

Smooth Operator – Modifying the Anhøj Rules to Improve Runs Analysis in Statistical Process Control

by Jacob Anhøj, Tore Wentzel-Larsen

Abstract An abstract of less than 150 words.

Introduction

Within statistical process control (SPC) runs analysis is being used to detect persistent shifts in process location over time.

Runs analysis deals with the natural limits of number of runs and run lengths in random processes. A run is a series of one or more consecutive elements of the same kind, for example heads and tails, diseased and non-diseased individuals, or numbers above or below a certain value. A run chart is a point-and-line chart showing data over time with the median as reference line. In a random process, the data points will be randomly distributed around the median, and the number and lengths of runs will be predictable within limits. All things being equal, if the process shifts, runs tend to become longer and fewer. Consequently, runs analysis may help detect shifts in process location. Process shifts are one form of non-random variation in time series data that are of particular interest to quality control and improvement. If a process shifts, it may be the result of planned improvement or unwanted deterioration.

Several tests (or rules) based on the principles of runs analysis for detection of shifts exist. In previous papers we demonstrated that the currently best performing rules with respect to sensitivity and specificity to shifts in process location are two simple tests (Anhøj and Olesen, 2014; Anhøj, 2015; Anhøj and Wentzel-Larsen, 2018):

- Shifts test: one or more unusually long runs of data points on the same side of the centre line.
- Crossings test: the curve crosses the centre line unusually few times.

Collectively, we refer to these tests as the Anhøj rules. Critical values for run length and number of crossings depend on the total number of data points in the chart. The number of crossings follow a binomial distribution, $b(n - 1, 0.5)$, where n is the number of data points and 0.5 the success probability. Thus, the lower prediction limit for number of crossings may, for example, be set to the lower 5th percentile of the corresponding cumulative binomial distribution (Chen, 2010). However, no closed form expression exists for the distribution of longest runs. Consequently, the upper prediction limit for longest runs has traditionally been either a fixed value (usually 7 or 8) (Carey, 2002) or an approximate value depending on n as with the Anhøj rules: $\log_2(n) + 3$ rounded to the nearest integer (Schilling, 2012).

Using simulations, we have shown that runs analysis using the Anhøj rules are comparable or even better at detecting minor to moderate *persistent* shift than the widely used Western Electric control chart rules (Anhøj and Wentzel-Larsen, 2018).

Each of the two tests has an overall specificity (true negative proportion) around 95%. The sensitivity (true positive proportion) of a test depends on the size of the shift (signal) relative to the random variation inherent in the process (noise). When applied together, the sensitivity increases, while the specificity decreases a bit and fluctuates between 89% and 97% depending on the number of available data points in the chart. See the red line in (Figure 1).

Historically, runs tests have mainly been studied in isolation as individual tests. But what is really of interest – because the rules are linked, when one goes up, the other goes down – is the properties of the joint distribution of longest runs and number of crossings.

We recently released an R package, **crossrun** (Wentzel-Larsen and Anhøj, 2018), that includes functions for calculating the joint probabilities of the number of crossings (C) and longest runs (L) in random data series of different lengths (N) and with and without shifts in process location expressed in standard deviation units (SD). Figure 2 illustrates this for a run chart with $N = 11$ and $SD = 0$ (no shift). To avoid very small numbers, the probabilities are shown using the times representation, that is the probabilities times 2^{N-1} . The red box encloses the combinations of C and L that would indicate random variation according to the Anhøj rules (true negatives). The area outside the box represents combinations of C and L that would indicate non-random variation (false positives).

