

Health Protection Agency Water Microbiology EQA

Guidance Notes for Laboratory Self-Assessment

EQUAL Scheme, Newcastle

*This leaflet has been prepared at the request of some participants who have sought guidance on self-assessment, using their results from distributions of the EQUAL Scheme. The notes provided here are suggestions and are **not** intended to describe compulsory procedures. The fundamental of self-assessment is that it comes from within the laboratory and is driven by the aims and requirements of that particular facility.*

Introduction

Each distribution has three samples (in the form of LENTICULE discs) which contain some, but not necessarily all, of the indicator organisms which participants are asked to test for and enumerate. Occasionally a sterile sample is included. The samples are prepared under strict quality assurance procedures. The development checks ensure that there is as little variation in numbers of organisms between samples as possible. But this variation is bound to be at least random (**Ref.1**). Therefore, for each individual water sample and for each parameter the “correct” or “true” count is never known.

Use of scheme reports

The Scheme organisers provide three types of documented feedback as well as the opportunity for correspondence. These are the **Interim Reports**, the **Distribution Reports** and regular **Performance Assessment Reports**. Approximately every 18 months there is a User Group Meeting. Participants in the Scheme should consider having documented procedures for recording that they have read and assessed the information in these reports, any actions derived and any feedback to their staff and/or Scheme Organisers, giving details of that response..

1. Immediately after the closing date of a distribution there is an **Interim Report**, which gives details of the intended counts for each of the parameters for each of the three samples analysed in the distribution, and confirming whether the Scheme Organisers have received results from the participant. The participant can record whether they found any false positive results (which are errors) and whether they detected the organisms listed in the statement. Only an approximate indication of the order of magnitude of the positive counts will be apparent at this stage. Gross transcription or calculation errors may be detected. Apparent false negatives may require scrutiny, although chance zeros can occur.
2. When all the data from participating laboratories returning results has been aggregated a **Distribution Report** is published which contains detailed descriptive statistics. Participants should plan to scrutinise their results, comparing their counts with the overall picture - not to see if they did "better" than others but because the overall picture gives the best information on what was present in the LENTICULE discs. The organisers state *that ‘Results compiled from all participants should show a concentration of results around the median result. This overall median should be close to the ‘intended’ result, which is the average count achieved by the organisers during pre-distribution quality assurance procedures. The participants’*

median is usually the best estimate of the average numbers of organisms in the LENTICULE

discs sent to participants'. However, as is inevitable with living organisms, there can be departures from the usual pattern. One strain may produce wider variation in counts than another or may prove more challenging to culture. It must also be remembered that the populations of bacteria present will represent a range of metabolic and stressed states, although this variability will be less for the LENTICULE discs used in the Scheme compared to that which would be obtained from natural samples. Therefore, participants must ensure that they read the Comments in the report and take note of these when assessing their results.

3. At regular intervals long-term and short-term **Assessment Reports** are sent out. These are parameter specific and have been devised to take into account natural variation in numbers of organisms between LENTICULE discs. There has been considerable debate about whether a scoring system is justified and the reasons for choosing not to use z-scores are discussed in **Ref. 2**.

As stated in the Assessment reports "Low and high 'tail-end' counts are arbitrarily defined according to the average result from all laboratories. Poisson 1% tail values are now being used. Typically about 5% to 10% of results fall into tail-areas due to natural overdispersion or to laboratory inaccuracies. These low and high counts could be reported occasionally and correctly from any laboratory through chance. However, if they cluster in particular laboratories over time then inadequate methodology should be suspected."

The report will identify how many "low" or "high" results the participant had over the particular series of samples for that parameter. The number will be tabulated as probably, possibly or unlikely to be due to chance. The latter is strong evidence of poor performance. These statistical assessments aim to identify poor performance which is microbiologically unacceptable. The analysis scheme deliberately makes generous allowance for variation in numbers of organisms between LENTICULE discs. This variation is bound to be at least random (i.e. Poisson distribution) but additional overdispersion can be found in microbiology. There is no attempt to seek out every statistically significant difference between laboratories.

In many situations a participant with highlighted poor performance will already be aware that they may have a problem. Their procedures for monitoring data from the Distribution report will have alerted them. The Assessment report always states that *"It is recommended that self-assessment is used as the first evaluation and this can be done by using the descriptive statistics supplied in the report which is circulated as soon as possible after each distribution. Participants can observe whether their results are, on average, close to the overall median results. Exact parity is not expected but large average differences or sustained increases in variability should be investigated as they may indicate rectifiable problems. In addition, major problems should be detected by these formal assessments made by the organisers."*

What is good self-assessment?

