). The primary respondent was a parent or guardian. The control population ($n=265$) consisted of children without recent 'severe' trauma or severe chronic disease, recruited from different groups: children from hospital personnel, a nursery and a secondary school. We used propensity score matching and created matched pairs ($n=133$) between the trauma and control groups to correct for observed imbalances in the covariate distributions (age and gender of child, marital status and education level of proxy; see Table 1) [30]. Further analysis was done only in these matched groups.

For the trauma cases the median Injury Severity Score (ISS) was 9 (min 1; max 43). Thirty point eight percent had an ISS above 15; 21% were categorised as 'polytrauma'

Table 1 Descriptive statistics for the covariates in the IROS data set after the propensity score matching

		Control	Trauma	<i>p</i> value for difference
Age child	Mean (SD)	9.99 (5.12)	9.32 (5.34)	0.29
Gender child	Male	85 (63.9%)	85 (63.9%)	NA
Education proxy	Higher education or university	51 (38.3%)	48 (36.1%)	0.80
	Primary or secondary school	82 (61.7%)	85 (63.9%)	
Marital status proxy	Married/living together	116 (87.2%)	107 (80.5%)	0.18
	Single, widow(er)	17 (12.8%)	26 (19.5%)	

Age is given in years, for the other covariates we give absolute numbers (and percentage between brackets). We further give *p* values for the differences between the two matched groups. For gender, no *p* value is available (NA) because exactly the same number of males and females were included in the trauma and control group

[7]. In 38.6% of the patients, a serious to critical brain injury was scored (Abbreviated Injury Scale 1990). On the other hand, 17.9% had solely a minor to moderate injury of the extremities or pelvis. The median length of stay in hospital was 6 days (range 2–124 days). Circumstances of trauma were among others traffic in 35.7%, fall in 19.3% and burn in 12.8%. Paediatric Overall Performance Category (POPC) upon discharge was normal in 59.3%, mild disability in 27.9%, moderate disability in 7.9% and severe disability in 3.6% [11]. Trauma cases did not differ significantly from non-responders or excluded cases for ISS, gender, age or main diagnosis (*t* test for ISS or age, Fisher's exact for difference in proportions).

ICF and the IROS instrument

The ICF is a classification of health in its broadest sense. It is cross-cultural developed and truly generic. Recently, a version for children and youngsters became available (ICF-CY) [39]. The ICF-CY, as the ICF, is a taxonomy and consists of four separate constructs: bodily dysfunction (b), structural changes (s), daily activity/participation (d) and environmental influences (e). Codes are made up using an alphanumeric system starting with the letter of the construct; followed by a first digit for the particular chapter, then two for the second level description, and finally an optional third level single digit, e.g. d4 mobility, d450 walking, d4502 walking on different surfaces. The nature of somebody's functioning and health can thus be described in more or less detail. Severity scaling, by adding a number after the decimal point (e.g. 450.2), is similar for all items and inherently asymmetrically constructed: 0 none (0–4%), 1 mild (5–24%), 2 moderate (25–49%), 3 severe (50–95%) and 4 total (96–100%).

IROS includes 54 questions representing 99 'second-level' ICF-CY codes (e.g. d760 problems with family relations), as well as questions on medical consumption, family impact and environmental influences ([Electronic Supplementary Material: IROS.wmf](#)). In addition, four 'sum' scores (physical, mental, social and total) measure

the overall burden of health problems on an 11-point adjectival scale between 0 'no burden' and 10 'maximum burden'.

For reasons of feasibility, many problems are only scored under a more general code, ending with 8 or 9: 'otherwise or not further specified', e.g. b729 problems with functions of joint and bones, other specified and unspecified, which then also covers for instance b710 problems with the mobility of joints. The 99 ICF codes used thus capture the whole spectrum of the ICF (and therefore of 'health') [32, 37]. Since several questions correspond to more than one code, space for free text clarification is provided. To promote uniformity in scaling, we added a scoring description sheet.

