Description of the rules for declaring no special cause variation made from building domains, sequence length n=10 to 80:

10

cmin cmax l

1 2 2 5

2 3 9 7

11

cmin cmax l

1 2 2 5

2 3 3 6

3 4 10 7

12

cmin cmax l

1 2 2 5

2 3 4 6

3 5 11 7

13

cmin cmax l

1 2 3 5

2 4 4 6

3 5 12 8

14

cmin cmax l

1 2 3 5

2 4 4 6

3 5 13 9

15

cmin cmax l

1 2 4 5

2 5 5 7

3 6 14 9

16

cmin cmax l

1 3 4 5

2 5 5 6

3 6 6 8

4 7 15 9

17

cmin cmax l

1 3 3 5

2 4 6 6

3 7 16 10

18

cmin cmax l

1 3 4 5

2 5 6 6

3 7 17 11

19

cmin cmax l

1 3 5 5

2 6 6 6

3 7 18 12

20

cmin cmax l

1 3 6 5

2 7 7 8

3 8 19 12

21

cmin cmax l

1 3 7 6

2 8 8 7

3 9 20 12

22

cmin cmax l

1 3 7 6

2 8 8 7

3 9 21 13

23

cmin cmax l

1 4 6 5

2 7 8 6

3 9 22 14

24

cmin cmax l

1 4 7 5

2 8 8 6

3 9 23 15

25

cmin cmax l

1 4 9 6

2 10 10 11

3 11 24 14

26

cmin cmax l

1 4 8 6

2 9 10 7

3 11 25 15

27

cmin cmax l

1 4 9 6

2 10 10 7

3 11 26 16

28

cmin cmax l

1 5 8 5

2 9 10 6

3 11 27 17

29

cmin cmax l

1 4 10 6

2 11 11 7

3 12 12 9

4 13 28 16

30

cmin cmax l

1 4 10 6

2 11 11 7

3 12 12 8

4 13 29 17

31

cmin cmax l

1 5 10 6

2 11 12 7

3 13 30 18

32

cmin cmax l

1 5 11 6

2 12 12 7

3 13 31 19

33

cmin cmax l

1 5 12 6

2 13 13 8

3 14 32 19

34

cmin cmax l

1 5 12 6

2 13 13 7

3 14 14 9

4 15 33 19

35

cmin cmax l

1 5 12 6

2 13 13 7

3 14 14 8

4 15 34 20

36

cmin cmax l

1 5 12 6

2 13 14 7

3 15 35 21

37

cmin cmax l

1 6 13 6

2 14 14 7

3 15 36 22

38

cmin cmax l

1 6 14 6

2 15 15 8

3 16 37 22

39

cmin cmax l

1 6 14 6

2 15 16 8

3 17 38 22

40

cmin cmax l

1 6 14 6

2 15 16 7

3 17 39 23

41

cmin cmax l

1 6 15 6

2 16 16 7

3 17 40 24

42

cmin cmax l

1 6 16 6

2 17 17 10

3 18 41 24

43

cmin cmax l

1 7 16 6

2 17 17 8

3 18 18 9

4 19 42 24

44

cmin cmax l

1 7 14 6

2 15 18 7

3 19 43 25

45

cmin cmax l

1 7 16 6

2 17 18 7

3 19 44 26

46

cmin cmax l

1 7 17 6

2 18 18 7

3 19 45 27

47

cmin cmax l

1 7 18 6

2 19 19 9

3 20 46 27

48

cmin cmax l

1 7 12 6

2 13 20 7

3 21 47 27

49

cmin cmax l

1 8 17 6

2 18 20 7

3 21 48 28

50

cmin cmax l

1 8 18 6

2 19 20 7

3 21 49 29

51

cmin cmax l

1 8 20 6

2 21 21 11

3 22 50 29

52

cmin cmax l

1 8 20 6

2 21 21 8

3 22 22 10

4 23 51 29

53

cmin cmax l

1 8 17 6

2 18 22 7

3 23 52 30

54

cmin cmax l

1 8 20 6

2 21 22 7

3 23 53 31

55

cmin cmax l

1 9 21 6

2 22 22 7

3 23 54 32

56

cmin cmax l

1 9 22 6

2 23 23 9

3 24 55 32

57

cmin cmax l

1 8 23 7

2 24 24 8

3 25 56 32

58

cmin cmax l

1 9 21 6

2 22 24 7

3 25 57 33

59

cmin cmax l

1 9 22 6

2 23 24 7

3 25 58 34

60

cmin cmax l

1 9 24 6

2 25 25 12

3 26 59 34

61

cmin cmax l

1 8 25 7

2 26 26 8