The Assessment report makes it clear that a participant with appropriate performance assessment procedures will already have found helpful information for self-assessment in the first two documents - the Interim Report and the Distribution Report - and that these formal Assessment reports are an additional service from the Scheme, acting as a safety net.

Good self-assessment will be planned beforehand, be well-documented and will set out clear lines of responsibility for the assessment of performance and undertaking and verifying any remedial action that may be required. The procedures should be able to evolve in the light of experience and thus grow to be appropriate to that laboratory's work.

If the procedures and log of events are to be reviewed by an external assessor or by a manager not closely involved in the day-to-day running of the laboratory then it may be helpful to have a written introduction highlighting the issues relating to the diversity in microbial behaviour and natural variation that can be expected in counts from microbiological methods. This should clarify why the EQA assessments cannot be conducted along the lines which were developed for chemistry. It should also emphasise that this EQA participation is but one management tool in a larger Quality Assurance package and should always be interpreted as such.

Some participants find it helpful to construct monitoring charts in addition to those provided in the reports. Some useful examples of these were presented by Dr. Simon Cole of Wessex Water at the 2002 User Group Meeting and his presentation is included as part of the report of the meeting and available at the scheme website (www.hpaweqa.org.uk).

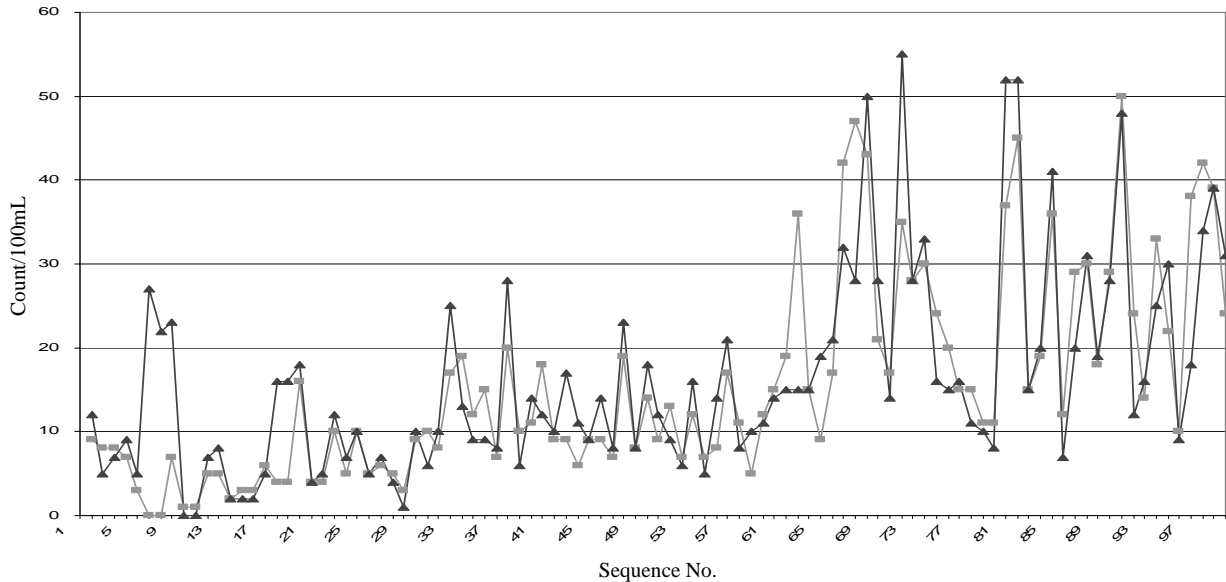
Recording results

Plots, parameter specific, of results over time can be considered. They are best made in relation to the overall median because this usually represents the best estimate of the average numbers of organisms in that distribution of LENTICULE discs. But beware of anomalous medians (look out for relevant comments in Distribution and Assessment reports) and allow for these, perhaps by highlighting them on the charts.