Data structure and statistical methodology

Given the unlimited amount of possible health states, grouping of codes and severity levels is necessary for any evaluation beyond individual consulting and for further statistical analysis. We grouped codes according to their 'first level', in line with the ICF construct (e.g. item D1 'learning' covering all codes starting with d1, e.g. d160 focussing attention; see also the item list in Table 2). However, b1 and b2 were split arbitrary into two items, so that respectively emotional (b1a: [b130, b134 and b152]) and other mental (b1b) codes, as well as pain (b2a: [b280]) and other sensation (b2b) could be appreciated separately. We also created an item summarising all 'family burden' codes (F). Further, severity levels were regrouped to a three-point scale (0 'normal to mild', 1 'moderate' and 2 'severe to very severe'). The eventual score in each item was defined as the highest level of severity in any of the ICF codes that corresponded to that item. For the 'burden' questions, scores were merged to four categories (0–1, 2–3, 4–5, 6–10).

First, to test for significant differences between the trauma and the control group, we used the Generalized Estimating Equations approach for ordinal data [16]. Statistical inference was obtained using the Wald test.

Table 2 Differences between matched pairs

Item	O.R. (95% CI)	<i>p</i> value	Percent of trauma cases with important problems
B1a: emotional	4.03 (1.57; 10.38)	0.004*	18.8
B1b: other mental	2.77 (0.93; 8.27)	0.067	28.6
B2a: pain	4.05 (0.48; 34.3)	0.200	9.8
B2b: other sensations	1.46 (0.52; 4.10)	0.468	20.3
B3: speech and voice	2.52 (0.49; 12.96)	0.268	6.8
B4: cardiorespiratory	0.96 (0.35; 2.63)	0.937	14.3
B5: intestinal, metabolic and horm.	3.27 (1.01; 10.56)	0.047*	15.8
B6: urogenital and reproduction	0.49 (0.09; 2.78)	0.423	3
B7: locomotor system	17.38 (2.54; 118.91)	0.004*	21.1
B8: skin	0.99 (0.20; 4.78)	0.985	5.3
D1: learning	3.07 (0.91; 10.39)	0.071	12.8
D2: tasks and demands	6.74 (1.68; 27.08)	0.007*	19.5
D3: communication	1.52 (0.41; 5.66)	0.530	4.5
D4: mobility	14.42 (3.25; 63.92)	<0.001*	18
D5: self-care	9.03 (1.99; 41.04)	0.004*	12
D6: domestic life	2.58 (0.48; 13.79)	0.268	3.8
D7: social relations	6.93 (0.94; 51.12)	0.057	15
D9: societal life	23.57 (3.44; 161.76)	0.001*	15
S: structural	1.52 (0.28; 8.17)	0.624	12.8
F: family	4.72 (1.40; 15.94)	0.012*	24.8
PB: physical burden	13.52 (1.89; 96.85)	0.010*	21.1
MB: mental burden	6.19 (0.86; 44.60)	0.007*	15.8
SB: social burden	11.34 (1.46; 88.10)	0.020*	19.5
TB: total burden	3.77 (1.09; 12.99)	0.036*	20.3

P values and odds ratios (OR) with 95% confidence intervals (CI) for differences between the two groups in the 24 items, based on a Generalized Estimating Equations analysis for matched pairs (* if $p < 0.05$). There is no d8 in this paediatric version. As an illustration, we added in the last column the percentages of individuals out of the trauma group ($n=133$) that had a 'moderate to very severe' score for that item or a score of 3 and more in case of the four 'burden' items

Differences between matched groups in, e.g. medical consumption or individual codes were further evaluated using Liddell's exact test [19].

Second, to evaluate the reproducibility of our findings, a double evaluation was made in 39 children: in 11 'control' children, IROS was scored simultaneously by both parents; in 28 'trauma' children, both evaluations were made by the same proxy (97% mother) within a 2–3-week interval. We also obtained evaluations from both child and parent in 113 adolescents (84 controls, 29 trauma cases). We tested for marginal homogeneity and agreement between both raters, using the powerful Bhapkar test [3].

Third, we investigated the characteristics of the IROS items by means of polytomous Item Response Theory (IRT) models [31]. IRT provides a powerful class of models that can be used to identify items that are informative for specific levels of health, to evaluate the ability of an instrument to measure health and most importantly to derive a score measuring the health status of a child based on his/her responses. For our analysis, we used Samejima's graded response model [31]. We plotted item and test information curves that determine how accurate each level of health status is being estimated by each item and by the whole set of items. We then obtained an estimate for the

level of health (burden) of each individual (i.e. for each IROS response pattern): the factor score z . For interpretability reasons, these factor scores were rescaled in $[0,1]$ using the formula $\exp(z) / (1 + \exp(z))$, where 0.5 represents the median burden and lower values, better health (less burden).