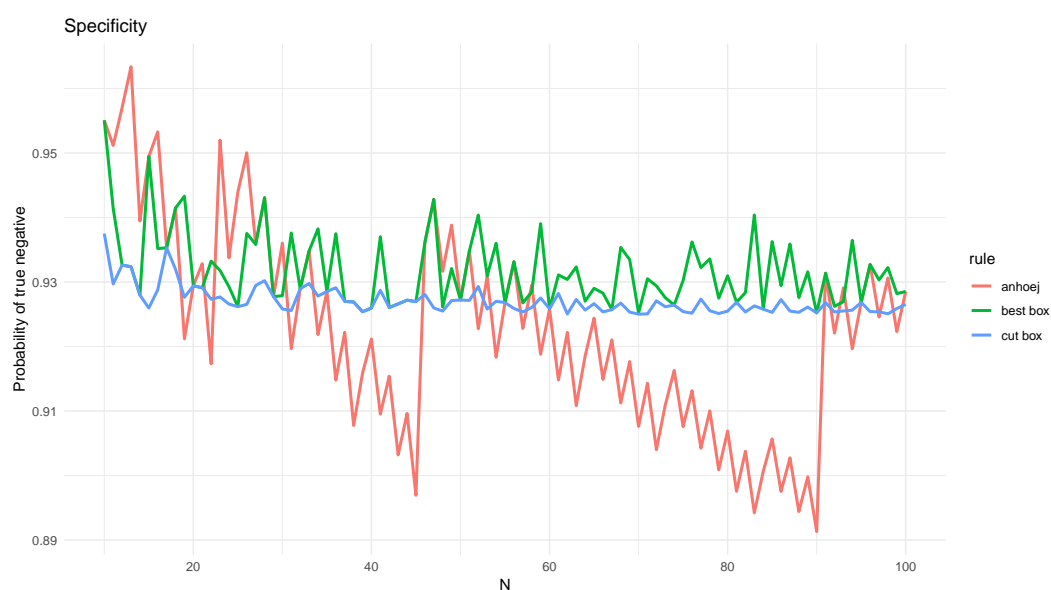


Figure 1: Specificity of the anhøj, best box, and cut box rules. N = number of data points in run chart.

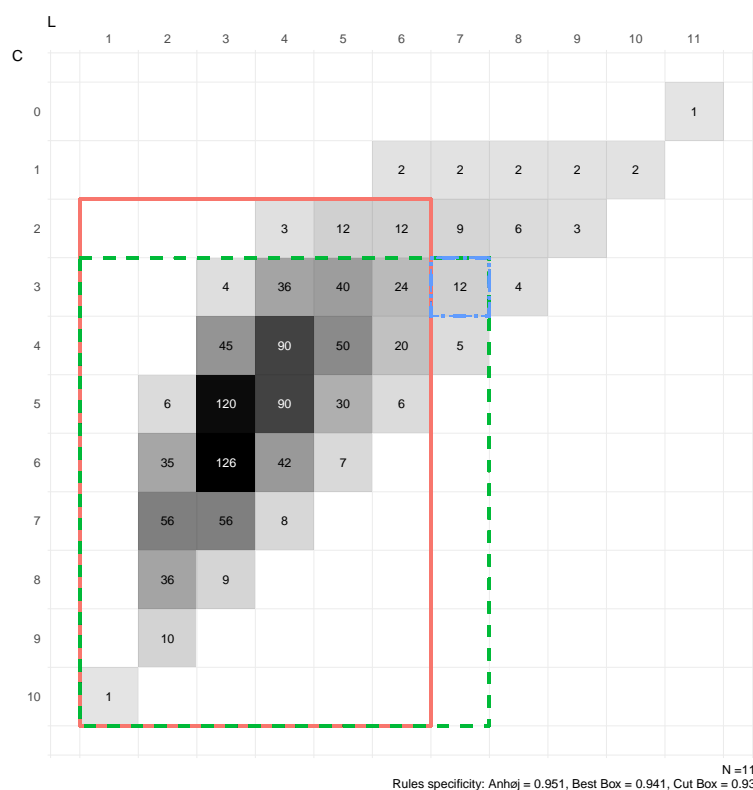


Figure 2: Borders of the anhøj, best box, and cut box rules for $N = 11$ data points. C = number of crossings, L = longest run. The numbers in the cells are times representation of of the joint probabilities of longest run and number of crossings. Anhøj = red solid, best box = green dashed, cut box = blue dot-dashed.

With the `crossrun` package it became feasible to calculate exact joint probabilities of C and L for different N and SD. And consequently, it is feasible to investigate the diagnostic properties of run charts using exact values for specificity and sensitivity rather than values based on time consuming and complicated simulation studies.

As is clearly visible in Figure 1 the specificity of the Anhøj rules (red line) jumps up and down as N changes. This is a result of the discrete nature of the two tests – especially the shifts test. Although the specificity of the Anhøj rules does not decrease continuously as N increases as is the case with other rules, we hypothesised that it would be possible to improve the diagnostic value further by smoothing the specificity using minor adjustments to C and L depending on N.

