Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

02-Nov-2015

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

STATISTICAL ISSUES RELATED TO OECD IN VITRO GENE MUTATION TESTS TEST GUIDELINES (TG476)

Series on Testing & Assessment
No. 224

This document is available in pdf format only.

JT03385593

ENV/JM/MONO(2015)33
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 224

STATISTICAL ISSUES RELATED TO OECD IN VITRO GENE MUTATION TESTS TEST GUIDELINES (TG476) (No. 224)

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

Several projects for updating the set of OECD Test Guidelines (TG) on genotoxicity were submitted by a group of four countries, i.e., Canada, France, the Netherlands and the US, and included in the Test Guideline workplan in April 2011 by the WNT. Among others, there was the TG 476 for the *In Vitro* Mammalian Cell Gene Mutation Tests, that actually was limited to the Hprt and xprt genes only. The tk gene based mutation assays were instead considered in the new TG 490.

To support the update of TG 476, statistical analyses were performed based on data collected from calls for data organised between September and December 2012. This document presents the Statistical analysis supporting the revision of the OECD TG 476, carried out by a Consultant to the Secretariat. The main goal was to determine the optimal number of cells to be scored during the execution of the test.

This statistical analysis report was endorsed by the Working Group of the National Coordinators of the Test Guidelines Programme (WNT) at its 27th meeting in April 2015. The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 3 July, 2015.

This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

STATISTICAL ISSUES RELATED TO OECD *IN VITRO* GENE MUTATION TESTS TEST GUIDELINES (TG 476)

David Lovell, St George's Medical School, University of London

Introduction

1. This report discusses statistical issues associated with Test Guideline (TG 476). The purpose of this work was to perform statistical analyses to support the determination of the optimal number of cells to be scored in Test Guideline 476 (*in vitro* gene mutation tests).

2. This report provides a summary of the data sets provided by laboratories based upon a call for data made in September 2012 from Nathalie Delrue of OECD (and Tim Singer of Health Canada) for negative historical control data with a further call made in December 2012.

3. The objective of the data collection process was to obtain a manageable set of data from over 12 and no more than 20 laboratories for each test with a geographical spread and representation of the users of the tests. The call allowed laboratories to restrict their returns to recent data from upto their last 20 experiments. Some laboratories provided their most recent data from, say, their last 20 experiments; others provided their full historic control database.

4. A number of laboratories responded with data from their negative control cell mutation assay data sets. Although individual culture data were requested, not all laboratories provided their data in this form. In practice, the reporting of this information varied from laboratory to laboratory.

5. Laboratories were asked to provide some limited information on the conditions that the tests were carried out on. The amount of extra information provided was somewhat limited. As some laboratories requested their data should remain anonymous the laboratories have been given codes to designate them

6. Test Guideline 476 covers in vitro mammalian cell gene mutation tests using the hprt gene with CHO-WBL and V79 cells. Data were obtained from 11 (A to J) laboratories (Table 1). Data were collected and checked by Robert Heflich (NCTR) who provided codes for the laboratories to anonymize them.

7. Following a discussion in July 2013, Heflich undertook to obtain further information by asking for further details from the laboratories on the experimental techniques and the source of their cells. This allowed an update of the laboratory information with this further information provided in an email from him of 16th September 2013. One laboratory (F) specifically noted that cells were 'cleansed' with HAT, others referred to HAT treatment. This extra information has been complied into a Table (Table 2).

8. Heflich also received some further sets of data from two laboratories. Firstly, Lab B provided some additional CHO-Hprt data produced by another principal investigator in the same laboratory (here identified as from lab B') but running the studies using a different protocol. Secondly, a new laboratory, Lab K, provided a small data set on hprt mutations in TK6 cells. Lab J also provided an update of their data file which identified the technical replicates.

Summary of background mutant frequency data provided

9. Data were obtained from about 11 laboratories. The following sets of results were initially received:

6 sets of CHO-Hprt data	Laboratories A, B, C, D, E, F
4 sets of V79-Hprt data	Laboratories G, H, I, J
1 set of L5178Y-Hprt data	Laboratory F
2 sets of AS52-gpt data	Laboratories A, B

10. A further set of data were received by Heflich in September 2013

1 set of CHO-Hprt data	Laboratory B'
1 set of TK6 Hprt data	Laboratory K

Tables 1 and 2 gives details of the datasets and the number and conditions of the studies carried out.

The quantity of data provided varied between laboratories

11. Data were cut and pasted from the Excel work sheets and, in some cases, Word documents provided and entered into the package Minitab. In the case of data from one laboratory (Laboratory F) part of the dataset had to be transcribed because it could not be cut and pasted.(double entry was carried out to avoid errors.)

12. Some laboratories (e.g. Laboratories C and F) provided appreciable information on the conditions used in the tests. Other laboratories initially gave no other information other than the MF values but provided more details in response to Heflich's requests. These are summarized in Table 2.

13. The way the MF is derived depends in part on the specific design used by the particular laboratory. The 'basic' endpoint is the mutation frequency (MF x 10^{-6}). The MF values were presumed to have been calculated as the (number of mutants /number of cell plated/ plating efficiency (PE), although most laboratories did not explicitly state this, The data are expressed in terms of mutants per 10^6 clonable cells. Not all laboratories provided the numerators and denominators used to derive the MF. In only one case (Lab J) can the underlying data associated with the derivation of the MF be identified from the data sent and the MF value checked as being correct.

14. Some laboratories provided just the MF value for those particular study/experiments; others provide more of the underlying data that the MF value was based upon. In some cases appreciable supplementary data are provided, which ultimately feeds into the calculation of the MF. Laboratory C, for instance, gave the specific counts from each of 6 replicate plates and also provided both the uncorrected and the corrected (after adjustment by the PE) MF values while Laboratory J reported the summary data as the mean ±standard deviations for both MF and %PE plating efficiency (PE) values.

15. One laboratory (Lab D) provided 4 replicate values for each experiment. Some laboratories appear to have used two replicates (C, F, I), others (A, B, E, G, H, J) just a single replicate. Laboratory B and H provided a single measure based upon the mean values of two technical replicates. Laboratory B reported that their hprt (but not their gst values) MF values were based upon a mean of technical replicates. (In the extra data for Laboratory F each value is derived from two replicate cultures.) Laboratory F subsequently resent their data with replicates clearly) identified.

16. Replicates appear in all cases to be technical replicates derived from a single isolate of cells. The closeness of the replicates to the initial isolate probably varies from laboratory to laboratory.

17. In some cases information on solvents were provided. Laboratories B, C and I noted the different solvents used in their studies. Laboratory I provided information on studies where different solvent were used in the same study. Laboratory C provided a breakdown of the results for each solvent.

Laboratory I also provided positive control data. These data were not used in any analyses.

18. Data were often provided in a form where they were separated into whether S9 mix was used or not (+S9 or –S9). One laboratory (H) reported data based upon two different expression times (6 days and days 8 - 9 days). Initially information on different conditions were analyzed separately and then combined across S9 and other conditions because no significant differences were, in general, identified between the different conditions.

19. In some laboratories the results with and without S9 mix can clearly be identified as being from the same experiment. In others, this seems to be implicit in how the results are reported. In other cases, there does not appear to be any linkage between the sets of data. At present, no analyses have been carried out to try to make use of the linked data information across experiments.

Summary of results

20. In general, only the MF x 10^{-6} data were available. No data on cloning efficiency (CE) were provided (except for one laboratory (J) where data on plating efficiency (PE) was provided). No information on cellular suspension growth (SG) comparable with that for the tk assays was provided. Criteria such as those developed for the tk assays for acceptability were not available so it was not possible to say that the studies were acceptable or not based upon supplementary data.

21. There were no significant differences in the MF values between the +S9 and –S9 data sets or different lengths of expression time or different solvents based upon a series of one-way analyses of variance (anova on untransformed data).

22. From visual inspection of the histograms of MF values for each laboratory there was evidence of variability between laboratories for both the CHO and V79 cells. *(This is discussed in more detail later).*

23. Also, in contrast to the reported tk assays data, there were many cases where the distributions were not normally distributed (often positively skewed) and zero or low values occurred. In most cases the values of the mean minus 2SD were negative.

24. Despite the skewness of the data there are not appreciable differences between the mean and median values although, as would be expected, the medians are slightly lower given the positive skewness of most datasets.

25. Three (GPT, CHO, V79) sets of data (plus some other smaller sets from Laboratory F) are thus available. It is clear from an initial inspection of the results that the MFs for the hprt of the CHO are about 3 times lower than those of the MFs for the hprt of the V79. However, there is an overlap between laboratory means for the CHO (range 2.8 to 11.8) and the V79 (10.2 to 17.9). (This is discussed in more detail later).

26. Note that the Anderson-Darling (AD) test has low power to detect deviation from a normal distribution when n is small. The null hypothesis is that the data are from a normal distribution. An implication of this is that if n is large enough all datasets will be significantly different from a normal distribution as the statistical test will be capable of picking up even small deviations from a normal distribution.

27. The results from the MFs from the AS292 gpt laboratories are presented separately.

Table of means, SD and Anderson-Darling (AD) test

Lab	N	Mean	StDev	AD	P
A GPT	16	16.03	7.22	0.889	0.017
B GPT	506	22.905	13.571	5.807	<0.005
A CHO-K1-BH4	41	4.234	2.808	0.747	0.048
B CHO-K1-BH4	63	5.061	4.551	2.483	<0.005
C CHO-K1	340	2.815	3.044	15.134	<0.005
D CHO-K1	128	11.771	6.812	1.355	<0.005
E CHO-K1-BH4	537	6.743	4.646	7.471	<0.005
F CHO-K1-BH4	306	4.129	3.084	8.643	<0.005
G V79	135	17.907	9.442	3.807	<0.005
H V79	268	10.183	8.453	18.183	<0.005
I V79	528	17.739	11.569	6.828	<0.005
J V79	131	10.530	8.759	3.633	<0.005
F L5178Y	80	4.787	2.984	3.483	<0.005
F CHO	4	7.03	4.40	0.325	0.30

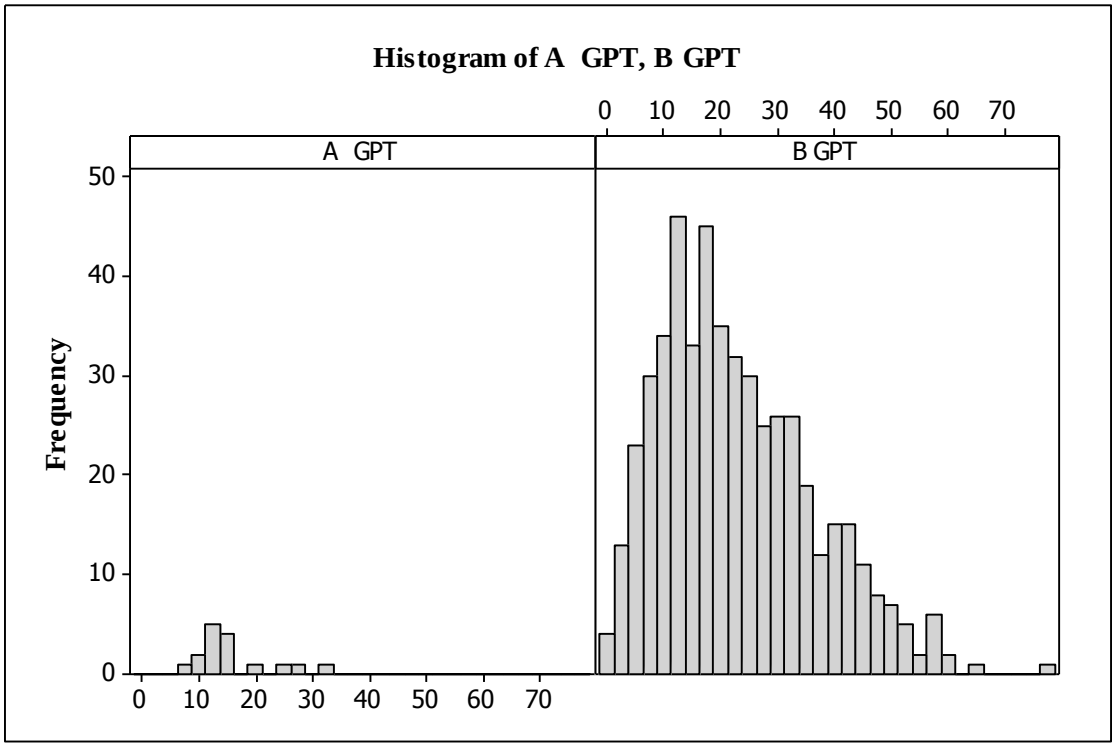
GPT data: 2 laboratories

28. Only two laboratories provided data so it is difficult to draw any general conclusions One laboratory (A) submitted small datasets, the other (B) was a larger set with some zeros and appreciable skewness.