Finally, based on these factor scores, we investigated the effect of collected covariates (e.g. patient characteristics, ISS, diagnosis) on the eventual patient health status by means of simple linear regression and corresponding *F* test.

Since the final purpose is to describe health in all children irrespective of their level of illness, control and trauma cases were analysed together without considering differential item functioning. Importantly, we made in our IRT analysis a distinction between items that are treated as *effect* indicators (D 1–9, G and burden 'sum' scores) and items that are more likely *causal* indicators (B1–B8, S) [10, 24, 33]. In clinical trial data, a subset of items may indeed rather act as causal indicators, i.e. variables that directly affect the underlying latent variable 'health' and not the other way round (e.g. pain will give 'bad' health but not all patients with 'bad' health will have pain). All results were estimated with the statistical environment R (v2.8-1) and specifically, the ltm package [29]. Additional functions were written to fit the model that accounts for the effect of

causal indicators; p values were produced by means of the Wald test. A p value less than 0.05 was considered significant. As this is an exploratory study, no correction for multiple testing was envisaged.

Results

Cases versus controls

We first looked at differences in health perception (Table 2). Statistical significant differences were seen for among others emotional problems, mobility, societal life, burden on family and for all burden sum scores. To illustrate this further, for instance for the F (family) item, 24.8% trauma cases scored a ‘moderate’ to ‘very severe’ problem, while only 6.8% in the control group. In line with these results, children in the trauma group had significant more health visits in the last month. The estimated relative risk R' for “visiting more than once in the last month” was 1.7 (exact 95% CI=[1.1, 2.8]) for physician visits, 13 (exact 95% CI=[3.3, 113]) for physiotherapist visits and 3.5 (exact 95% CI=[1.1, 14.6]) for psychologist visits.

Reproducibility

Using the Bhapkar test, we found no statistically significant difference in the between-parents comparison, except for sensational problems and self-care in the trauma cases (in four out of 28 trauma cases scores differed; Table 3). Differences between parent and child were far more pronounced. In the trauma group, these differences reached statistical significance for among others pain, task and demands. In each of these items differences were caused by a higher severity score given by the parent, e.g. for pain eight (out of 29) parents scored moderate to severe problems whereas the child scored only mild. On the contrary, for the control cases significant differences were caused by higher scores for the child, e.g. for emotional 26 children scored moderate to severe problems and the parent only mild. Overall, differences were observed most often around the mild to moderate cut-off.

Characteristics of the IROS items and covariate effects

We first estimated how well the ‘effect indicators’ on their own would capture a unique construct, i.e. the underlying

Table 3 Observer variability

Item	Parent–parent trauma cases	Parent–parent controls	Parent–child Trauma cases	Parent–child controls
B1a: emotional	0.067	NA	0.143	<0.001*
B1b: other mental	0.309	0.294	0.246	0.763
B2a: pain	0.165	0.118	<0.001*	0.04
B2b: other sensations	0.031*	0.118	0.738	0.248
B3: speech and voice	0.142	NA	0.309	NA
B4: cardiorespiratory	0.561	0.294	0.409	0.015*
B5: intestinal, metabolic and horm.	>0.999	0.294	0.654	0.017*
B6: urogenital and reproduction	0.309	0.294	NA	0.314
B7: locomotor system	0.067	0.118	>0.999	0.013*
B8: skin	0.309	NA	NA	0.739
D1: learning	NA	0.294	0.031*	0.739
D2: tasks and demands	0.309	NA	0.021*	0.412
D3: communication	0.309	0.294	0.067	0.563
D4: mobility	0.067	NA	0.409	0.078
D5: self-care	0.031*	NA	0.309	NA
D6: domestic life	NA	NA	NA	0.563
D7: social relations	0.561	NA	0.309	0.09
D9: societal life	0.142	0.142	0.561	0.175
S: structural	0.561	0.561	0.044*	0.592
F: family	0.142	0.142	0.654	0.0279*
PB: physical burden	0.466	0.560	0.051	0.068
MB: mental burden	0.195	0.294	0.377	0.003*
SB: social burden	0.185	NA	0.16	0.771
TB: total burden	0.073	0.295	0.108	0.229

P values for significant disagreement between raters by means of the Bhapkar test (* if <0.05). For some items, p values are not available (NA) because the same reply was given by all respondents

'health status'. Here, Cronbach's alpha was 0.890 (Bootstrap 95% CI=[0.844, 0.919]), indicating very good reliability of IROS.