3 27 60 34

62

cmin cmax l

1 10 20 6

2 21 26 7

3 27 61 35

63

cmin cmax l

1 10 23 6

2 24 26 7

3 27 62 36

64

cmin cmax l

1 10 25 6

2 26 26 7

3 27 63 37

65

cmin cmax l

1 9 26 7

2 27 28 8

3 29 64 36

66

cmin cmax l

1 9 27 7

2 28 28 8

3 29 65 37

67

cmin cmax l

1 11 24 6

2 25 28 7

3 29 66 38

68

cmin cmax l

1 11 26 6

2 27 28 7

3 29 67 39

69

cmin cmax l

1 9 28 7

2 29 29 8

3 30 30 9

4 31 68 38

70

cmin cmax l

1 9 29 7

2 30 30 8

3 31 69 39

71

cmin cmax l

1 11 24 6

2 25 30 7

3 31 70 40

72

cmin cmax l

1 10 30 7

2 31 31 9

3 32 71 40

73

cmin cmax l

1 10 30 7

2 31 31 8

3 32 32 9

4 33 72 40

74

cmin cmax l

1 10 30 7

2 31 32 8

3 33 73 41

75

cmin cmax l

1 10 31 7

2 32 32 8

3 33 74 42

76

cmin cmax l

1 10 32 7

2 33 33 10

3 34 75 42

77

cmin cmax l

1 10 32 7

2 33 34 9

3 35 76 42

78

cmin cmax l

1 11 32 7

2 33 33 8

3 34 34 9

4 35 77 43

79

cmin cmax l

1 11 33 7

2 34 34 8

3 35 78 44

80

cmin cmax l

1 11 34 7

2 35 35 12

3 36 79 44

For instance, for sequence length n=70, if the number of crossings is between 9 and 29, maximum run length is 7, if the number of crossings is 30 maximum run length is 8, and if there are 31 or more crossings maximum run length is 39:

70

cmin cmax l

1 9 29 7

2 30 30 8

3 31 69 39

Similarly, for sequence length n=30, if the number of crossings is between 4 and 10, maximum run length is 7, if the number of crossings is 11 maximum run length is 7, , if the number of crossings is 12 maximum run length is 8, and if there are 13 or more crossings maximum run length is 17:

30

cmin cmax l

1 4 10 6

2 11 11 7

3 12 12 8

4 13 29 17

The following diagram shows specificities and sensitivities (shift 0.8) for the various rules investigated, with red for the Anhøj rules, green bestbox, blue cutbox and orange “domain build” rules. The domain build rules are not as good as the cutbox rules for up to about n=14, but mostly better for longer sequences. It is to be noted that this is for shift 0.8, the target shift used for generating the rules



Sequence lengths above 80 have not been included now since the computations for the domain build rules are quite time consuming, for n=80 about 2 and a half hour. The domain build rules have a less smooth appearance by sequence length than the cutbox rules, but their sensitivity in many cases even exceed the Anhøj rules (red) even for some sequence lengths where these rules have a pronounced dip in specificity. Sensitivities for other shifts, and likelihood ratios, have so far not been computed for the domain build rules, but these computations are not substantially more complicated than for the Anhøj, bestboc and cutbox rules.