29. Laboratory mean MFs were 16.0 x 10^{-6} and 22.9 x 10^{-6}

Fig 1 GPT

Lab	Total Count	N	Mean	SE Mean	StDev	Minimum	Median	Maximum	Zero
A GPT	16	16	16.03	1.80	7.22	8.00	13.75	33.10	0
B GPT	506	506	22.90	0.60	13.57	0.00	20.38	77.31	2



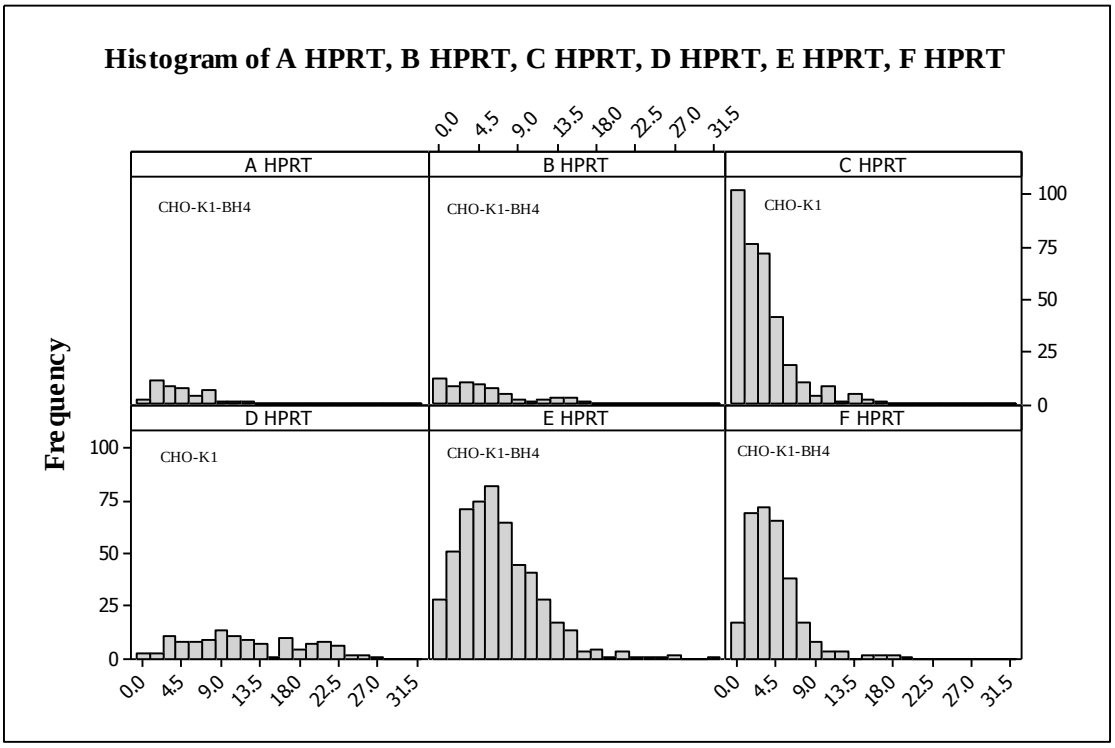
CHO HPRT data: 6 laboratories

30. Laboratory means ranged from 2.8 (C) to 11.8 (D).

31. The patterns of the distributions differed appreciably between laboratories some with appreciable skewness (Fig 2). Means ranged from 2.8 to 11.8. All laboratories had zero values, one (C) with 76! All the distributions were non-normal (P<0.005 on the Anderson-Darling test of fit except the small sample of A (P=0.048 on the AD).

Fig 2 CHO

Variable	Total Count	N	Mean	SE Mean	StDev	Minimum	Median	Maximum	Zero
A CHO-K1-BH4	41	41	4.234	0.439	2.808	0.000	3.300	12.000	2
B CHO-K1-BH4	63	63	5.061	0.573	4.551	0.000	3.800	16.900	5
C CHO-K1	340	340	2.815	0.165	3.044	0.000	2.045	16.430	76
D CHO-K1	128	125	11.771	0.609	6.812	0.000	10.596	27.536	3
E CHO-K1-BH4	537	537	6.743	0.200	4.646	0.000	6.000	32.200	15
F CHO-K1-BH4	306	302	4.129	0.177	3.084	0.000	3.550	19.100	4

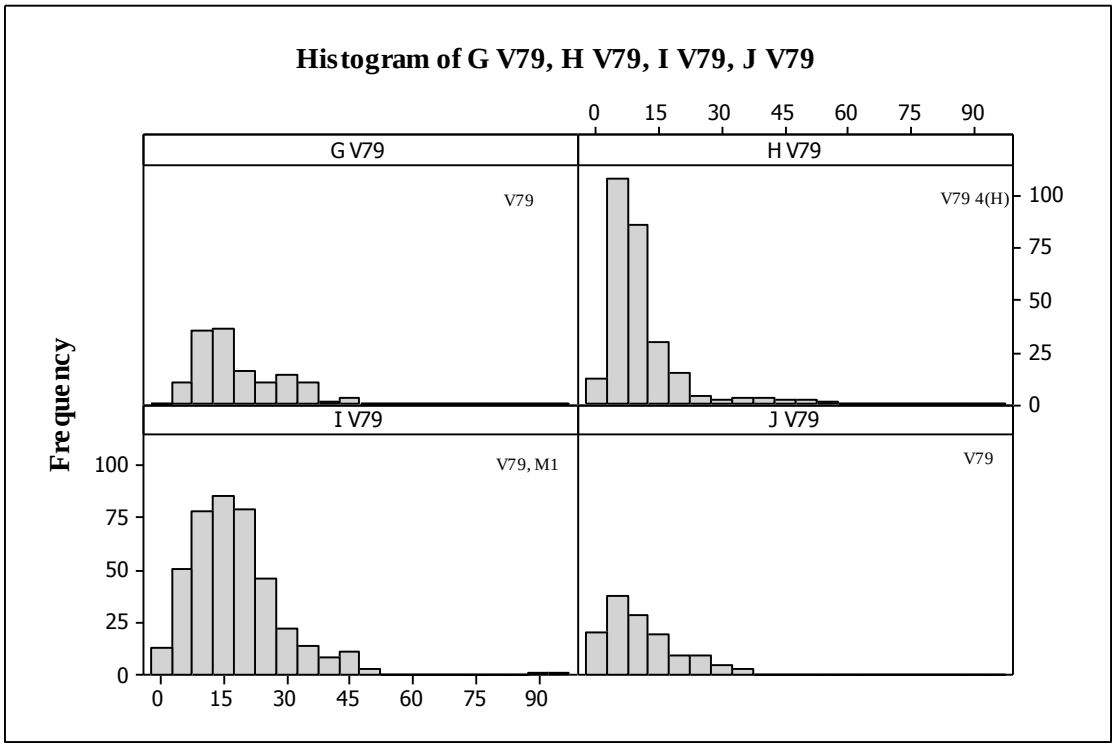


V79 HPRT data: 4 laboratories

32. Laboratory means were from 10.2 (H) to 17.9 (G). Laboratory J had 9 zeros. The other three laboratories had just one zero. The distributions of all 4 datasets looked approximately normal but with some positive skewness. (All were P<0. 005 on the Anderson-Darling test of goodness of fit to the normal distribution.)

Fig 3 V79

Variable	Total Count	N	Mean	SE Mean	StDev	Minimum	Median	Maximum	Zero
G V79	135	135	17.907	0.813	9.442	3.000	14.700	45.000	0
H V79	268	268	10.183	0.516	8.453	1.010	7.995	56.100	0
I V79	528	411	17.739	0.571	11.569	0.000	16.200	96.600	1
J V79	131	131	10.530	0.765	8.759	0.000	9.188	36.765	9

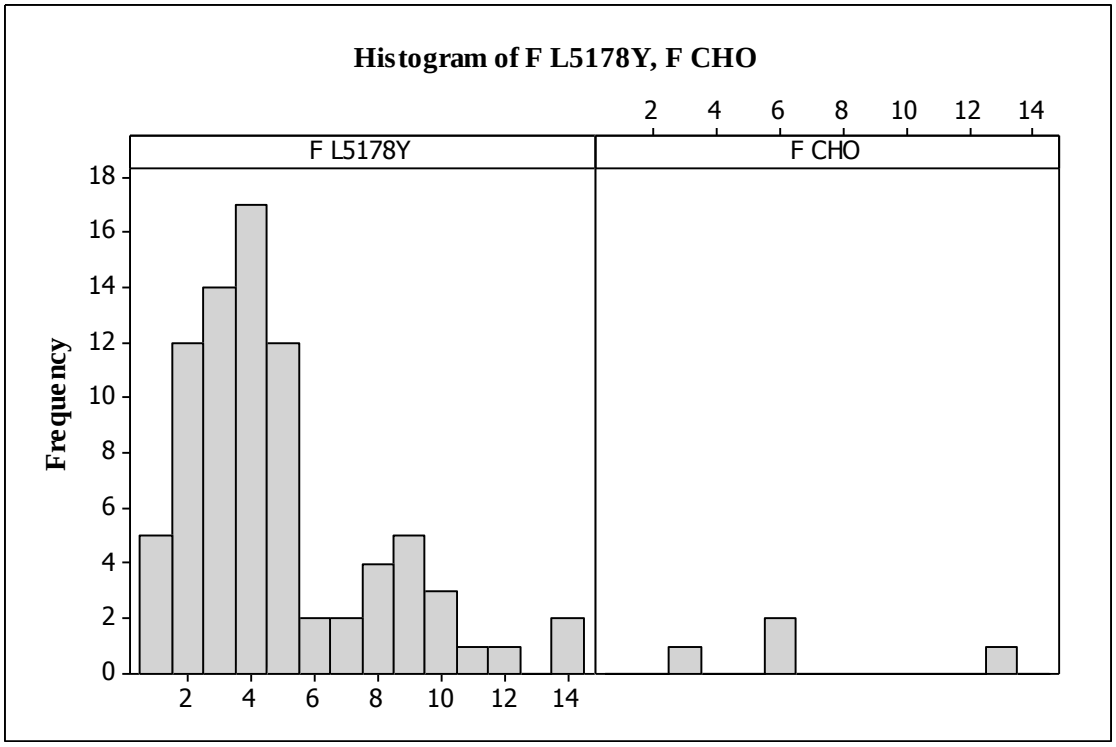


Other data

33. For completeness the extra data from Lab F is presented (Fig 4). Lab F L5178Y-Hprt data had a mean 7.0 with no zeros. The Lab F CHO sample consisted of just 4 data points.

Fig 4 Sundries: L5178Y F CHO

Lab	Total Count	N	Mean	SE Mean	StDev	Minimum	Median	Maximum	Zero
F L5178Y	80	80	4.787	0.334	2.984	0.870	3.800	14.170	0
F CHO	4	4	7.03	2.20	4.40	2.83	6.04	13.20	0



Comments on the experimental designs

34. The design of studies in each experiment is probably based upon an original single isolate of cells from which a number of aliquots are frozen down. About each month one of these aliquots is thawed for use and then the following month another aliquot is thawed to produce cells for use and so on. This can continue until a new isolate of cell needs to be frozen down. So there can be a complicated hierarchy of replicate values over time: within experiment, between experiments, between aliquots, between isolates etc.