To evaluate the information content of the IROS instrument, we plotted the item information curves for each of the effect indicators and sum scores and the overall test information curve. The IROS scale primarily provides information from medium to high health burden levels: information in (-4, 0)=1.14%; in (0, 4)=96.94%. We then obtained an estimate for the health state of each individual based on his/her response pattern, i.e. the factor score z_{01} . There were in total 168 unique response patterns in our sample. To illustrate this further: a patient with a response pattern of 0000000000000/00000000000 (none to mild for all 13 effect and 11 causal indicators) will obtain a factor score z_{01} of 0.41; on the contrary a patient with a response pattern of 2212221213333/1202000200 on the 24 items will have a factor score close to 1 (0.967). Importantly, we observed no considerable changes in factor scores before versus after accounting for the causal effects.

Finally, based on the factor scores, we investigated the effect of several collected variables with simple linear regression. In Table 4 we present F test-derived p values for the impact of the selected covariate on the latent trait 'health status'. Significant impact could only be proven for State at discharge (POPC), although there was a clear tendency towards worse factor scores for children that were older, had a higher ISS or after traffic injury.

Discussion

'Outcome' is a multidimensional (biopsychosocial), contextual (person, environment) and dynamic (function of the a priori health state) concept. It is inherently subjective in concept and measurement. Instruments capturing outcome should equally be multidimensional and contextual and describe health from a patient's perspective. When evaluating paediatric populations, they should be relevant for children, including infants and their families. As IROS is developed out of the ICF-CY, its underlying construct is well-defined and its content validity very high [39].

To evaluate health status, one could make comparison with an a priori state. Yet most often, this has to be scored in retrospect with a high risk of recall bias and halo effects. A good alternative is to use a representative population sample or a matched-control group. Still, appreciating differences between groups is not easy. Existing differences might be obscured by, e.g. intervening diseases, developmental changes, adaptation and response shift [33]. We observed a significant difference between the trauma and matched-control group, especially in items that were expected to be impaired: mobility, social life, family, emotional etc. [32, 37]. Contrary to expected and available adult evidence, this was not the case for chronic pain [28]. Trauma cases had higher healthcare needs than controls, as illustrated by the number of healthcare visits. Importantly, children in the trauma group were not all that severely injured and 'control' children were not all 100% healthy, they just did not have a recent injury. In fact, the trauma

Table 4 Covariate analysis

Variable	categories	F test: p value
Age	0–10y; 11–18y	0.103
Gender	male; female	0.359
Proxy education	Primary/secondary school; higher education/university	0.358
Proxy marital state	Married/living together; single/divorced/Widow(er)	0.697
Main diagnosis	Brain; abdomen; spine; extremities/pelvis; facial/neck; thorax; vascular; other	0.565
ISS	1–43	0.16
Length of stay	2–124 days	0.39
State at discharge	POPC [42]: normal, mild, moderate, severe	0.015*
Trauma circumstances	Traffic–sports–burn–fall–other	0.304
Circumstances: traffic	Traffic; not traffic	0.091
Circumstances: burn	Burn; not burn	0.894
Diagnosis: abdomen	Any AIS Abdomen 3–5; not AIS abdomen 3–5	0.301
Diagnosis: spine	Any AIS Spine 3–5; not AIS Spine 3–5	0.471
Diagnosis: brain	Any AIS Brain 3–5; not AIS Brain 3–5	0.932
Diagnosis: thorax	Any AIS Thorax 3–5; not AIS Thorax 3–5	0.497
Diagnosis: extremities	Any AIS extremities 3–5; not AIS extremities 3–5	0.86
Participating centre	18 different hospitals	0.384

F test-derived p values (* if <0.05), based on simple linear regression, for the impact of selected variables on the health status' estimate based on the factor scores from the IRT model

group seemed to be divided into two subgroups: one ‘larger’ group that had but minimal residual problems after 12 months and one ‘smaller’ subgroup with important, long-lasting problems in a wide range of health items for patient and caregivers. More restrictive inclusion criteria might have generated a more homogeneous sample but since health problems are not necessarily or solely related to, e.g. injury severity, it might as well have given significant information loss.