The aim of this study was to suggest a procedure for smoothing the diagnostic properties of the Anhøj runs rules.

Methods

Likelihood ratios to quantify the diagnostic value of runs rules

The value of a diagnostic test is traditionally described using terms like sensitivity and specificity. These parameters express the probability of detecting the condition being tested for when it is present and not detecting it when it is absent:

$$\text{Specificity} = P(\text{no signal} \mid \text{no shift}) = P(\text{true negative})$$

$$\text{Sensitivity} = P(\text{signal} \mid \text{shift}) = P(\text{true positive})$$

However, we usually seek to answer the opposite question: what is the likelihood that a positive or negative test actually represents the condition being tested for, which in our case is a shift in the underlying process? Likelihood ratios (LR) do this:

$$\text{LR+} = \text{TP/FP} = \text{sensitivity}/(1 - \text{specificity})$$

$$\text{LR-} = \text{FN/TN} = (1 - \text{sensitivity})/\text{specificity}$$

A likelihood ratio greater than 1 speaks in favour of the condition, while a likelihood ratio less than 1 speaks against the condition. As a rule of thumb, a positive likelihood ratio greater than 10 is considered strong evidence that the condition is present. A negative likelihood ratio smaller than 0.1 is considered strong evidence against the condition (Deeks and Altman, 2004). Thus, likelihood ratios are useful measures of the diagnostic value of run charts (Anhøj, 2015; Anhøj and Wentzel-Larsen, 2018).

Best box and cut box adjustments to improve the Anhøj rules

To fix some terms, we define a box as a square region $C \geq c, L \leq l$. These are the square regions that may be used to define non-random variation. The corner of the box is its upper right square $C = c, L = l$. In Figure 2 the box $C \geq 2, L \leq 6$, marked with red, specifies the Anhøj rules for $N = 11$. The corner of this box is the square $C = 2, L = 6$.

Based on the `crossrun` package we developed two functions, `bestbox()` and `cutbox()` that automatically seek to adjust the critical values for C and L to balance between sensitivity and specificity requirements. Specifically, the `bestbox()` function finds the box with highest sensitivity for a pre-determined shift (the target shift), among boxes with specificity \geq a pre-determined value (the target sensitivity). The `cutbox()` function subsequently cuts squares from the upper and right borders of the best box, starting from the upper right corner while keeping specificity \geq its target value, and the sensitivity for the target shift as large as possible. The result of `cutbox()` is not necessarily a box, but a reasonable region for declaring non-random variation with only a few squares close to the corner possibly removed.

In this study we used a target specificity of 92.5%, which is close to the actual average specificity for the Anhøj rules for $N = 10$ -100, while the target shift was set at 0.8.

Figure 2 illustrates these principles for a run chart with 11 data points. Thus, for $N = 11$, the Anhøj rules would signal a shift if $C < 2$ or $L > 6$; best box would signal if $C < 3$ or $L > 7$; and cut box would signal if $C < 3$ or $L > 7$ except when $C = 3$ and $L = 7$.

Results

We calculated the limits for the Anhøj, best box, and cut box rules together with their corresponding positive test proportions and log-likelihood ratios for $N = 10-100$ and $SD = 0-3$ (in 0.2 SD increments). We stored the log of likelihood ratios to preserve numerical precision. To get the actual likelihood values back, use `exp(log-likelihood)`. The limits are presented in Table 1. The full results set is available in the supplementary R data set ‘`cr_bounds.rds`’. The R code to produce the results and the figures in this article are available in the supplementary materials ‘`crossrunbox.R`’ and ‘`figs.R`’ respectively.

Figure 1 illustrates the effect of the best box and cut box procedures on the specificity of the runs analysis. As expected, the variability in specificity with varying N is markedly reduced and kept above and closer to the specified target – more with cut box than with best box.

Figure 3 shows the probabilities of getting a signal as a function of N and SD . The upper left facet ($SD = 0$) contains the same data as Figure 1. As expected and shown previously in our simulations studies, the power of the runs analysis increases with increasing N and SD . The smoothing effect of best box and cut box appears to wear off as N and SD increases.

