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 (Anhøj and Wentzel-Larsen, 2018).

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, which are the default rules used for run and control chart analysis with the **qicharts2** package (Anhøj, 2019). 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 to the widely used Western Electric control chart rules at detecting minor to moderate *persistent* shifts (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.

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}, which is 1024 for N = 11. 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).

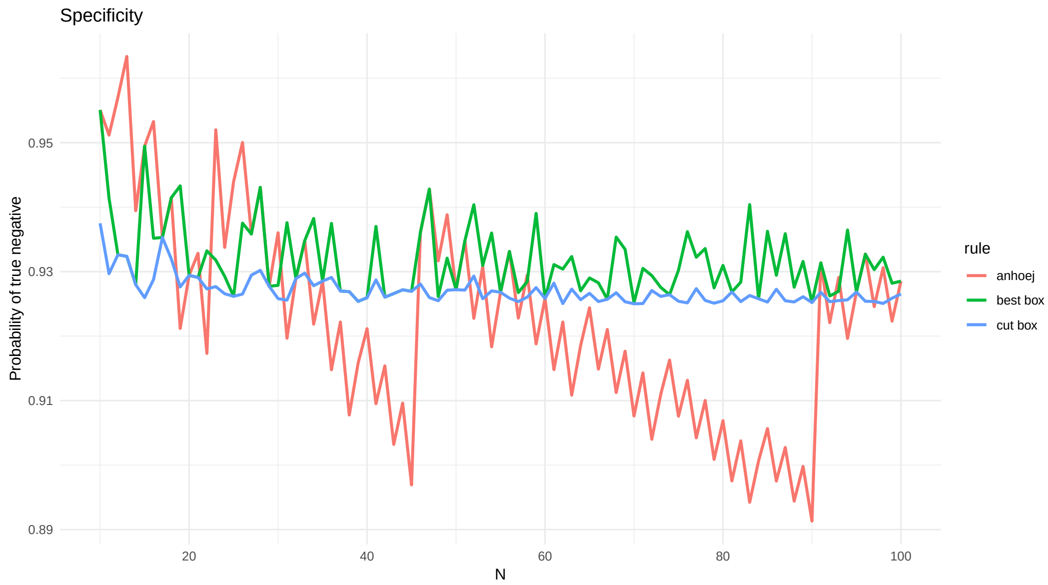

Figure 1: Specificity of the anhøj, best box, and cut box rules. N = number of data points in run chart.

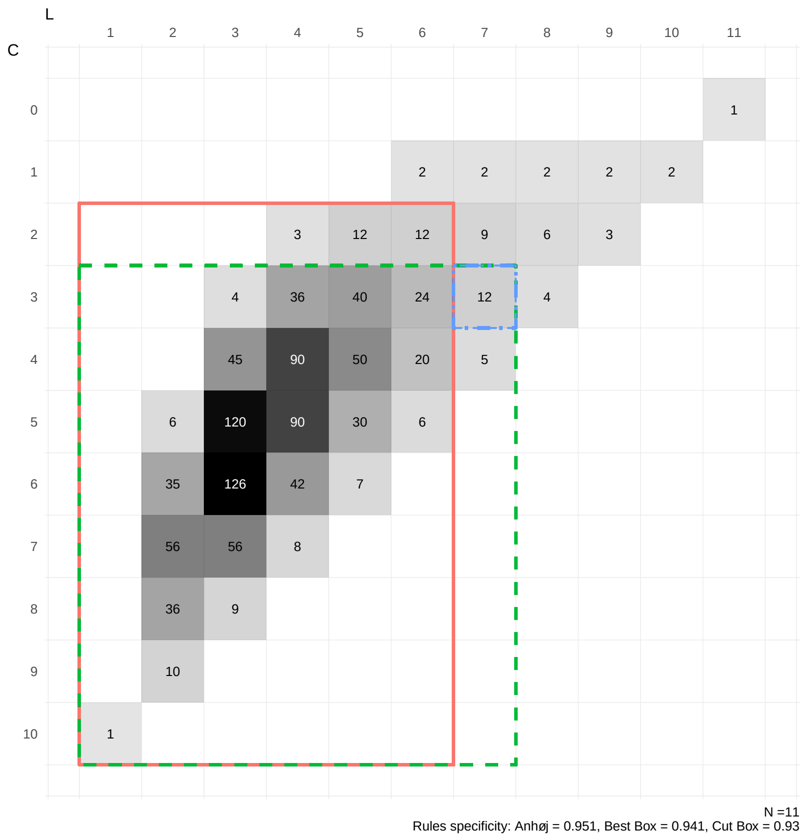

Figure 2: Borders of the anhøj, best box, and cut box rules for N = 11 data points. C = number of crossings, L = longest run. The numbers in the cells are times representation of of the joint probabilities of longest run and number of crossings. Anhøj = red solid, best box = green dashed, cut box = blue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