A certain degree of observer variability is unavoidable (because of among others' cognitive and memory effects). In our sample, parent–parent differences did not reach significance except in two items. If there were differences, then these were most often situated around the cut-off mild vs. moderate. In line with existing literature, parent–child differences were more pronounced [9, 13, 36]. In the trauma group, children tended to underscore compared to their parents. This might be related to more adaptation and satisficing by the child. For the controls, on the contrary, existing differences were most often related to higher severity scores by the child. Overall in relation to their medical consumption, we found these youngsters to report a higher level of problems and show a higher level of item interdependence (‘halo bias’). Although child reporting might provide important additional information, obtaining a proxy-reported questionnaire in every patient is probably the most consistent and reliable option for among others trauma registries [17, 25]. First of all, for a lot of patients, only a proxy report will be available (selection bias) [25, 36]. Second, while parents might be underestimating certain non-observable problems and are themselves influenced by their child’s health state, they are more accessible and less influenced by knowledge or short-time memory effects, response shift or satisficing [5, 9, 23, 34]. Most often, it is them who decide about health care resource use [30].

Item properties and the effect of covariates

The IROS instrument primarily provides information from medium to high health burden levels. In line with this, effective factor scores (z01) ranged from 0.41 to 0.967. This implies that IROS items distinguish less well between patients from optimal health to mild or moderate health problems, but are very appropriate in distinguishing patients with more severe health states. This is not surprisingly given the asymmetric severity scaling and the inherent skewed distribution. However, where left-sided sensitivity might be important for individual counselling, the primary focus of any health measurement tool should be on the above median burden patients and for this IROS performance is great.

To further identify the characteristics of the ‘worse outcome’ subgroup, we evaluated the effect of several

covariates, based on the factor scores from the IRT model. Significant impact could only be proven for the patient state at discharge as scored by the POPC. Demographic parameters, trauma circumstances, injury diagnosis or length of stay all failed to reach significance. However, children that were older, had a higher ISS or after traffic injury tended to have worse factor scores. Most likely, these children represent a distinct subgroup. The fact that the ISS on its own, although it describes severity of injuries, is not strongly related with long-term outcome is not that surprising. Similar observations have been made by other authors and especially in children the usefulness of the ISS to predict outcome remains unproven, if not doubtful [12].

Current limitations and future directions

Although not matched for a priori health state, evaluating the trauma children vs. a control group, matched for age, gender and proxy characteristics, is probably the closest we can get to ‘attributable burden’—more so than with normative data. One well-acknowledged host factor in childhood injury is the existence of attention deficit hyperactivity disorder or other conduct disorders [4]. However, in our population, no difference between both groups in methylphenidate use was observed.

We evaluated outcome at 12 months post-injury, as this is a point of presumed steady state. Ideally, we would have had intermittent points of evaluation, to also evaluate recovery patterns, but for this we lacked the logistics. We are planning a second evaluation 5 years after the initial trauma.

We evaluated the effect of selected covariates on the outcome of these severely injured children and could only identify a significant impact of health state at discharge. However, other covariates are known to be of possible influence on long-term outcome after trauma and were not available here: social support, a priori mental health, etc. [20, 32]. Further, exploring these covariates is imperative, as they might provide us with clues to help patient and surroundings: better outcome, less unmet needs....

As for the IRT model itself, the distinction between causal and effect indicators is of course arbitrary. From a theoretical point of view however, this allows for a more realistic approach to the measurement of health status and acknowledges the importance of each of the items considered. Still, accounting for causal effects did not change factor scores considerably and thus, for future samples, only using the effect indicators to calculate factor scores might be an acceptable alternative. Importantly, since the estimated factor scores value (the burden of) an individual person’s health between 0 and 1, they can be used to compare individuals and groups, for instance in randomised controlled trials, economical evaluations, etc.

Overall our sample is small, especially given the skewed distribution of the responses. IROS is robust in construct and content, but evidence for reproducibility, discriminating capacity and impact of covariates is therefore still only indicative. We will need to test this further in future patient cohorts. In view of the truly generic and multicultural underlying construct, IROS should be equally valid for other medical conditions, age groups and/or in other languages or countries. It is our aim to explore these possibilities but also we encourage (and are happy to aid) others to incorporate IROS in their research or clinical setting.