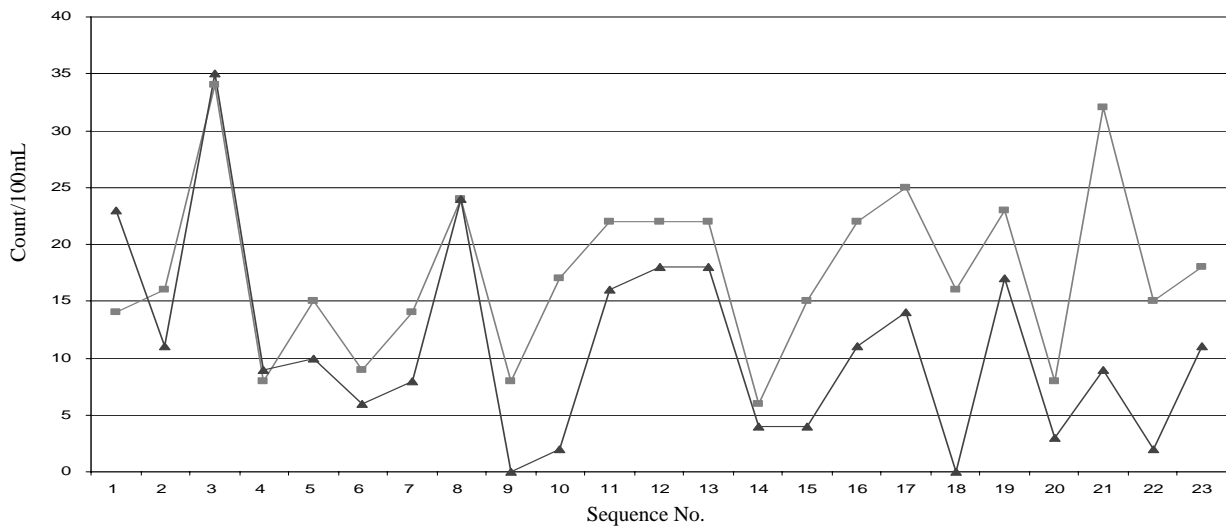
Three styles of plot have been suggested which are applied to a parameter and use cumulated results from samples which should have contained that parameter. Each gives visual prominence to a slightly different aspect of monitoring although all three use the same information.

- (i) A **line graph** with the time sequence of samples as the x-axis and the count as the y-axis. One symbol plots the participant's results and another the median result calculated from the results returned by all participants. A visual aid is to join up the points with lines although strictly speaking the lines, do not represent anything as there are no "results" being reported between the times of samples.

Line graph of laboratory and median counts for positive samples



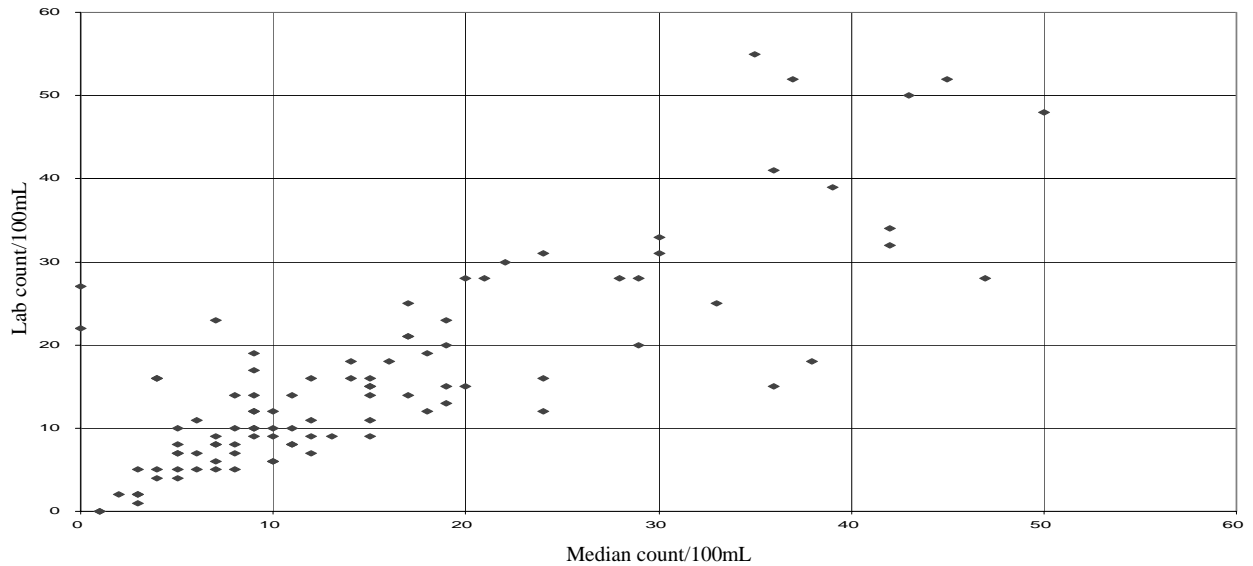
Line graph showing significant deviation from median