Most data sets probably include technical replicates (although what this is a replicate of is not always clear).

35. The +S9 experiments were carried out over 'short treatment times' e.g. 3-5hrs. Most –S9 experiments were also carried out at these 'short treatments times' of 3-5hrs but in some cases a longer 24hr treatment time was also used.

.
Different solvents were used as the vehicle. No evidence was found using anovas that these vehicles affected the MF values.

Supplementary analyses

36. In most case there was no evidence of differences between +S9 or –S9 results within a laboratory (or using different treatment or expression times). Most anovas were non-significant. Data were subsequently pooled over conditions within laboratories.

1) Variability between experiments

37. Three laboratories (D, F, H and the extra data from J) provided data where it was relatively easy to test whether there was evidence of variability between experiments.

Lab D

38. In the case of Lab D the results were in the form of data from cultures with either +S9 or –S9 conditions for 16 separate experiments, thus allowing a test for differences between the experiments ($P<0.001$) and the S9 mix x Experiments interaction ($P=0.002$). The overall difference between +S9 and -S9 mix cultures was not significant indicating that the size and direction of the differences between the conditions depended on the particular experiments (Fig 5).

Untransformed data

```
Analysis of Variance for MF, using Adjusted SS for Tests

Source     DF   Seq SS   Adj SS  Adj MS      F      P
Expt       15  3154.22  3150.47  210.03  10.91  0.000
S9          1    29.70    29.82   29.82   1.55  0.216
Expt*S9    15   779.59   779.59   51.97   2.70  0.002
Error      93  1789.99  1789.99   19.25
Total     124  5753.50
```

Transformed data (log+1)

```
Analysis of Variance for Log MF, using Adjusted SS for Tests

Source     DF    Seq SS   Adj SS   Adj MS     F      P
Expt       15   5.86955  5.86456  0.39097  9.26  0.000
S9          1   0.07725  0.07543  0.07543  1.79  0.185
Expt*S9    15   1.46001  1.46001  0.09733  2.30  0.008
Error      93   3.92775  3.92775  0.04223
Total     124  11.33456
```

39. Based upon the analysis of variance an estimate of the pooled standard deviation can be obtained which provides an estimate of the coefficient of variation (CV%).

Lab D Mean = 11.77 Pooled SD = 4.387 CV% = 37.2%

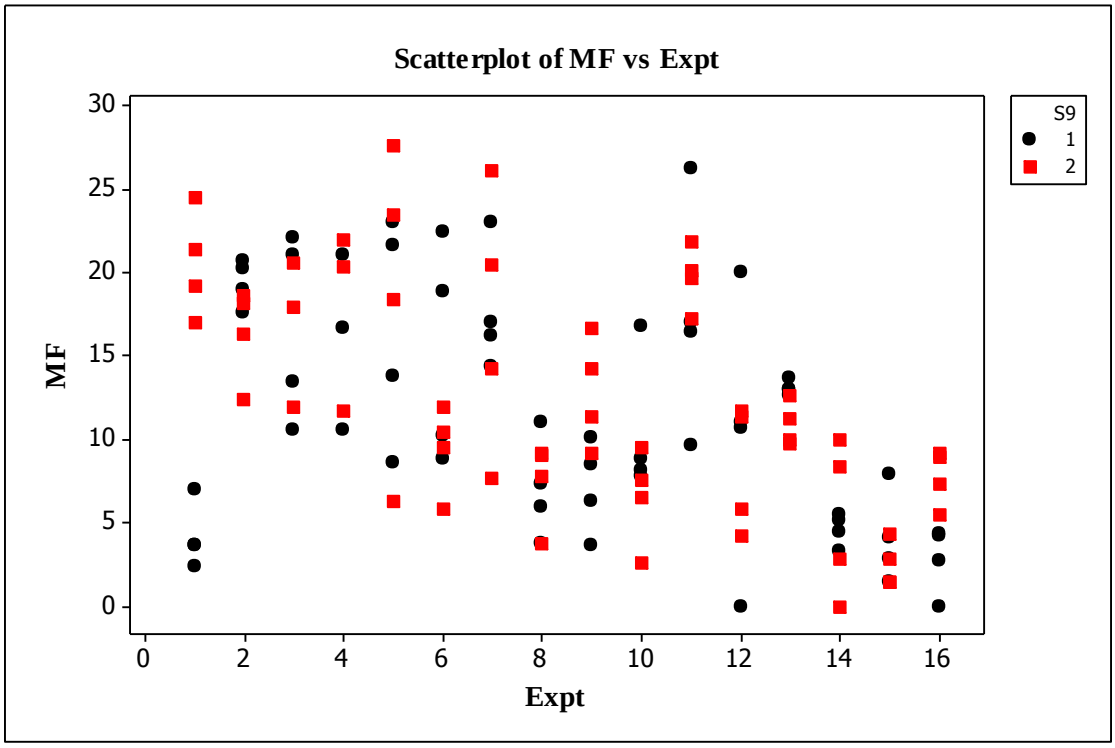


Fig 5 A scatterplot of the individual MF values for 4 replicates with either S9 added (Black circle:) or S9 not added (Red circle: 2) for 16 (independent) experiments

Lab F

40. In the case of Laboratory F, highly significant differences in MF values between experiments (P<0.001) were seen for both +S9 and –S9 conditions. (This analysis is an approximation because although the laboratory stated that the data represented 2 replicates per experiment there were 153 individual replicate values and it was, therefore, not clear exactly which pairings related to the 76 or 77 independent experiments.)

41. In the case of Laboratory F the following further statistics can be derived from the anova:

-S9: Mean = 4.055 Pooled SD = 1.873 CV = 46.2%
+S9 Mean = 4.216 Pooled SD = 2.319 CV = 55.0%

(Note that the values in the smaller Lab F data set are the means of 2 replicate cultures)

Lab H

42. The data set appeared to consist of 67 experiments carried out under 4 difference conditions. Significant differences were found between the experiments and between the different conditions (both P<0.001)

Analysis of Variance for Lab H MF, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Lab H expt	66	15920.71	15920.71	241.22	17.41	0.000
Lab H S9	3	413.61	413.61	137.87	9.95	0.000
Error	198	2743.48	2743.48	13.86		
Total	267	19077.81				

43. In the case of Lab H the following statistics can be derived from the anova.

All Mean = 10.18 Pooled SD = 3.722 CV = 36.6%

2) Relationship between mutation frequency MF and plating efficiency (PE)

44. One laboratory (J) provided PE data. The correlation between the PE and the MF was -0.306 (P=0.012) for the –S9 mix (Fig 6) and -0.067 (P= 0.60) for the +S9 mix (Fig 7).

The linear regression of Mutation Frequency on Plating efficiency for the –S9 mix studies was:
MF0% = 35.4 - 0.319 PE0%
And for the +S9 mix studies was:
MF5% = 17.4 - 0.096 PE5%

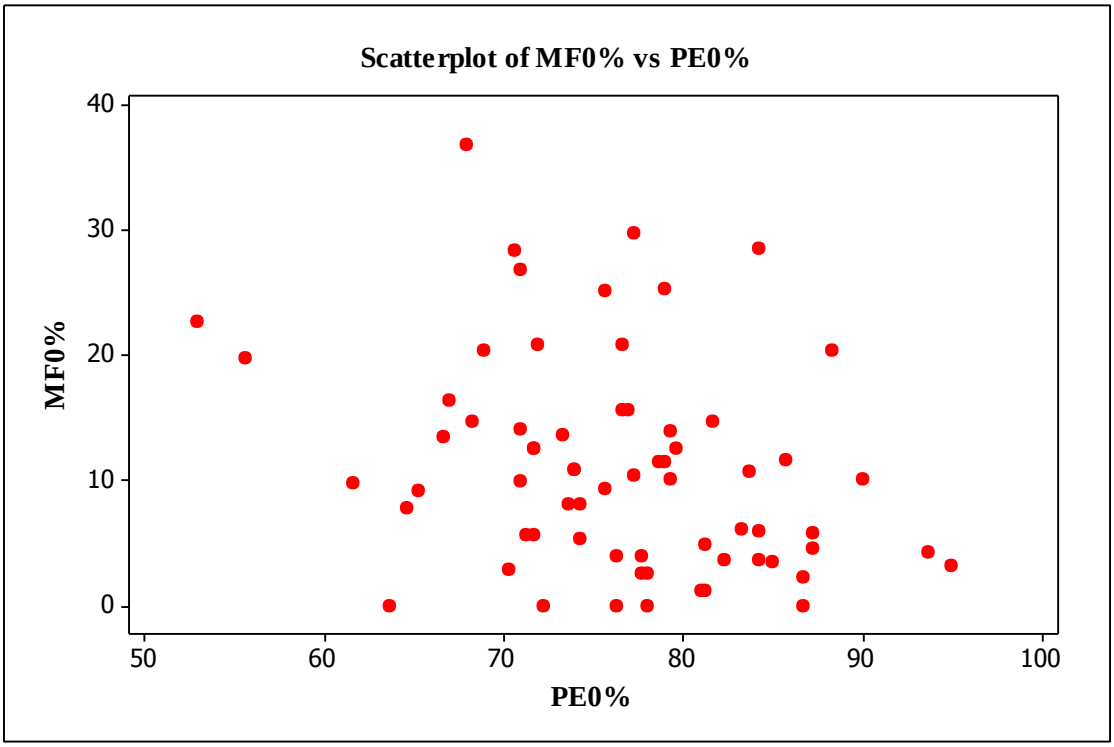


Fig 6 Scatterplot of Mutation Frequency against Plating Efficiency for the –S9 studies

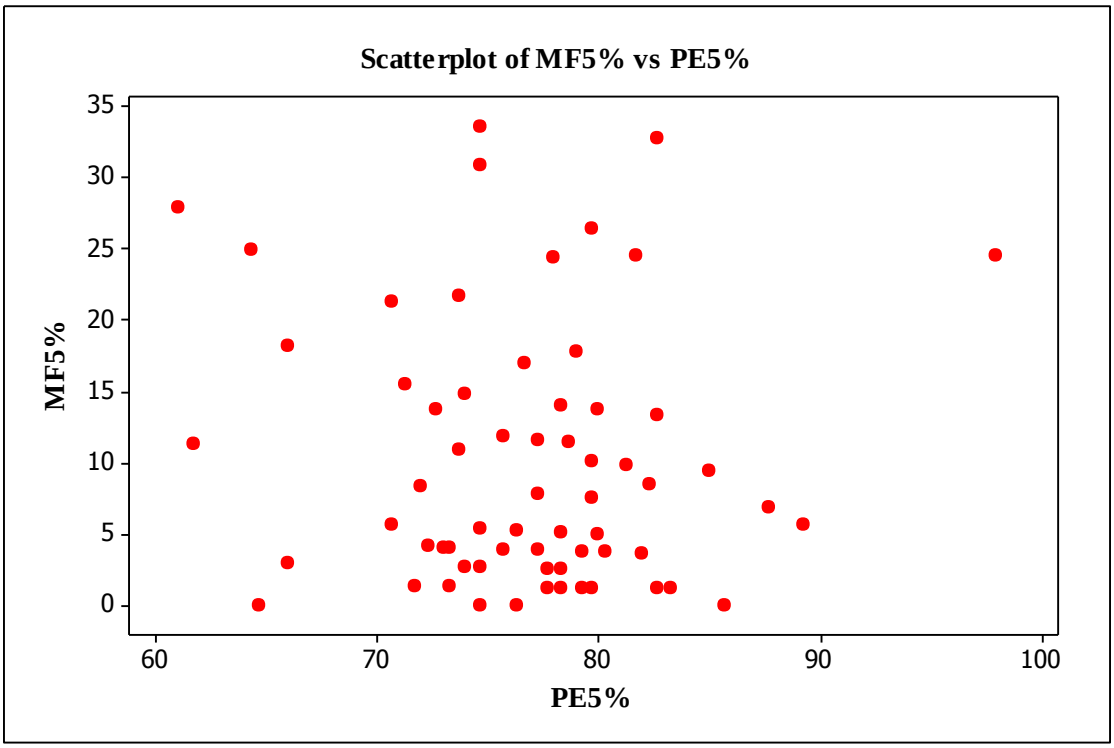


Fig 7. Scatterplot of Mutation Frequency against Plating Efficiency for the +S9 studies

Analyses on the extra data provided to Heflich in September 2013

45. There were some interesting aspects to these extra set of data which have been explored by further analyses.

Laboratory B'

46. (Note that compared with the original data that was sent Lab B reported a small amount of extra data: one extra culture -S9 #31 WATER 1.66 and one extra culture +S9 #32 WATER 1.27. These extra data were not added to the analyses done previously.)