Conclusion

IROS provides an improved scoring system to evaluate the burden of health problems after injury or critical illness. We described long-term outcome for children after severe trauma in Flanders. Although the perceived health state after 12 months was for several ‘trauma’ children comparable to controls, for a specific subgroup the burden of health problems was still high. More specifically, these problems were physical, as well as social or psychological in nature (emotions/behaviour, intestinal, mobility, self-care, executing tasks, societal life). For this group, the burden on family was significant. It is not easy to predict which child will have a bad long-term outcome after trauma as in our study only the health state at discharge was significantly related. Still the latter might be of importance in view of for instance trajectory assistance. Future research is needed to describe recovery patterns after severe paediatric trauma; to evaluate additional covariates and their relation to outcome and also to further explore the ICF format as a way to report that outcome.

Acknowledgments We like to thank Daphne Christiaens, Anneleen Vantomme and Helena Kourdioukova, research staff of the PENTA study centre, for their participation in data collection and validation.

The following Flemish hospitals (Belgium) participated in the PENTA registry: ASZ Aalst, St. Blasius Dendermonde, UZ Gent, AZ MMSJ Gent, Jan Palfijn ZH Gent, AZ St. Lucas Gent, AZ Oudenaarde, AZ MM St. Niklaas, AZ Groeninge Kortrijk, HH ZH Menen, HH ZH Roeselare, AZ St. Maarten Mechelen, AZ St. Maarten Duffel, St. Vincentius ZH Antwerpen, Klina Brasschaat, AZ Stuivenberg Antwerpen, ZH Noord-Limburg Lommel, ZOL Genk, UZ Gasthuisberg Leuven.

All of the above authors were actively involved in study design or data gathering/analysis and revisions of the initial draft. They all gave final approval to the version currently submitted.

No author declared any conflict of interest. The PENTA study as a whole was funded by the Flemish fund for scientific research FWO partim Levenslijn. Dr. Tsonaka further acknowledges support from the Research Fund KU Leuven; PDMK/08/179. There has been no further involvement from the funding sources or in the study design, or in the manuscript writing or submission.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Aitken ME, Tilford JM, Barrett KW, Parker JG (2002) Health status of children after admission for injury. *Pediatrics* 110:337–342
2. Angus DC, Carlet J, on behalf of the 2002 Brussels Roundtable Participants (2003) Surviving intensive care: a report from the 2002 Brussels roundtable. *Intensive Care Med* 29:368–377
3. Bhapkar VP (1966) A note on the equivalence of two test criteria for hypotheses in categorical data. *J Am Stat Assoc* 61: 228–235
4. Bruce B, Kirkland S, Waschbusch D (2007) The relationship between childhood behavior disorders and unintentional injury events. *Paediatr Child Health* 12:749–754
5. Capuzzo M, Grasselli C, Carrer S et al (2000) Quality of life before intensive care admission: agreement between patient and relative assessment. *Intensive Care Med* 9:1288–1295
6. De Vet HC, Ader HJ, Terwee CB, Pouwer F (2005) Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual Life Res* 14:1203–1218
7. Dick WF, Baskett PJF (1999) Recommendations for uniform reporting of data following major trauma—the Utstein style. A report of a Working Party of the International Trauma, Anaesthesia and Critical Care Society (ITACCS). *Resuscitation* 42:81–100
8. Dijkers MPJM (2005) Quality of life of individuals with spinal cord injury: a review of conceptualization, measurement, and research findings. *J Rehabil Res Dev* 42:S87–S110
9. Eiser C, Morse R (2001) Can parents rate their child’s health-related quality of life? Results of a systematic review. *Qual Life Res* 10:347–357
10. Fayers PM, Hand DJ (2002) Causal variables, indicator variables and measurement scales: an example from quality of life. *J R Stat Soc* 165:233–261
11. Fiser DH (1992) Assessing the outcome of pediatric intensive care. *J Pediatr* 121:68–74
12. Foreman BP, Caesar RR, Parks J et al (2007) Usefulness of abbreviated injury score and the injury severity score in comparison to the Glasgow Coma Scale in predicting outcome after traumatic brain injury. *J Trauma* 62:946–950
13. Frost MH, Reeve BB, Liepa AM, Mayo/FDA patient-reported outcomes consensus meeting group et al (2007) What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 10:S94–S105
14. Gabbe BJ, Williamson OD, Cameron PA, Dowrick AS (2005) Choosing outcome assessment instruments for trauma registries. *Acad Emerg Med* 12:751–758
15. Hanita M (2000) Self-report measures of patient utility: should we trust them? *J Clin Epidemiol* 53:469–476
16. Heagerty P, Zeger S (1996) Marginal regression models for clustered ordinal measurements. *J Am Stat Assoc* 91:1024–1036
17. Hofhuis J, Hautvast JL, Schrijvers AJ, Bakker J (2003) Quality of life on admission to the intensive care: can we query the relatives? *Intensive Care Med* 29:974–979
18. Levy MM (2004) The burden of caregiving on families of ICU survivors. In: Angus DC, Carlet J, Vincent JL (eds) *Surviving intensive care*, 1st edn. Springer, Berlin, pp 63–72