$$\text{Specificity} = \text{P(no signal} \mid \text{no shift)} = \text{P(true negative)} = 1 - \text{P(false positive)}$$
$$\text{Sensitivity} = \text{P(signal} \mid \text{shift)} = \text{P(true positive)} = 1 - \text{P(false negative)}$$

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 (LR+) greater than 10 is considered strong evidence that the condition is present. A negative likelihood ratio (LR-) 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 predetermined shift (the target shift), among boxes with specificity $\geq$ a pre-determined value (the target specificity). 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 one or 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, and also 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 and specifities are presented in Table 1. The R code to reproduce the full results set and the figures from this article is provided in the supplementary file 'crossrunbox.R'.

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.

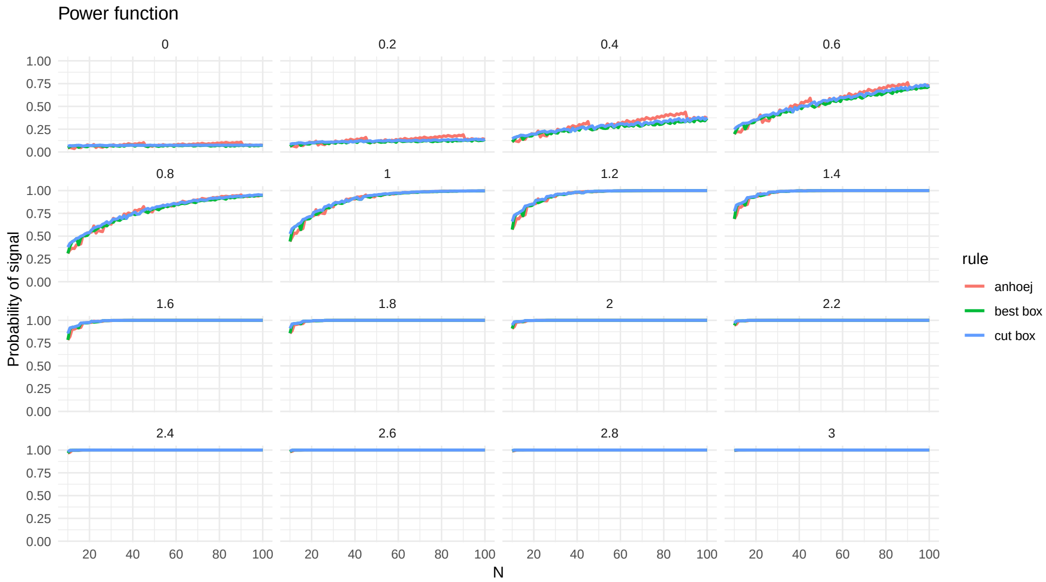

Figure 3: Power function of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units (SD) that is present in data.

Figures 4 and 5 compare 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 and conclusion

This study provided exact values for the diagnostic properties of the Anhøj rules for run charts with 10-100 data points including shifts up to 3 standard deviation units.

To our knowledge, and with the exception of our own **crossrun** package, the properties of the joint distribution of number of crossings and longest runs in random data series have not been studied before.

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

Most importantly, figures 4 and 5 confirm what we expected from years of practical experience using runs analysis, that the Anhøj rules is a useful and robust method for detection of persistent shifts only slightly larger than 1 standard deviation units and with as little as 10-12 data points. This can be seen by the fact that LR+ > 10 for SD > 1 and N $\geq$ 10. Although, the best box and cut box procedures will not change this, the box adjustments may potentially improve the practical value of runs analysis by reducing sudden shifts in sensitivity and specificity when the number of available data points changes. However, this remains to be confirmed in practice.