47. The rest of the results were data submitted by another researcher (Lab B') in the organisation but using a different protocol. These consisted of:

8 studies without S9 (-S9) with 4 vehicles (EtOH, di-H2O, Saline, DMSO) n=22 with 2-4 replicates/study and 8 studies with S9 (+S9) with 4 vehicles (EtOH,di-H2O, Saline, DMSO) n=24 with 2-4 replicates/study and some called 'confirmatory studies'.
Some interesting results were provided with I charts (Fig 8 & 9) and histograms for both MF (Fig 10) and CE% (Fig 11).

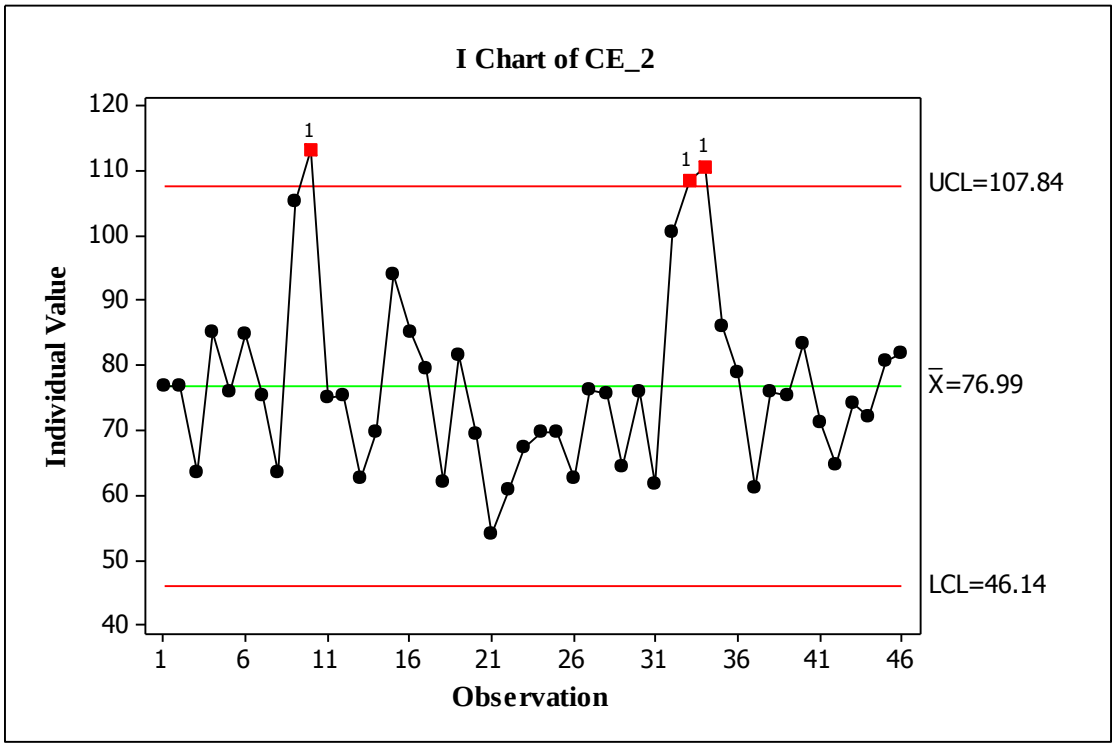


Fig 8 I Chart of Cloning Efficiency for Laboratory B'

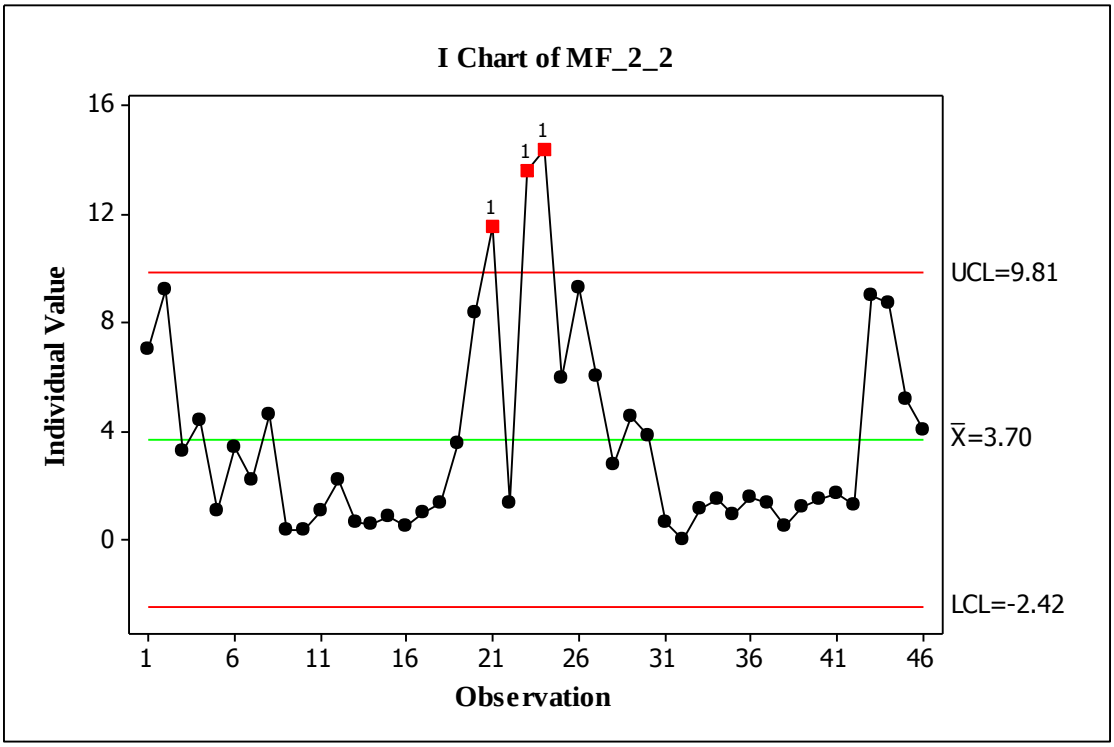


Fig 9 I Chart of Mutation Frequency for Laboratory B'

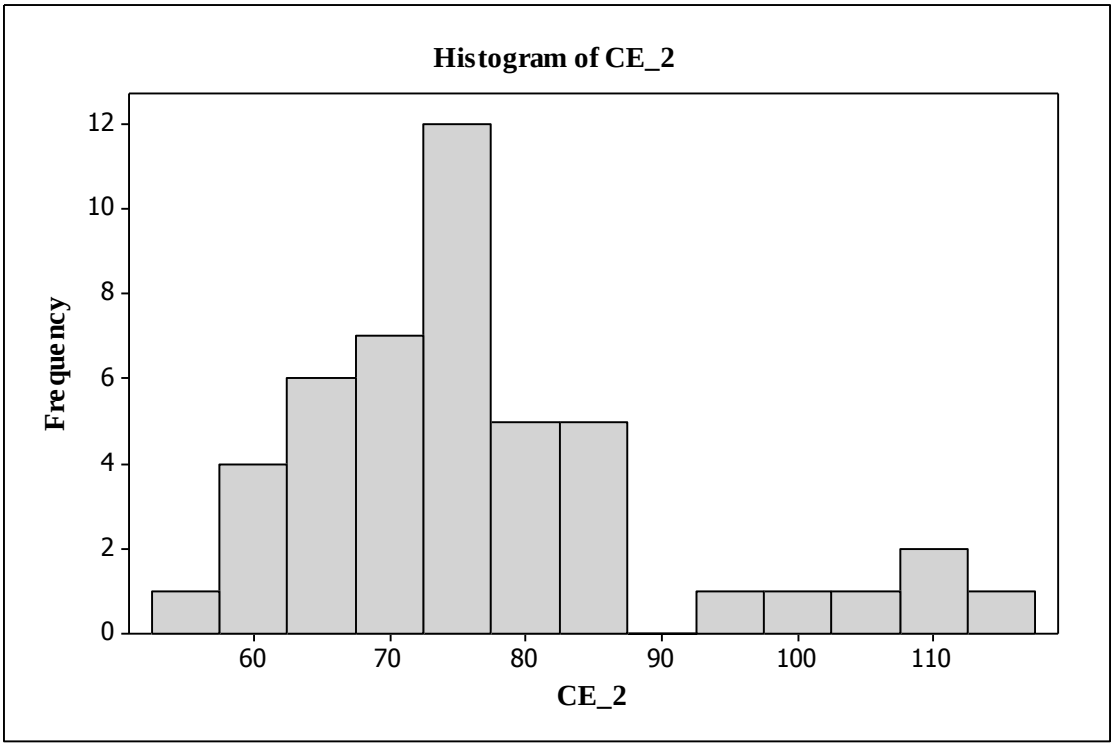


Fig 10 Distribution of Cloning Efficiency data for Laboratory B'

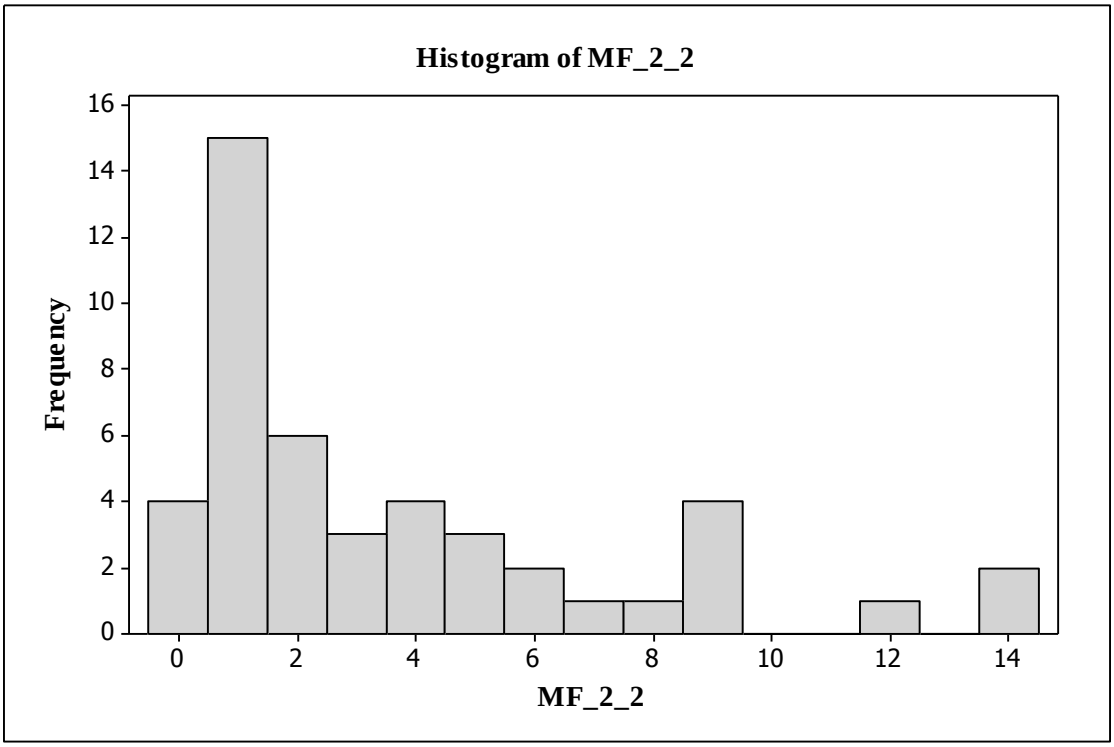


Fig 11 Distribution of Mutation Frequency data for Laboratory B'

Some CE values were >100% and are associated with very low MFs (Fig 12). The Pearson correlation of CE_2 and MF_2_2 = -0.363 P = 0.013.