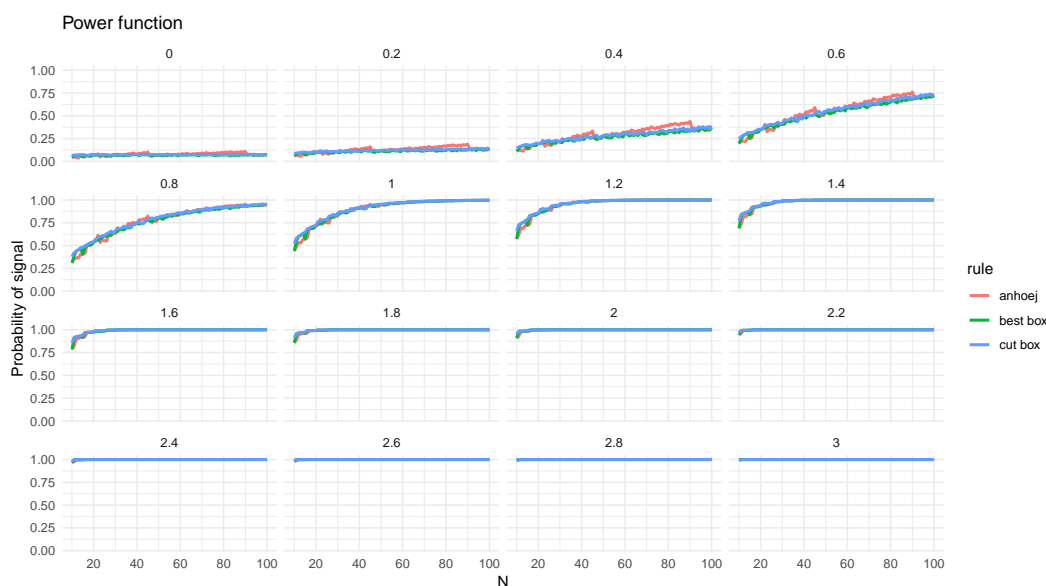


Figure 3: Power function of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units (SD) that is present in data.

Figures 4 and 5 compares the positive and negative likelihood ratios of the Anhøj rules to the box adjustments. The smoothing effect appear to be of practical value only for positive tests.

Discussion

In this study we demonstrated that it is feasible to reduce variation in run chart specificity from varying number of data points by using the best box and cut box adjustments of the Anhøj rules.

Confirmation that runs analysis is a valuable tool even with small N .

What is already known on this topic? (nothing)

What are the limitations of this study?

What are the strengths of this study?

What are the practical implications of our findings? (maybe none)

How can our findings be operationalised (method argument to `qic()`)

Conclusion

In this study we demonstrated that it is feasible to reduce variation in run chart specificity from varying number of data points by using the best box and cut box adjustments of the Anhøj rules.

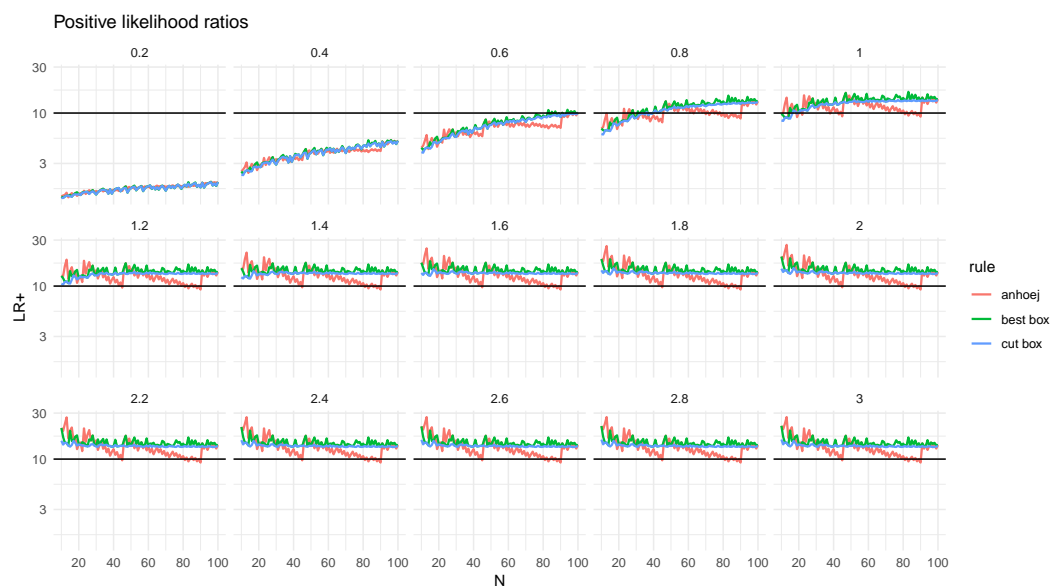


Figure 4: Positive likelihood ratio of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units that is present in data.