The study has two important limitations. First, the calculations of box probabilities require that

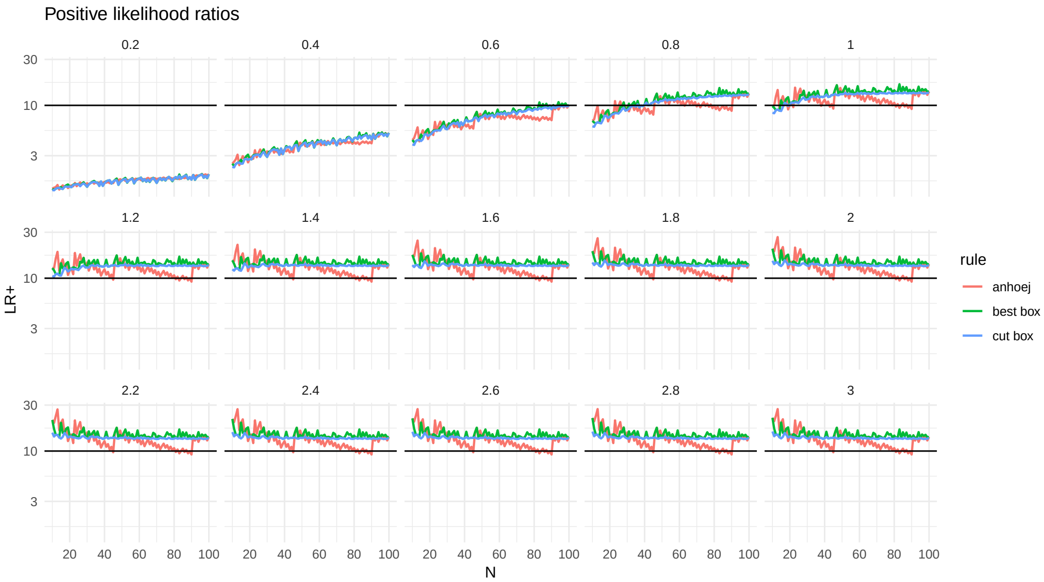

Figure 4: Positive likelihood ratio of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units that is present in data.

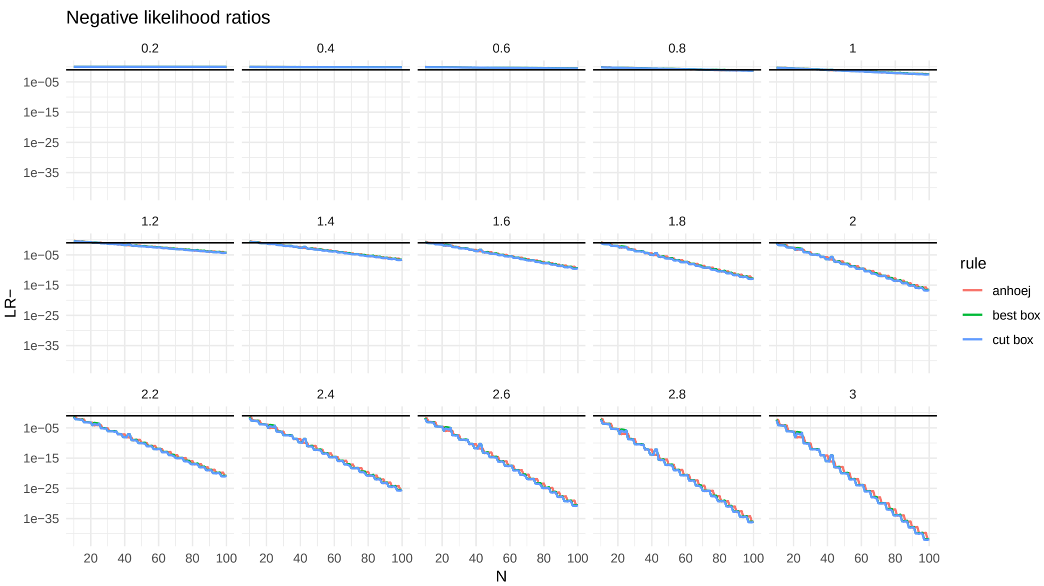

Figure 5: Negative likelihood ratio of Anhøj, best box, and cut box rules. N = number of data points in run chart. Numbers above each facet represent the size of the shift in standard deviation units (SD) that is present in data.

the process centre is fixed and known in advance, for example, the median from historical data. In practice the centre line is often determined from the actual data in the run chart, in which case the calculations of box probabilities do not apply. Preliminary studies suggest that this is mostly relevant for short data series. We plan to include a function in a future update of **crossrun** to calculate the box probabilities with empirical centre lines.