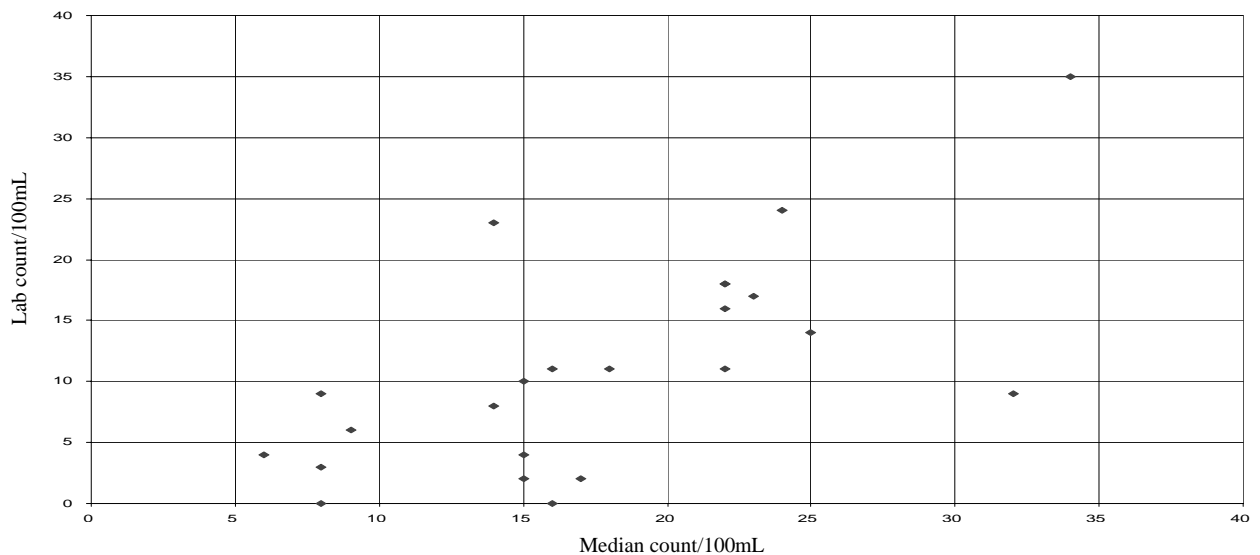
A simple assessment of performance would be to monitor whether there is a consistent trend of results on one side of the median. A satisfactory performance will be when the two lines criss-cross but are seldom widely separate and that average separation does not increase over time. A large variability of results around the distribution medians could indicate poor control of the analytical process. This type of chart shows all available information - time sequence, actual value of counts and difference between this laboratory and the median. It is, therefore, quite complex to interpret at a glance and can be supplemented by two other charts.

- (ii) An x/y **scatter plot** of the median count (x-axis) against the laboratory's results (y-axis). Satisfactory performance is when the scatter is around the diagonal line of equality with approximately similar numbers below and above. It will also be possible to spot whether the pattern changes for higher median values, although allowance must be made for the fact that the magnitude of the scatter will inevitably increase. Random scatter is proportional to the average count (with respect to Poisson distribution).