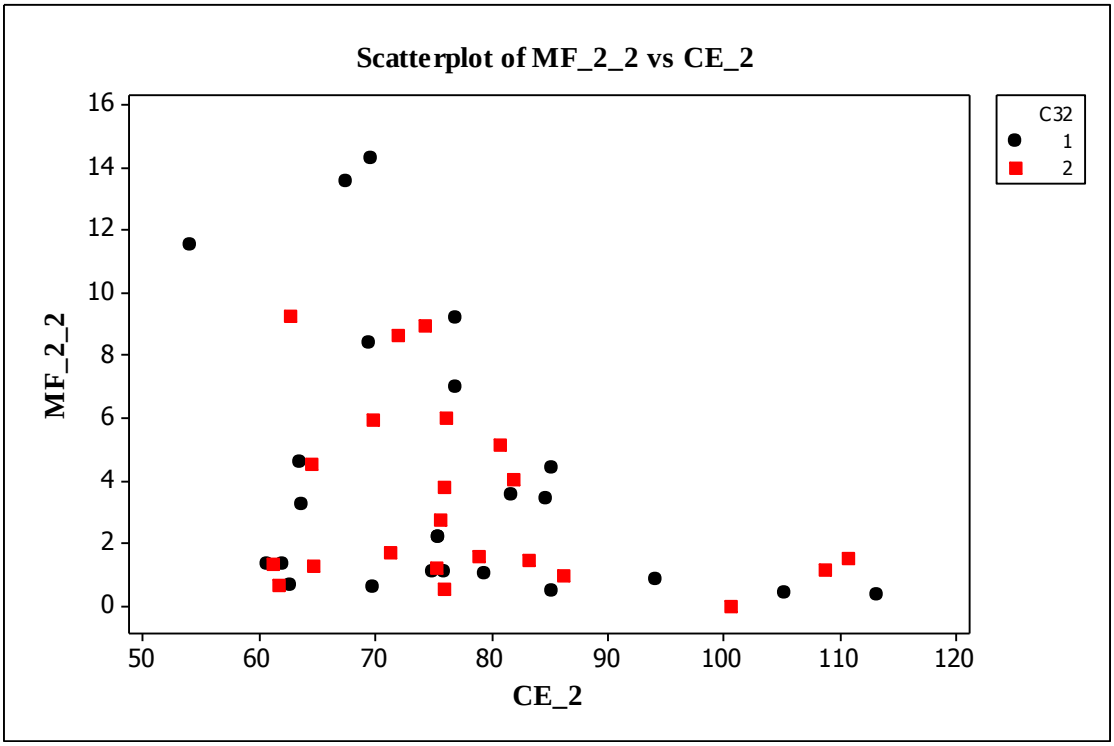


Fig 12 Scatterplot of Mutation Frequency against Cloning Efficiency for Laboratory B'

48. There was no significant difference between the replicates which were with (+S9) or without (-S9) S9 condition for the CE or MF values.

49. Three studies (AD65Z AD70SA and AD71UE) had higher MF values than for the other studies with the between study component in the anova having P =0.058 for –S9 and P<0.001 for +S9 conditions).

Laboratory K

50. A new laboratory, Lab K, sent in a small data set of MF values of hprt mutation in TK6 cells. This data was provided from possibly two studies: C175-004 consisting of 40 cultures and 50045-0001 consisting of 20 cultures. These two sets of data allow a comparison between different studies and between different vehicles within studies.

C175-004 40 cultures
Vinyl Acetate Media 10
Vinyl Acetate DMSO 10
Acetaldehyde Media 5
Acetaldehyde HBSS 5
Acetaldehyde Media 5
Acetaldehyde HBSS 5

50045-0001 20 cutures
IDX20963 (+S9) Media 5
IDX20963 (+S9) DMSO 5
IDX20963 (+S9) Media 5
IDX20963 (+S9) DMSO 5

51. The overall data show a distribution skewed to the right (Fig 13).

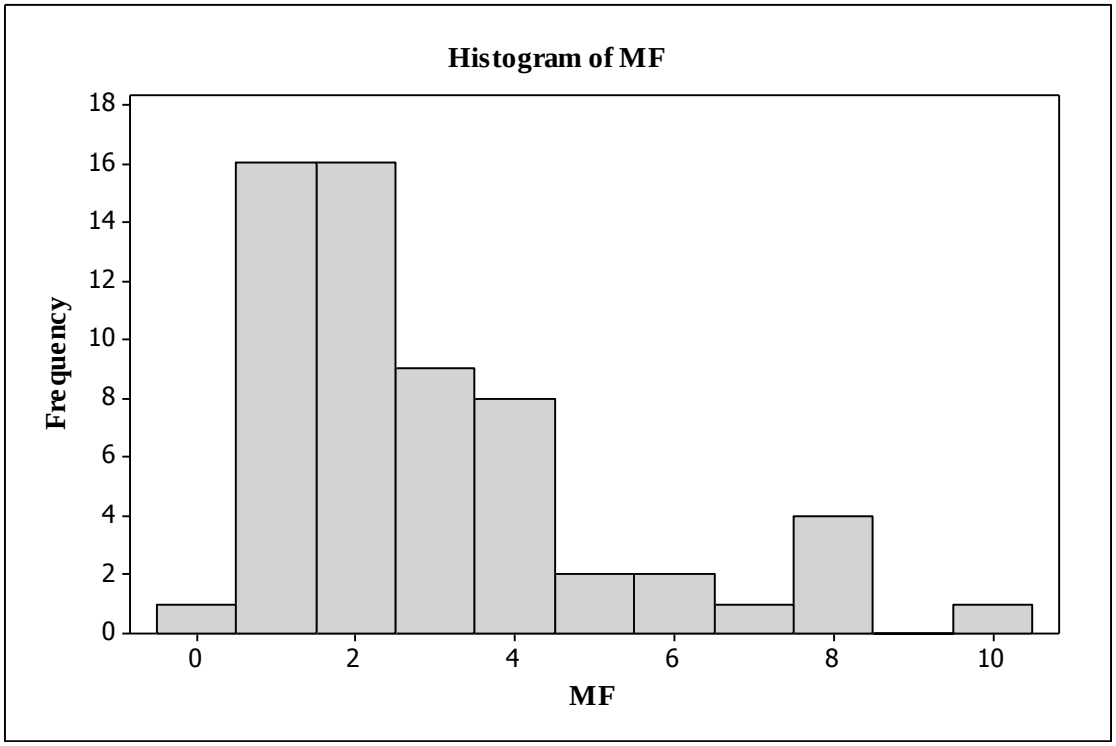


Fig 13 Distribution of Mutation Frequency data for Laboratory K

52. The difference in MF between the two studies was just significant (P=0.026).

One-way ANOVA: MF versus Study Number

```
Source        DF      SS     MS     F      P
Study Number   1   23.64  23.64  5.23  0.026
Error         58  262.21   4.52
Total         59  285.85

S = 2.126   R-Sq = 8.27%   R-Sq(adj) = 6.69%

                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level        N   Mean  StDev  ------+---------+---------+---------+---
50045-0001  20  2.047  1.084  (-----------*----------)
C175-004    40  3.378  2.480                    (-------*--------)
                               ------+---------+---------+---------+---
                                   1.60      2.40      3.20      4.00
```

53. This seems to be traceable to highly significant differences between the three experiments in the first study (P<0.001) with the controls for the vinyl acetate experiment being significantly different from the other two sets of controls.

One-way ANOVA: MF versus Experiment

Source	DF	SS	MS	F	P
Experiment	2	69.06	34.53	9.08	0.000
Error	57	216.79	3.80		
Total	59	285.85			

S = 1.950 R-Sq = 24.16% R-Sq(adj) = 21.50%

Level	N	Mean	StDev
Acetaldehyde	20	2.313	1.284
IDX20963 (+S9)	20	2.047	1.084
Vinyl Acetate	20	4.444	2.930

54. An I chart (Fig 14) showed that one set of cultures were very different from the others (around culture #s 10 -20).

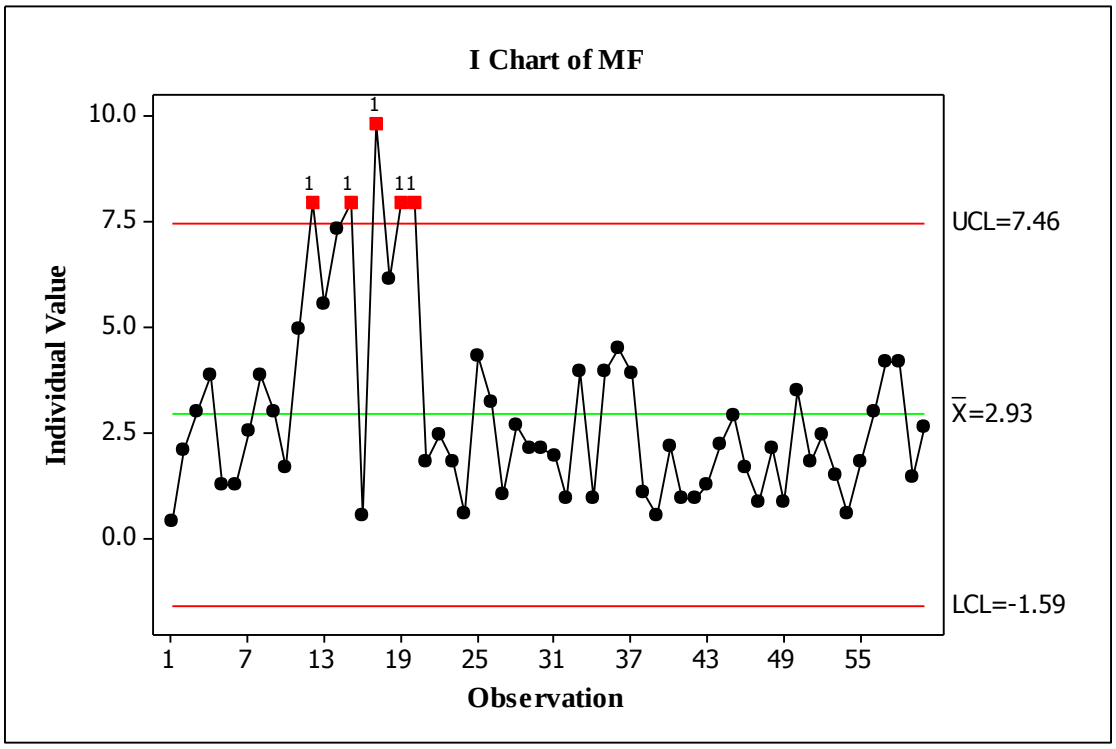


Fig 14 I Chart of Mutation Frequency for Laboratory K

55. An analysis of the different sub groups within Experiment 1 showed a highly significant difference between 'doses' with the second group of 10 cultures very different from the other 30. These were the controls for vinyl acetate with DMSO which had MF of 6.6 x 10^{-6} vs. approx 2.3 x 10^{-6} in the other groups. It is not clear what the reason for this is.