Second, the procedures have so far only been checked for up to 100 data points. Because of the iterative procedures and use of high precision numbers using functions from the **Rmpfr** package (Maechler, 2019) to calculate the joint distributions for varying N, the computations are time consuming. On a laptop with an Intel Core i5 processor and 8 GB RAM, it takes about one hour to complete `crossrunbox.R` for N = 10-100 and SD = 0-3, and the objects created consume over 6 GB of memory. However, we have no reason to believe that the procedures are not valid for N > 100, but the application of the box procedures for larger N may be impractical at the moment.

Also, one should be aware that the value of the box procedures rely on the choice of target specificity and target shift values. Other target values may give very different diagnostic properties. By supplying the R code, we encourage users to adapt our findings to their own needs.

Regarding the practical application of the box adjustment of the Anhøj rules, we are in the process of testing a `method` argument for the `qic()` function in the **qicharts2** package allowing the user to choose between `"anhoej"`, `"bestbox"`, and `"cutbox"` methods to identify non-random variation in run and control charts with up to 100 data points. This will allow us and others to quickly gain practical experience with the box adjustments on real life data.

In conclusion, this study provided exact values for the diagnostic properties of the Anhøj rules for run charts with 10-100 data points with shifts up to 3 standard deviation units, and demonstrated that it is feasible to reduce the variability in run chart specificity from varying number of data points by using the best box and cut box adjustments of the Anhøj rules.

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. *qicharts2: Quality Improvement Charts*, 2019. URL `https://CRAN.R-project.org/package=qicharts2`. R package version 0.6.0. [p1]

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. Maechler. *Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable*, 2019. URL `https://CRAN.R-project.org/package=Rmpfr`. R package version 0.7-2. [p6]

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]

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

	Anhøj		Best box		Cut box		Specificity		
N	L	C	L	C	Cbord	Lbord	Anhøj	Best box	Cut box
10	2	6	2	6	3	5	0.9551	0.9551	0.9375
11	2	6	3	7	4	6	0.9512	0.9414	0.9297
12	3	7	3	6			0.9570	0.9326	0.9326
13	3	7	3	6			0.9634	0.9324	0.9324
14	4	7	3	6			0.9395	0.9280	0.9280
15	4	7	4	7	6	6	0.9495	0.9495	0.9260
16	4	7	5	8	6	7	0.9533	0.9352	0.9288
17	5	7	5	7			0.9353	0.9353	0.9353
18	5	7	5	7	6	6	0.9415	0.9415	0.9320
19	6	7	5	7	6	5	0.9212	0.9433	0.9276
20	6	7	6	7			0.9294	0.9294	0.9294
21	6	7	7	8			0.9328	0.9291	0.9291
22	7	7	6	7	7	6	0.9173	0.9332	0.9273
23	7	8	6	7	7	6	0.9520	0.9318	0.9277
24	8	8	6	7	7	6	0.9338	0.9293	0.9266
25	8	8	6	7			0.9439	0.9262	0.9262
26	8	8	9	9	10	7	0.9500	0.9375	0.9265
27	9	8	9	8	10	7	0.9358	0.9358	0.9295
28	9	8	9	8	11	7	0.9431	0.9431	0.9302
29	10	8	10	8			0.9277	0.9277	0.9277
30	10	8	11	10	12	9	0.9360	0.9279	0.9258
31	11	8	11	9	14	8	0.9197	0.9376	0.9256
32	11	8	11	8			0.9289	0.9289	0.9289
33	11	8	11	8	12	7	0.9348	0.9348	0.9298
34	12	8	11	8	13	7	0.9218	0.9382	0.9278
35	12	8	12	8			0.9285	0.9285	0.9285
36	13	8	13	9	15	8	0.9148	0.9375	0.9291
37	13	8	14	10			0.9222	0.9270	0.9270
38	14	8	13	8			0.9078	0.9269	0.9269
39	14	8	15	11			0.9158	0.9254	0.9254
40	14	8	15	9			0.9212	0.9260	0.9260
41	15	8	15	9	17	8	0.9095	0.9370	0.9287
42	15	8	14	8			0.9154	0.9260	0.9260
43	16	8	14	8			0.9032	0.9266	0.9266
44	16	8	17	10			0.9096	0.9272	0.9272
45	17	8	17	9			0.8969	0.9270	0.9270
46	17	9	17	9	19	8	0.9361	0.9361	0.9281
47	17	9	17	9	20	7	0.9428	0.9428	0.9260
48	18	9	19	12	20	11	0.9317	0.9261	0.9255
49	18	9	19	10	21	9	0.9388	0.9321	0.9271
50	19	9	19	9			0.9272	0.9272	0.9272
51	19	9	19	9	21	8	0.9348	0.9348	0.9271
52	20	9	19	9	21	7	0.9228	0.9404	0.9293
53	20	9	21	11	23	9	0.9308	0.9310	0.9258
54	21	9	21	10	23	8	0.9183	0.9360	0.9270
55	21	9	21	9			0.9268	0.9268	0.9268
56	21	9	21	9	23	8	0.9331	0.9331	0.9259
57	22	9	23	12	25	11	0.9228	0.9268	0.9254
58	22	9	23	10	24	9	0.9295	0.9285	0.9260
59	23	9	23	10	26	8	0.9188	0.9390	0.9275
60	23	9	23	9			0.9258	0.9258	0.9258
61	24	9	23	9	24	8	0.9148	0.9311	0.9282