XY plot of laboratory vs. median count – positive samples



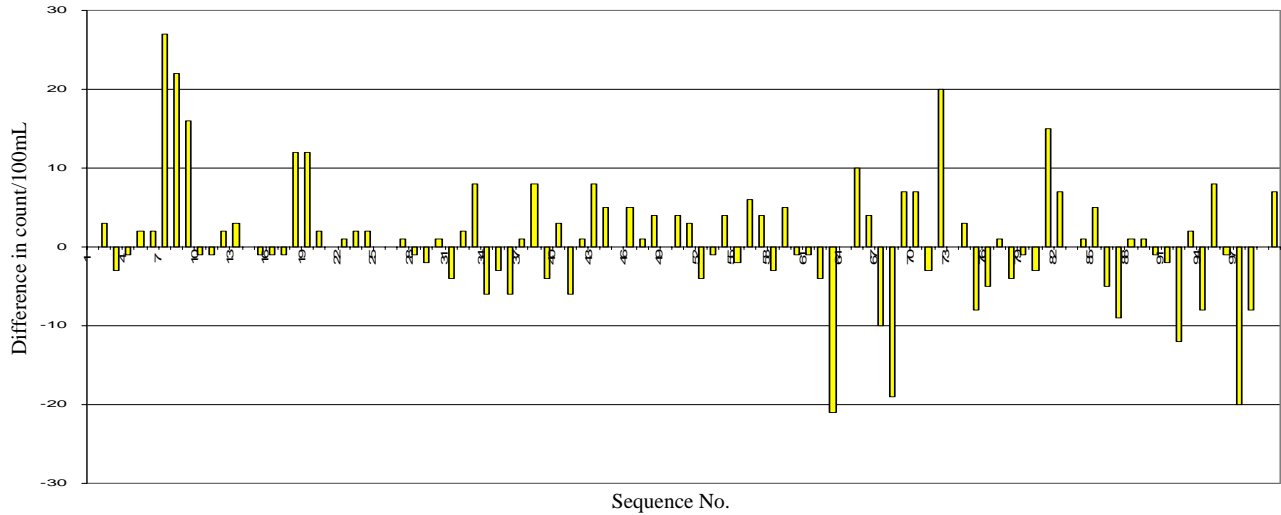
XY plot for parameter showing deviation from the median



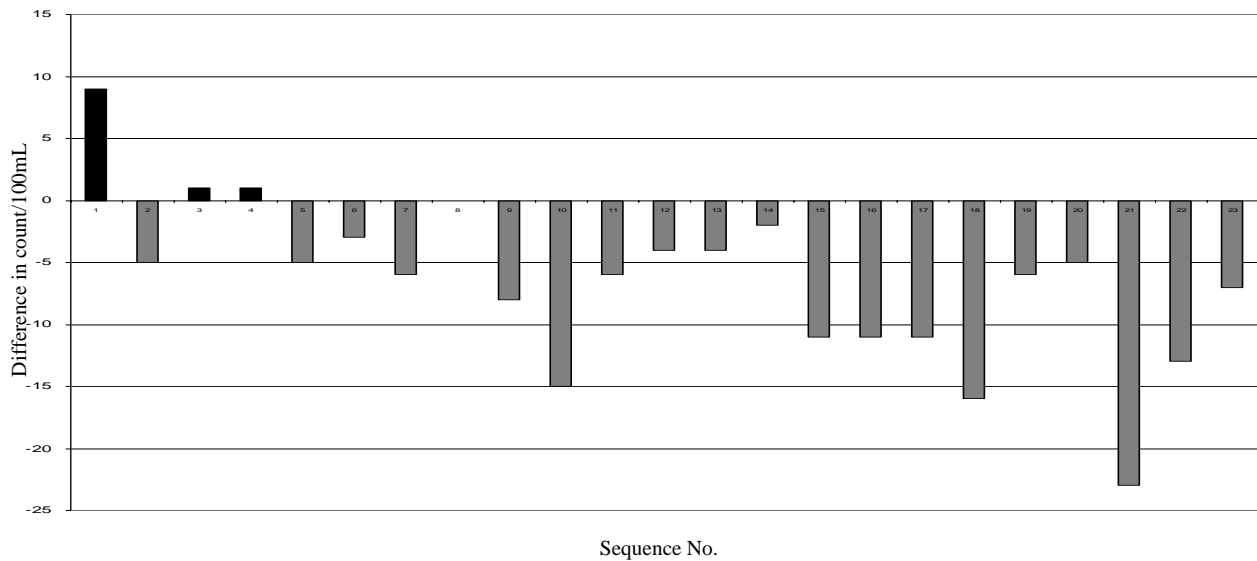
This plot will not indicate time sequence and so not provide an early warning that performance has changed. A participant may, for example, notice that they perform adequately with moderate or higher counts but tend to record a deficit with low counts.

- (iii) A **bar chart of differences**. This plots sequentially the absolute value of the difference between the laboratory's result and the median.

Bar chart of difference counts laboratory vs. median



Difference bar chart for the same parameter with negative bias



If high average counts are involved you might consider using a different scale - square root or logarithm - but with drinking waters actual counts may be the best. Remember that the choice of scale makes a huge difference to the visual impact of the differences, regardless of the true facts. This will give a quick visual warning if a laboratory is consistently finding more or less than the average numbers, i.e. whether there is a consistent or marked trend of results on one side of the median. Remember, however, that small "biases" may not be microbiologically significant and not cost effective to investigate beyond routine checks.