One-way ANOVA: MF versus Dose

```
Source DF      SS     MS     F      P
Dose    5  137.86  27.57  9.19  0.000
Error  34  102.02   3.00
Total  39  239.87

S = 1.732   R-Sq = 57.47%   R-Sq(adj) = 51.22%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level   N   Mean  StDev  -------+---------+---------+---------+--
1      10  2.296  1.162     (----*-----)
2      10  6.592  2.554                          (-----*-----)
3       5  2.194  1.362  (-------*-------)
4       5  2.254  0.808  (-------*-------)
5       5  2.355  1.519   (-------*-------)
6       5  2.448  1.724   (-------*-------)
                          -------+---------+---------+---------+--
                               2.0       4.0       6.0       8.0
```

56. There were no differences between the replicates or different vehicles in the IDX20963 (+S9) study

```
Analysis of Variance for MF_1

Source                DF       SS      MS     F      P
Replicate              1   2.0982  2.0982  2.24  0.154
Vehicle_1              1   3.1920  3.1920  3.41  0.083
Replicate*Vehicle_1    1   2.0698  2.0698  2.21  0.156
Error                 16  14.9778  0.9361
Total                 19  22.3378
```

Laboratory J

57. Laboratory J provided extra information in an update of their data file which identified the technical replicates. In the new file the same "Exp No." means that the experiments were conducted on the same day while the "replicate no." refers to the number of the culture performed on the same day under the same test condition.

There were significant differences between the 28 experiments for studies carried out in conditions both with (+S9) and without (-S9) S9.

One-way anova for –S9

One-way ANOVA: MF0% versus E0%

```
Source  DF      SS     MS     F      P
E0%     27  3698.2  137.0  5.18  0.000
Error   39  1030.4   26.4
Total   66  4728.7

S = 5.140   R-Sq = 78.21%   R-Sq(adj) = 63.12%
```

One-way anova for +S9

One-way ANOVA: MF5% versus E5%

```
Source  DF      SS     MS     F      P
E5%     27  3302.5  122.3  2.30  0.010
Error   36  1912.2   53.1
Total   63  5214.7

S = 7.288   R-Sq = 63.33%   R-Sq(adj) = 35.83%
```

3) Application of QC statistical methods

58. QC methods can be applied to MF data. The MF values are continuous / quantitative values and can be investigated using I graphs. The values are plotted consecutively and action levels are drawn up based upon lines approximately 2 and 3 SD from the sample mean. (In practice, the calculation of the SD is slightly more complex than just using the sample SD) QC graphs can be produced for both individual samples or for groups of samples. In the cases here individual samples have been plotted.

59. The graphs for groups of samples could be used if there were groups of observations such as for an individual experiment where there may be replicated negative control cultures for the different experimental conditions of presence/absence of S9 and lengths of exposure. Such data can be analysed using X bar charts. (There are other methods such as Z-MR charts where the 'runs' are short.)

60. Other QC methods are possible such as R and S charts. Xbar-R charts are used for sample means and ranges. They track on the same chart the mean values and the within group variation of a series of groups of samples. The Xbar-S charts are similar except that they are more appropriate when the sub-group sizes are larger (i.e. more than 8). Both methods can be used with unequal sample sizes. In both cases special calculations are used for the estimates of the standard deviations used in the charts.) Another chart, called the Zone chart, is a simplified version where each point is scored for how unusual it is. Other approaches that could be used include the moving average, EWMA and CUSUM charts available in Minitab.

61. In this case I charts for individual cultures have been used. The control limits are drawn on the graphs and the individual values plotted. The numbers refer to a series of tests. These tests (based upon, for instance, the Western Electric criteria) provide indications of when a process has gone out of control. For example a '1' indicates that the point is more than 3 standard deviations from the central line., a '3' that there are 6 consecutive points in a row which are all increasing or are all decreasing.

62. Note that the sample numbers are entered in consecutive order for each of the experimental combinations in turn. So the order may, for example, be that the first set might be the -S93h condition,

then the +S93h conditions then the –S924h condition. As a result the individual experimental results are spread over the whole range. While it might be possible to reorganise the datasets into an order based upon experiments this would be time consuming and it is not always clear which cultures are from the same experiment. One consequence is that patterns such as a series of cultures which are very different from the overall mean will appear as a cyclical pattern.

63. A Moving Range (MR) chart could also be used. (Sometimes this is combined with a U chart to form an I-MR chart). This uses a chart of MRs based upon artificial subgroups created for consecutive measures. MR charts use an estimate of the 'process variation', s, with MR / d2. The MR is of length 2 as adjacent values have the best chance of being alike. A number of other methods can be used to estimate s such as the median of the MR.

64. The MF values are not, in general, normally distributed but rather the mean and variances are positively correlated. The usual transformations in such circumstances are the logarithmic or the square root transformation. These should probably be used if the I charts are to be widely used.

65. QC statistical methods can be useful for identifying patterns of change over time. They can also be used to explore inter-experimental variability. An example is given for one set of data (from Laboratory D) where an I chart (Fig 15) gives an example of how the experiment and experiment x S9mix interactions identified by an anova (see Fig 5) can be illustrated. QC statistics help identify temporal changes and help explain the non-normality of the data and show and highlight between experiment variability.

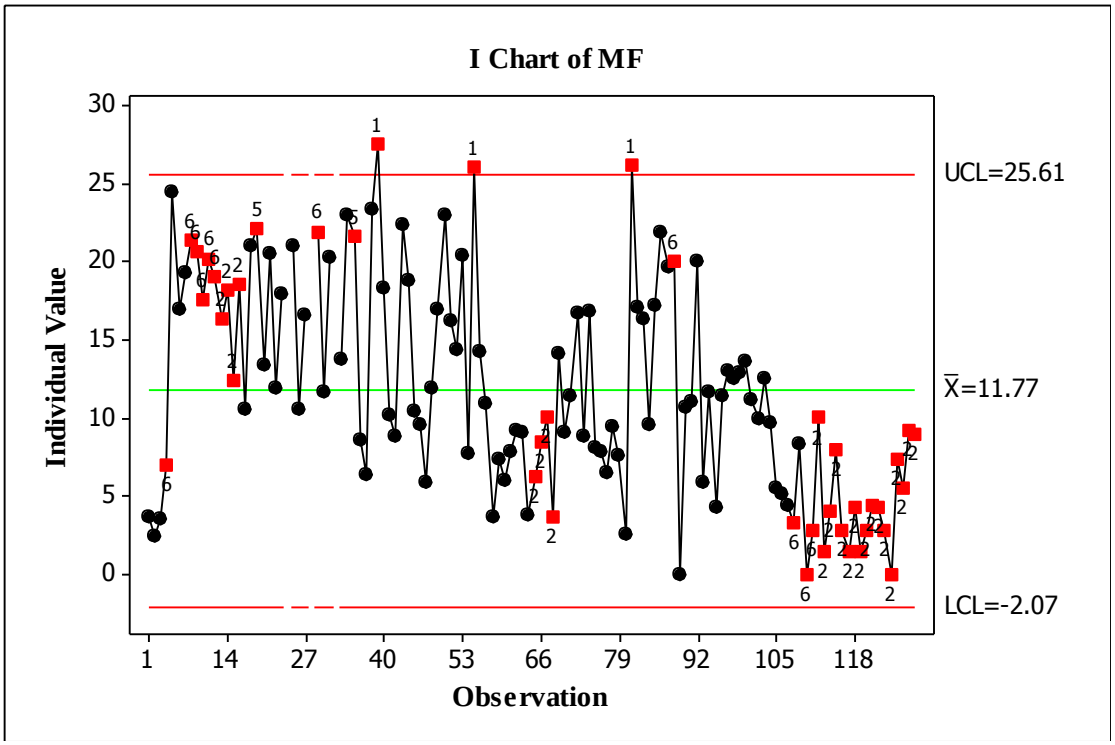


Fig 15: I chart of MF data for Lab D

4) Analysis of individual plate counts

66. Individual plate counts were available from one laboratory (C). These counts were split into 15 combinations of S9 conditions (+S94h –S94h –S924h) and 5 sets of vehicles (culture medium, DMSO, acetone, ethanol, tetrahydrofurane). In most cases the data in the datasets were a poor fit to a Poisson distribution with an excess of zero and large counts and fewer intermediate values than would be expected from a Poisson distribution. This may represent either a non-Poisson distribution for the individual plates or be representative of inter-experimental variability. (This was, though, less so for stacks 7, 8, 11 and 13 where the P value for the Goodness of Fit test was >0.05).

67. The 4212 plate counts from all the replicate experiment and all experimental conditions were combined. The distribution of this dataset was significantly different from a Poisson distribution (P<0.001) (Fig 16), with an excess of zeros and large counts (Fig 17),