19. Liddell FDK (1983) Simplified exact analysis of case-referent studies; matched pairs; dichotomous exposure. *J Epidemiol Community Health* 37:82–84
20. Mc Carthy ML, MacKenzie EJ, Durbin DR, for the children's health after trauma study group et al (2006) Health-related quality of life during the first year after traumatic brain injury. *Arch Pediatr Adolesc Med* 160:252–260
21. McCarthy ML (2007) Measuring children's health-related quality of life after trauma. *J Trauma* 63:S122–S129
22. McKenzie EJ, McCarthy ML, Ditunno JF, Pennsylvania Study Group on Functional Outcomes Following Trauma et al (2002) Using the SF-36 for characterizing outcome after multiple trauma involving head injury. *J Trauma* 52:527–534
23. Meiser-Stedman R, Smith P, Glucksman E et al (2007) Parent and child agreement for acute stress disorder, post-traumatic stress disorder and other psychopathology in a prospective study of children and adolescents exposed to single-event trauma. *J Abnorm Child Psychol* 35:191–201
24. Moustaki I (2003) A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *Br J Math Stat Psychol* 56:337–357
25. Needham DM, Dowdy DW, Mendez-Tellez PA et al (2005) Studying outcomes of intensive care unit survivors: measuring exposures and outcomes. *Intensive Care Med* 31:1153–1160
26. Polinder S, van Beeck EF, Essink-Bot ML et al (2007) Functional outcome at 2.5, 5, 9, and 24 months after injury in the Netherlands. *J Trauma* 62:133–141
27. Randolph AG, Graham R (2004) Measuring the health status of pediatric ICU survivors. In: Angus DC, Carlet J, Vincent JL (eds) *Surviving intensive care*, 1st edn. Springer, Berlin, pp 107–118
28. Rivara FP, MacKenzie EJ, Jurkovich GJ et al (2008) Prevalence of pain in patients 1 year after major trauma. *Arch Surg* 143:282–287
29. Rizopoulos D (2006) ltm: an R package for latent variable modeling and item response theory analysis. *J Stat Softw* 17:1–25
30. Rosenbaum P, Rubin D (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 39:33–38
31. Samejima F (1969) Estimation of latent ability using a response pattern of graded scores. *Psychom Monogr Suppl* 34:100–114
32. Slomine BS, McCarthy ML, Ding R, The CHAT Study Group et al (2006) Health care utilization and needs after pediatric traumatic brain injury. *Pediatrics* 117:663–674
33. Streiner DL, Norman GR (2003) *Health measurement scales: a practical guide to their development and use*. Oxford University Press, New York
34. Theunissen NCM, Vogels TGC, Koopman HM et al (1998) The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res* 7:387–397
35. Van de Voorde P, Sabbe M, Calle P, on behalf of the PENTA study group et al (2008) Paediatric trauma and trauma care in Flanders (Belgium). Methodology and first descriptive results of the PENTA registry. *Eur J Pediatr* 167:1239–1249
36. Willis CD, Gabbe BJ, Butt W, Cameron PA (2006) Assessing outcomes in paediatric trauma populations. *Injury* 37:1185–1196
37. Winthrop AL, Brasel KJ, Stahovic L et al (2005) Quality of life and functional outcome after pediatric trauma. *J Trauma* 58:468–474
38. World Health Organization (2001) *The international classification of functioning, disability and health*. World Health Organisation, Geneva
39. World Health Organization (2007) *The international classification of functioning, disability and health; children and youngsters version*. World Health Organisation, Geneva