Interpreting self-assessment

Undertaking self-assessments, interpreting results and implementing any required actions will be part of the documented procedures under the laboratory's Quality Manual. Certain observed single results or sequences of counts when evaluated against established performance criteria will trigger investigations and reports (either formal or in logs). If a problem is suspected it is important to consider the whole system before jumping to conclusions. Each assumption should be challenged and checked. Participants are encouraged to share their findings with the Scheme Organisers, who can pass on the lessons learned. EQA schemes are not an arena for laboratories to compete against each other, but an important forum for all participants to assure and improve their performance. In this respect feedback from participants on the outcome of any internal investigations would be an important part of a mutual scheme.

Participants may wish to incorporate into their self-assessment the arbitrary "low" and "high" classifications used in the formal statistical assessments. The tables used by the Scheme are provided in the Appendix. **It is very important, however, that no individual result is assessed this way because of the chance variation in numbers of viable and/or stressed bacteria per sample.**

EQA schemes are an integral part of any laboratory's performance assessment. Results from such schemes should be continuously monitored for any trends (whether upward or downward), and for whether results show a microbiologically significant bias to be one side or the other of the median. Whatever performance criteria a laboratory establishes, they must be based upon a full understanding of the analytical process and the areas that can significantly impact on the final result, including the level of capability and performance expected by the customer (or regulator). Ascertainment of a laboratory's performance is dependent upon the laboratory having robust and justifiable monitoring of performance from both internal quality control and participation in external proficiency testing schemes.

References

1. Tillett, H.E and Lightfoot, N.F. Quality Control in Environmental Microbiology Compared with Chemistry: What is Homogeneous and What is random? *Wat. Sci. Tech.* 1995 Vol 31, No. 5-6, 471-477.
2. Tillett, H.E., Lightfoot, N.F., Eaton, S. and Place. B.M. External Quality of Microbial Counts from Water: To Score or Not to Score for Proficiency. *J.CIWEM*, 2000, **14**, 304-308

Appendix
TABLE OF ARBITRARY DEFINITIONS OF TAIL-END COUNTS
(Poisson 0.005 probability per tail: *i.e.* 1% two-tailed probability)

N.B. Interpretation of low and high depends on cumulated results. This table should NOT be used to assess SINGLE results.

Median count	"low" less than	"high" greater than
5	1*	12*
6	1	13
7	1	15
8	2	16
9	2	18
10	3	19
11	4	20
12	4	22
13	5	23
14	5	25
15	6	26
16	7	27
17	7	28
18	8	30
19	9	31
20	10	32
21	10	34
22	11	35
23	12	36
24	12	38
25	13	39
26	14	40
27	15	41
28	15	43
29	16	44
30	17	45
31	18	46
32	18	47
33	19	49
34	20	50

DO NOT USE TO ASSESS SINGLE RESULTS

35	21	51
36	22	52
37	22	54
38	23	55
39	24	56
40	25	57
41	25	58
42	26	60
43	27	61
44	28	62
45	29	63
46	30	64
47	30	66
48	31	67
49	32	68
50	33	69
51	34	70
52	34	71
53	35	73
54	36	74
55	37	75
56	38	76
57	39	77
58	39	79
59	40	80
60	41	81
61	42	82
62	43	83
63	44	84
64	44	86
65	45	87
66	46	88
67	47	89
68	48	90
69	49	91
70	49	92
71	50	94
72	51	95
73	52	96
74	53	97

DO NOT USE TO ASSESS SINGLE RESULTS

75	54	98
76	54	99
77	55	101
78	56	102
79	57	103
80	58	104
81	59	105
82	60	106
83	60	107
84	61	109
85	62	110
86	63	111
87	64	112
88	65	113
89	66	114
90	66	115
91	67	116
92	68	118
92	69	119
94	70	120
95	71	121
96	72	122
97	72	123
98	73	124
99	74	126
100	75	127

DO NOT USE TO ASSESS SINGLE RESULTS

* i.e. <1 (zero) is "low" and >12 is "high"

N.B. Interpretation of low and high depends on cumulated results. This table should NOT be used to assess SINGLE results.