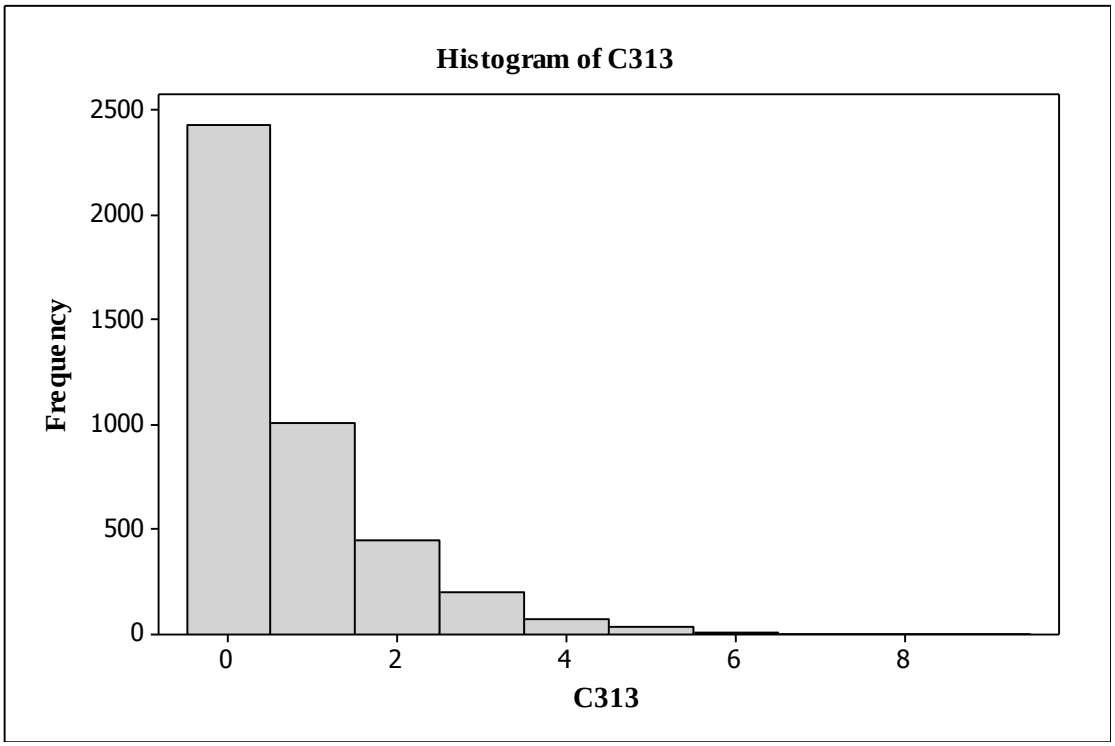


Fig 16 Distribution of plate counts from Laboratory C

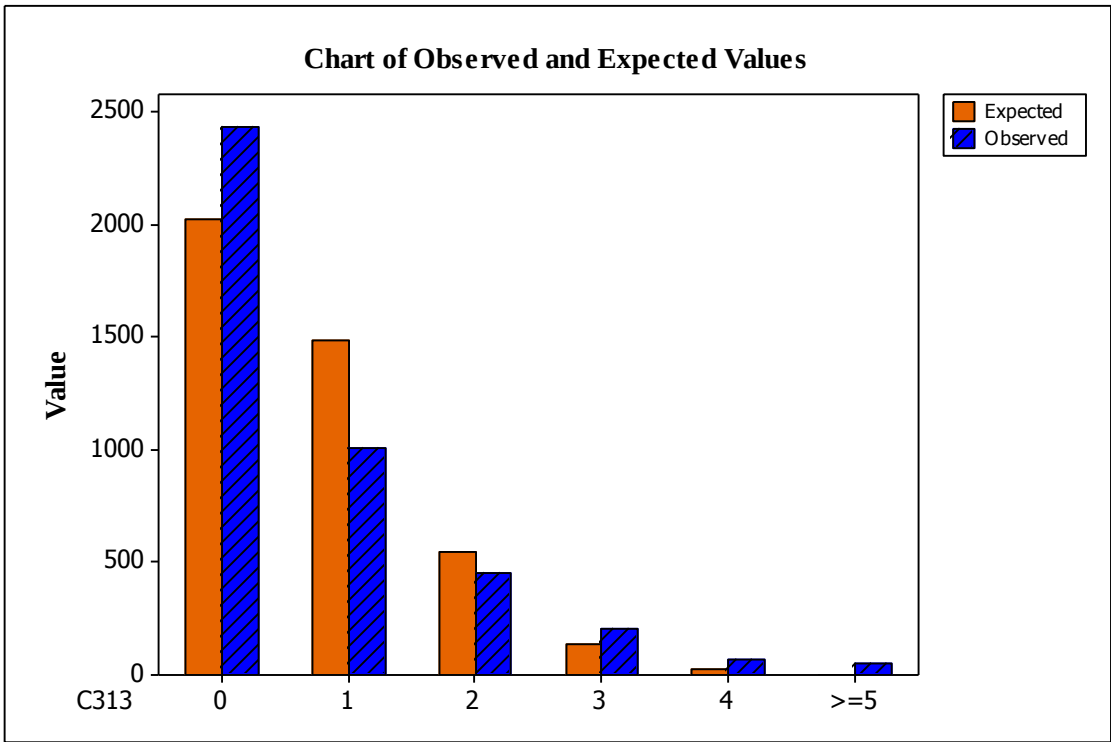


Fig 17 Plot of expected and observed counts based upon a Poisson distribution for Lab C

```
Poisson mean for C313 = 0.734330

                    Poisson              Contribution
C313   Observed  Probability   Expected     to Chi-Sq
0          2430     0.479827    2021.03        82.758
1          1005     0.352351    1484.10       154.666
2           449     0.129371     544.91        16.882
3           206     0.031667     133.38        39.536
4            69     0.005814      24.49        80.920
>=5          53     0.000970       4.09       585.360

>=5 counts
   5      35
   6      11
   7       2
   8       4
   9       1
  N=      53
```

68. In the case of Poisson distributed data, the mean should equal the variance. An attempt to test whether the poor fit to the Poisson was a consequence of excess variation between plates within an experiment was carried out by calculating the ratio, H, of the variance and the mean for each individual replicate and seeing if the distribution of the ratios (H) is symmetrically distributed around one, This is possible for Lab C where there are 6 plates in each replicate. Overall these do not fit a Poisson.

69. The distribution of the values of H is skewed to the right (Fig 18) but both the mean and median of the values of H are approximately 1. However note the large number of cases, 143 out of 702 or 20%, where zero counts were obtained from all 6 plates in the replicate.

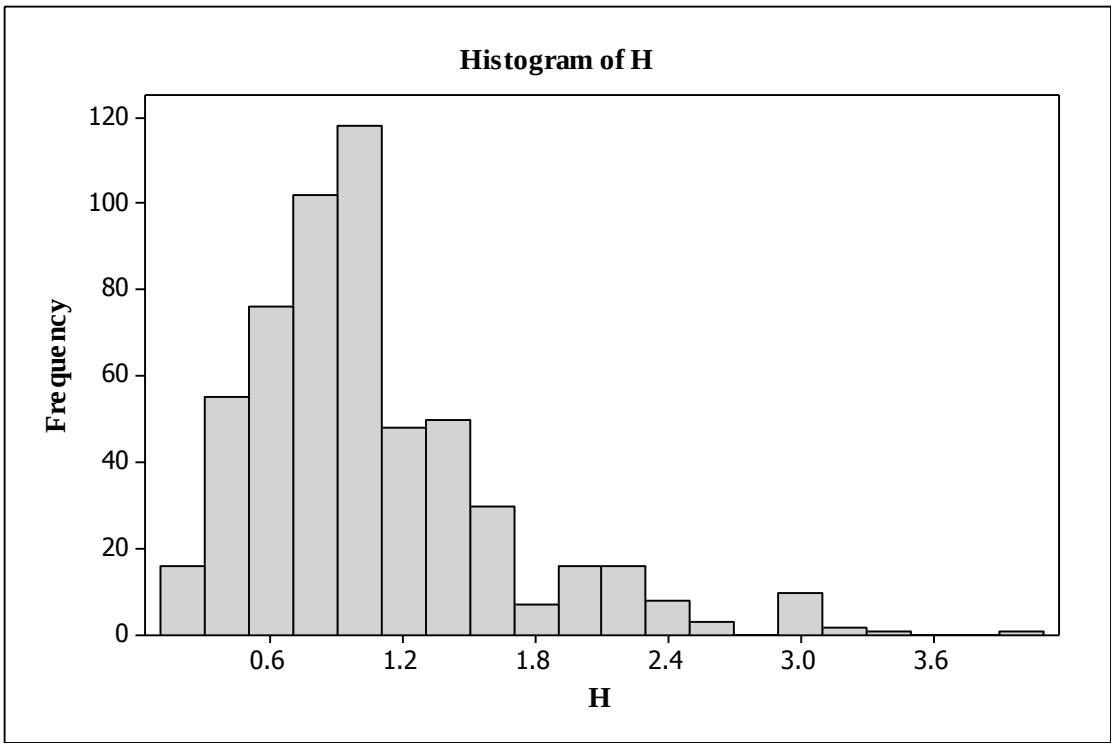


Fig 18 Distribution of H values from Laboratory C data

Descriptive Statistics: H

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Median	Maximum
H	559	143	1.0683	0.0246	0.5811	0.1200	1.0000	4.0000

Extension to the Global Equivalence Factor (GEF) approach

70. The GEF approach used for the Mouse Lymphoma Assay (MLA) test is, in effect, an absolute change. Based upon Moore *et al* (2003) paper, the GEF is defined as the mean plus one standard deviation based upon the distribution of the historical negative control data collected across laboratories. Provided the concurrent negative control falls within a predefined range, again based upon the historic negative control data, then an induced mutation frequency (IMF) value obtained from a treated group which equals or exceeds the GEF triggers a statistical analysis and a significant trend test signals a positive result.

71. It is not clear whether combining the data sets across the distributions would provide a suitable distribution to base considerations of the Global Equivalence Factor GEF approach.

72. Histograms of the individual MF culture values across all laboratories are shown for the five CHO laboratories (Fig 19) and the three V79 laboratories (Fig 20). In both cases there are highly significant differences in mean values across the laboratories ($P<0.001$).

Overall between lab variability

One-way ANOVA: CHO comb versus Lab cho

```
Source     DF       SS      MS      F      P
Lab cho     5   8841.4  1768.3  99.47  0.000
Error    1402  24923.9    17.8
Total    1407  33765.3

S = 4.216   R-Sq = 26.18%   R-Sq(adj) = 25.92%
```

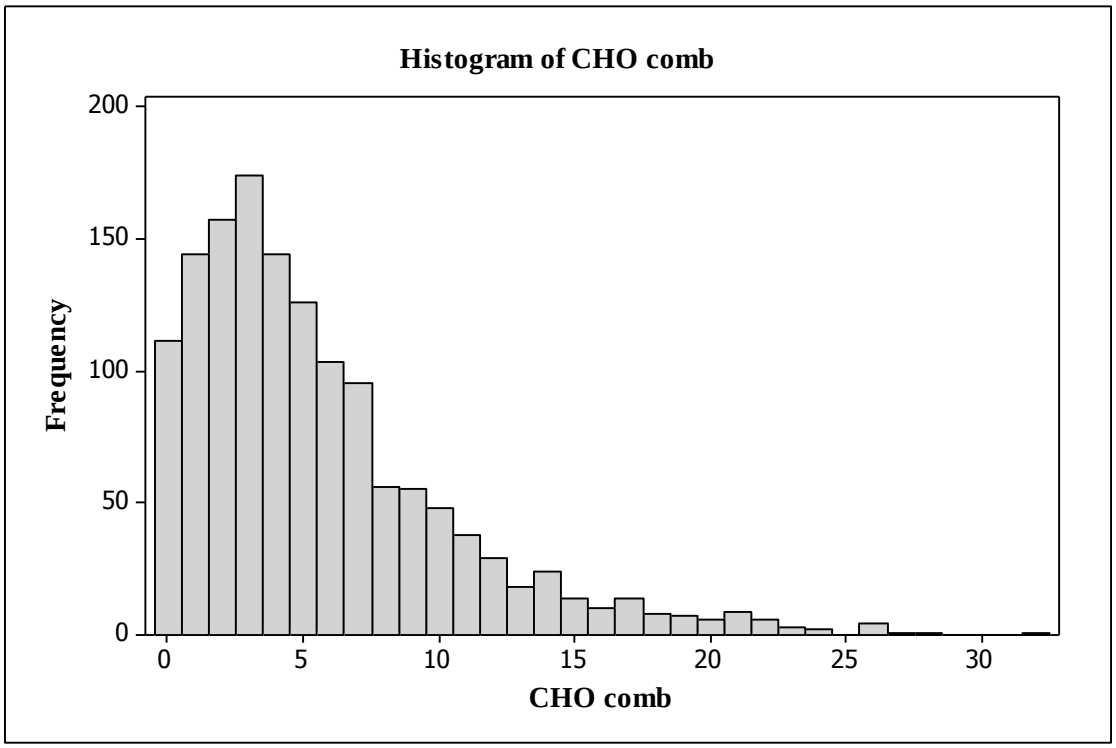


Fig 19 Distribution of MF values combined across all CHO laboratories

One-way ANOVA: V79comb versus Lab V79

```
Source    DF      SS    MS      F      P
Lab V79    3   12926  4309  42.29  0.000
Error    941   95872   102
Total    944  108798

S = 10.09   R-Sq = 11.88%   R-Sq(adj) = 11.60%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level    N   Mean  StDev   -+---------+---------+---------+--------
1      135  17.91   9.44                            (-----*----)
2      268  10.18   8.45    (---*---)
3      411  17.74  11.57                              (--*--)
4      131  10.53   8.76   (-----*-----)
                           -+---------+---------+---------+--------
                          9.0       12.0      15.0      18.0

Pooled StDev = 10.09
```

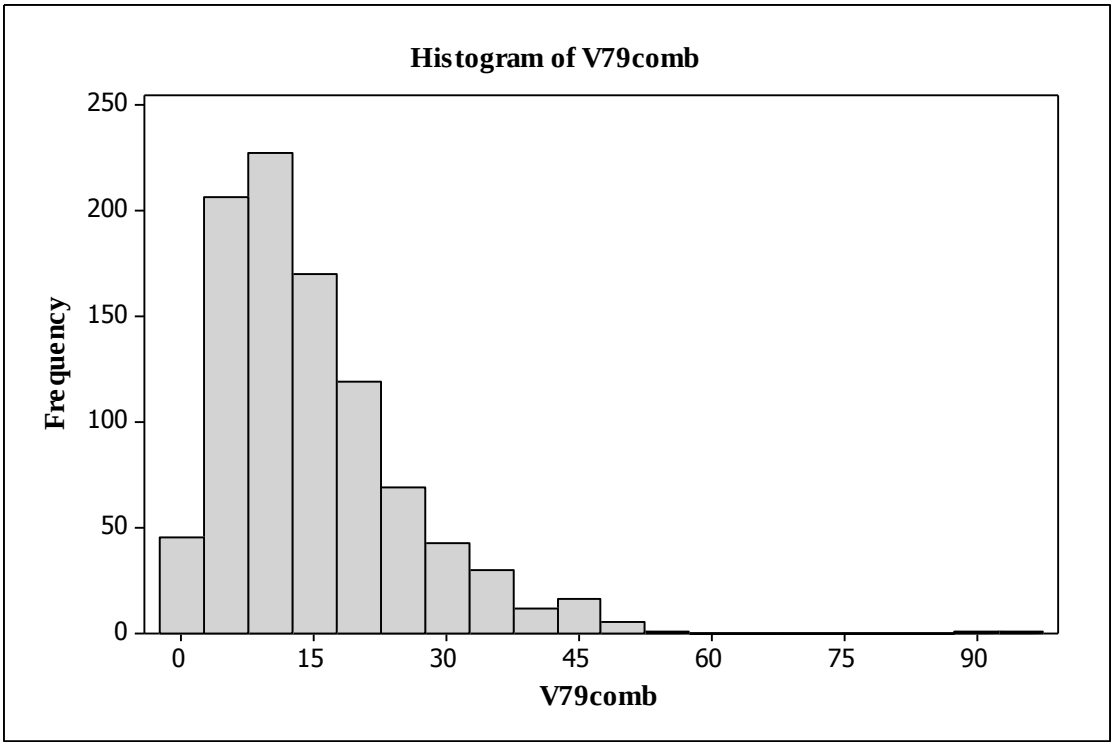


Fig 20 Distribution of MF values combined across all V79 laboratories

Number of cells necessary to avoid zero mutant frequencies

73. This aspect was considered by Arlett et al (1989) who provided two tables to identify the number of plates and cells to score

74. Firstly, they considered the probability of all the plates having zero mutants for different mutation frequencies and number of plates at that concentration level (Table 3.1)

75. Assume that the total number of cells scored is based upon 10 plates each with 100,000 cells plated per plate i.e. a total of 10 x 10^5 or a million or 10^6 cells.