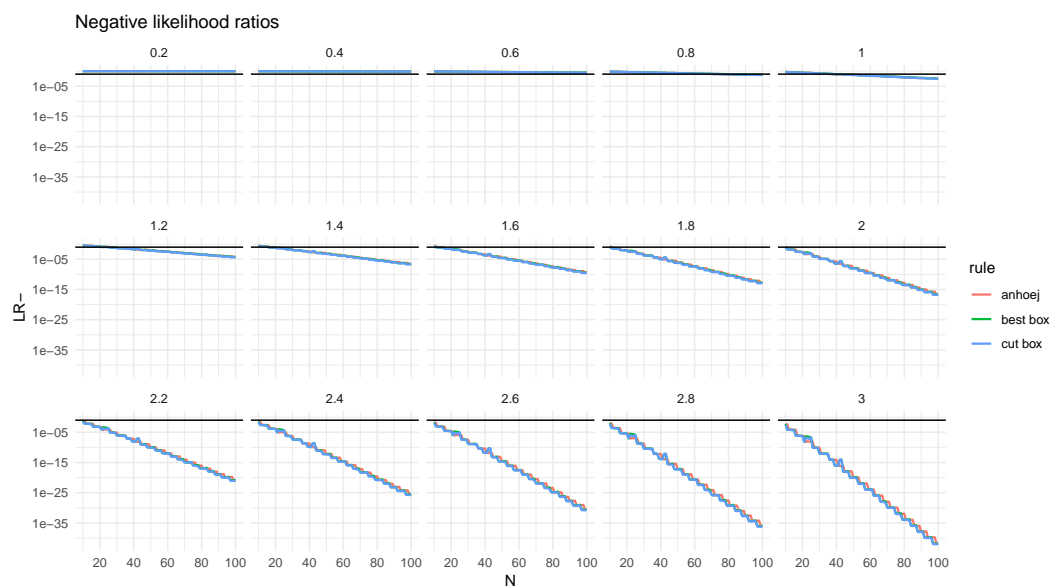


Figure 5: Negative likelihood ratio of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units (SD) that is present in data.

Bibliography

- J. Anhøj. Diagnostic value of run chart analysis: Using likelihood ratios to compare run chart rules on simulated data series. *PLoS ONE*, 2015. doi: 10.1371/journal.pone.0121349. URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121349>. [p1, 3]
- J. Anhøj and A. V. Olesen. Run charts revisited: A simulation study of run chart rules for detection of non-random variation in health care processes. *PLoS ONE*, 2014. doi: 10.1371/journal.pone.0113825. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0113825>. [p1]
- J. Anhøj and T. Wentzel-Larsen. Sense and sensibility: on the diagnostic value of control chart rules for detection of shifts in time series data. *BMC Medical Research Methodology*, 18(1):100, 2018. doi: 10.1186/s12874-018-0564-0. URL <https://doi.org/10.1186/s12874-018-0564-0>. [p1, 3]
- R. G. Carey. How do you know that your care is improving? part 1: Basic concepts in statistical thinking. *J Ambulatory Care Manage*, 25(1):80–87, 2002. [p1]
- Z. Chen. A note on the runs test. *Model Assisted Statistics and Applications*, 5:73–77, 2010. doi: 10.3233/MAS-2010-0142. [p1]
- J. J. Deeks and D. G. Altman. Diagnostic tests 4: likelihood ratios. *BMJ*, 329:168–169, 2004. doi: 10.1136/bmj.329.7458.168. URL <https://www.bmj.com/content/329/7458/168>. [p3]
- M. F. Schilling. The surprising predictability of long runs. *Mathematics Magazine*, 85:141–149, 2012. doi: 10.4169/math.mag.85.2.141. [p1]
- T. Wentzel-Larsen and J. Anhøj. *crossrun: Joint Distribution of Number of Crossings and Longest Run*, 2018. URL <https://CRAN.R-project.org/package=crossrun>. R package version 0.1.0. [p1]

Appendix: Bounds table

Table 1: Signal limits for the anhøj and best box rules and borders for the cut box rules. N = number of trials. L = upper limit for longest run, C = lower limit for number of crossings, Cbord and Lbord = cut box borders to keep.