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

	Anhøj		Best box		Cut box		Specificity		
N	L	C	L	C	Cbord	Lbord	Anhøj	Best box	Cut box
62	24	9	25	11	27	9	0.9222	0.9304	0.9250
63	25	9	25	10	27	9	0.9108	0.9323	0.9273
64	25	9	26	11	27	10	0.9185	0.9270	0.9256
65	25	9	26	10	27	9	0.9244	0.9290	0.9266
66	26	9	27	12	29	10	0.9149	0.9283	0.9254
67	26	9	27	10			0.9210	0.9257	0.9257
68	27	9	27	10	29	8	0.9112	0.9354	0.9267
69	27	9	28	11	29	8	0.9177	0.9335	0.9253
70	28	9	29	14	30	13	0.9076	0.9252	0.9250
71	28	9	29	11	31	9	0.9143	0.9305	0.9251
72	29	9	29	10	30	9	0.9040	0.9294	0.9271
73	29	9	30	11	31	10	0.9109	0.9276	0.9262
74	29	9	30	10			0.9163	0.9264	0.9264
75	30	9	31	12	32	9	0.9076	0.9302	0.9254
76	30	9	31	11	34	8	0.9132	0.9362	0.9252
77	31	9	31	10	33	9	0.9042	0.9322	0.9274
78	31	9	32	11	33	8	0.9100	0.9336	0.9255
79	32	9	33	13	37	11	0.9009	0.9275	0.9251
80	32	9	33	11	35	9	0.9069	0.9310	0.9255
81	33	9	33	10			0.8975	0.9269	0.9269
82	33	9	34	11	36	10	0.9038	0.9284	0.9254
83	34	9	33	10	36	7	0.8942	0.9404	0.9263
84	34	9	35	11			0.9006	0.9258	0.9258
85	34	9	35	11	38	8	0.9057	0.9363	0.9253
86	35	9	35	10	36	9	0.8975	0.9294	0.9273
87	35	9	35	10	38	8	0.9027	0.9359	0.9255
88	36	9	37	12	38	10	0.8944	0.9276	0.9253
89	36	9	37	11	39	9	0.8998	0.9316	0.9261
90	37	9	38	12			0.8913	0.9252	0.9252
91	37	10	37	10	39	9	0.9314	0.9314	0.9268
92	38	10	39	13	41	12	0.9221	0.9262	0.9254
93	38	10	39	11	40	10	0.9291	0.9270	0.9255
94	39	10	39	11	42	8	0.9196	0.9365	0.9256
95	39	10	39	10			0.9268	0.9268	0.9268
96	39	10	39	10	41	8	0.9327	0.9327	0.9254
97	40	10	41	12	42	9	0.9246	0.9303	0.9254
98	40	10	41	11	44	9	0.9306	0.9322	0.9251
99	41	10	42	12	43	10	0.9223	0.9282	0.9259
100	41	10	41	10	42	9	0.9285	0.9285	0.9265

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