N	Anhøj		Best box		Cut box	
	L	C	L	C	Cbord	Lbord
10	2	6	2	6	3	5
11	2	6	3	7	4	6
12	3	7	3	6		
13	3	7	3	6		
14	4	7	3	6		
15	4	7	4	7	6	6
16	4	7	5	8	6	7
17	5	7	5	7		
18	5	7	5	7	6	6
19	6	7	5	7	6	5
20	6	7	6	7		
21	6	7	7	8		
22	7	7	6	7	7	6
23	7	8	6	7	7	6
24	8	8	6	7	7	6
25	8	8	6	7		
26	8	8	9	9	10	7
27	9	8	9	8	10	7
28	9	8	9	8	11	7
29	10	8	10	8		
30	10	8	11	10	12	9
31	11	8	11	9	14	8
32	11	8	11	8		
33	11	8	11	8	12	7

Table 1: Signal limits for the anhöj and best box rules and borders for the (continued)

N	Anhøj		Best box		Cut box	
	L	C	L	C	Cbord	Lbord
34	12	8	11	8	13	7
35	12	8	12	8		
36	13	8	13	9	15	8
37	13	8	14	10		
38	14	8	13	8		
39	14	8	15	11		
40	14	8	15	9		
41	15	8	15	9	17	8
42	15	8	14	8		
43	16	8	14	8		
44	16	8	17	10		
45	17	8	17	9		
46	17	9	17	9	19	8
47	17	9	17	9	20	7
48	18	9	19	12	20	11
49	18	9	19	10	21	9
50	19	9	19	9		
51	19	9	19	9	21	8
52	20	9	19	9	21	7
53	20	9	21	11	23	9
54	21	9	21	10	23	8
55	21	9	21	9		
56	21	9	21	9	23	8
57	22	9	23	12	25	11
58	22	9	23	10	24	9
59	23	9	23	10	26	8
60	23	9	23	9		
61	24	9	23	9	24	8
62	24	9	25	11	27	9
63	25	9	25	10	27	9
64	25	9	26	11	27	10
65	25	9	26	10	27	9
66	26	9	27	12	29	10
67	26	9	27	10		
68	27	9	27	10	29	8
69	27	9	28	11	29	8
70	28	9	29	14	30	13
71	28	9	29	11	31	9
72	29	9	29	10	30	9
73	29	9	30	11	31	10
74	29	9	30	10		
75	30	9	31	12	32	9
76	30	9	31	11	34	8
77	31	9	31	10	33	9
78	31	9	32	11	33	8
79	32	9	33	13	37	11
80	32	9	33	11	35	9
81	33	9	33	10		
82	33	9	34	11	36	10
83	34	9	33	10	36	7
84	34	9	35	11		
85	34	9	35	11	38	8
86	35	9	35	10	36	9

Table 1: Signal limits for the anhøj and best box rules and borders for the *(continued)*

N	Anhøj		Best box		Cut box	
	L	C	L	C	Cbord	Lbord
87	35	9	35	10	38	8
88	36	9	37	12	38	10
89	36	9	37	11	39	9
90	37	9	38	12		
91	37	10	37	10	39	9
92	38	10	39	13	41	12
93	38	10	39	11	40	10
94	39	10	39	11	42	8
95	39	10	39	10		
96	39	10	39	10	41	8
97	40	10	41	12	42	9
98	40	10	41	11	44	9
99	41	10	42	12	43	10
100	41	10	41	10	42	9

Jacob Anhøj
 Rigshospitalet, University of Copenhagen
 Denmark
jacob@anhoej.net

Tore Wentzel-Larsen
 Centre for Child and Adolescent Mental Health, Eastern and Southern Norway & Centre for Violence and
 Traumatic Stress Studies, Oslo, Norway
 Norway
tore.wentzellarsen@gmail.com