76. Assume that mutations arise randomly (as in a Poisson distribution)

77. If a million cells are scored and there is 80% Plating Efficiency (PE) then with a MF = 6.25×10^{-6} there should be very few cases of zero (0) counts in all plates (approx 0.67 %.)

78. This is based upon the np=5 'concept'

79. For n = 0.8 x 10^6 and p = 6.25×10^{-6}

80. Then (6.25×10^{-6}) * 10^6 * 0.80 = 5 and 0.67% sets of plates will have zero.

81. A fuller set of tables can be found in Table 3.1 of the UKEMS chapter by Arlett et al.

82. This gives the probability that all the plates have zero mutants for a given MF/10^6

83. This is based upon the count taking into account n (the number of plates with 10^6 cells plated per plate) the %CE and the MF/10^6

84. So for MF = $5/10^6$ 50% CE and 2 plates the count is 5 and the prob of 0 is 0.00673

85. For MF = $1/10^6$ 50% CE and 2 plates the count is 1 and the prob of 0 is 0.368

(This can be compared with micronucleus calculations)

86. Table 3.2 in Arlett et al (1989) shows the number of cells to ensure that the probability is less than 0.05 that all the plates would be zero for different MFs and numbers of plates

87. To estimate this there is a need to organize the number of cells and the number of plates and the MF so that the expected count is 3. If so, then for mean of 3, we would expect exp (-3) = 0.049781 to be the proportion of sets with zero in. This agrees with the numbers in Arlett et al's Table 3.2.

Transformations

88. An example of the Box-Cox plot to identify an appropriate transformation is the use of the plot to identify the optimum value of lambda use in a Box-Cox power transformation. Using the data from Lab G's V79 MF data two plots (Fig 21, Fig 22) are produced (below).

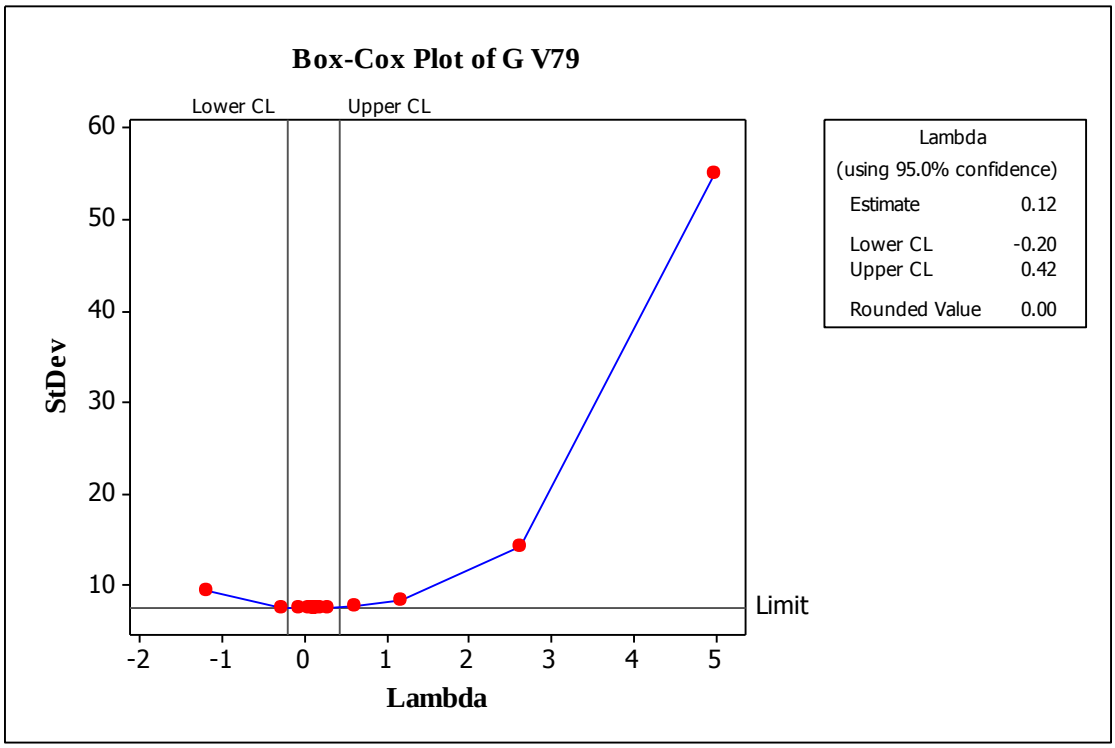


Fig 21 Box-Cox method of identifying appropriate transformation of Lab G data.

89. In this case the best estimate of lambda is 0.12 which is rounded down to 0 and is then equivalent to untransformed data. The 95% CIs on lambda is from -0.20 to 0.42 but the graph suggests that both the square root transformation (lambda=0.05) or the log transformation (lambda =1) could probably also be used satisfactorily.

90. The second plot (Fig 22) shows the Johnson transformation with the best transformation 'estimated' to be,

```
Y = -2.31193 + 1.49438 * Asinh( ( X - 4.29284 ) / 4.99466 )
```

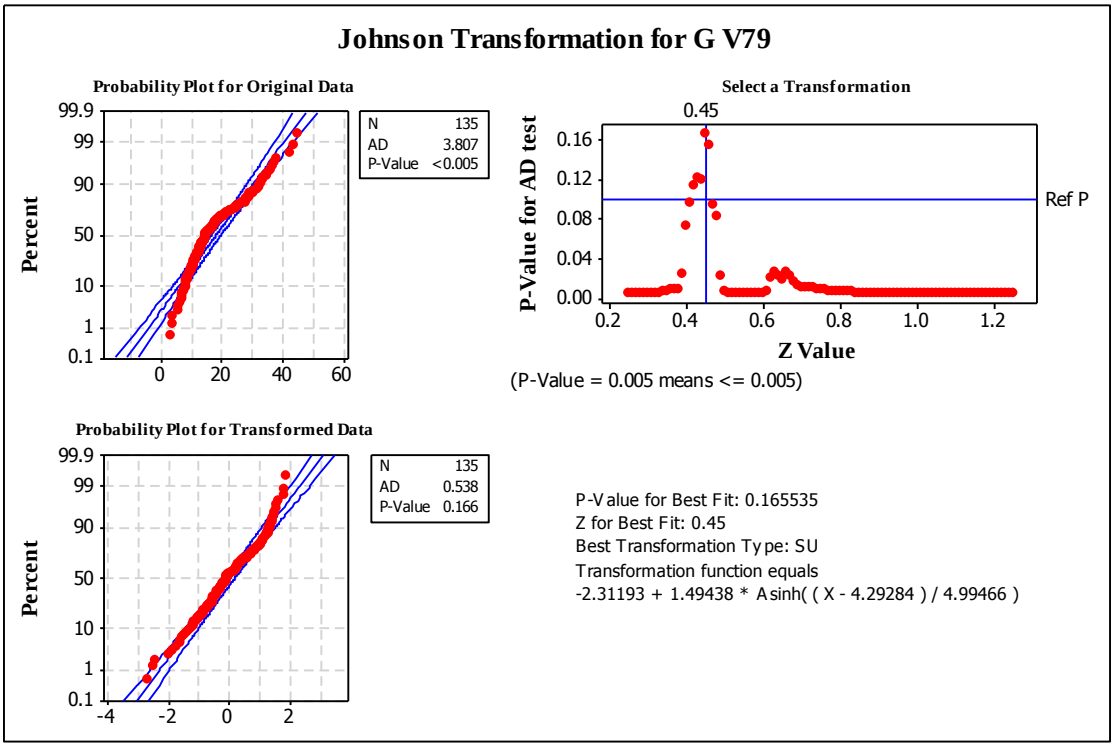


Fig 22. Johnson Transformation applied to Lab G data

Similar analyses can be done for other datasets.

References

Arlett, C.F., Smith, D.M., Clarke, G.M., Green, M.H.L., Cole, J., McGregor, D.B. and Asquith, J.C. (1989). Mammalian Cell Gene Mutation Assays Based upon Colony Formation. In: Statistical Evaluation of Mutagenicity Test Data, Kirkland, D.J. Ed., Cambridge University Press, pp. 66-101.

Hayashi, M., Dearfield, K., Kasper P., Lovell D. & Thybaud, V. (2011) Compilation and use of genetic toxicity historical control data. Mutation Research 723 87-90.

Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Burlinson, B., Cifone, M., Clarke, J., Delongchamp, R., Durward, R., Fellows, M., Gollapudi, B., Hou, S., Jenkinson, P., Lloyd, M., Majeska, J., Myhr, B., O'Donovan, M., Omori, T., Riach, C., San, R., Stankowski, L.F. Jr., Thakur, A.K., Van Goethem, F., Wakuri, S. & Yoshimura, I. (2006). Mouse lymphoma thymidine kinase gene mutation assay: follow-up meeting of the international workshop on genotoxicity testing, Aberdeen, Scotland, 2003, assay acceptance criteria, positive controls, and data evaluation, Environmental and Molecular Mutagenesis 47 1-5.

Table 1: Summary of datasets provided

Lab A MF
CHO – Hprt – S9 n= 23
CHO – Hprt + S9 n= 18
AS52-gpt –S9? n= 11
AS52-gpt +S9 n= 5

Lab B MF
CHO – Hprt – S9 n= 31
CHO – Hprt + S9 n= 32
AS52-gpt –S9 n=261
AS52-gpt+S9 n=245

Lab C MF from 6 plates and corrected MF
CHO-Hprt data-S9(4h) n= 86 (43)
CHO-Hprt data-S9(24h) n= 86 (43)
CHO-Hprt data+S9(4h) n=168 (84)
MF from 6 plates and corrected MF n= 351, Duplicate replicates (A and B)
Individual plate data available
Individual plate data available, number of different solvents
MF = mutant frequency (per 1 million cells) corrected with the cloning efficiency at the end of the expression period (CE2)

Lab D MF
CHO-K1- Hprt data n=124
(Individual MF from four cultures; only three cultures are included where one was deemed to be outside acceptable limits)
16 experiments with matched +S9 and –S9 4 MF values/experiment

Lab E MF
CHO-K1-BH4- Hprt data -S9 n=265
CHO-K1-BH4- Hprt data+S9 n=272

Lab F MF
CHO-Hprt-S9 n=153
CHO-Hprt+S9 n=153
L5178Y-Hprt-S9 n= 40
L5178Y-Hprt+S9 n= 40
CHO-WBL-S9 n= 2
CHO-WBL+S9 n= 2

Lab G MF
V79-Hprt-S9 n= 68
V79-Hprt+S9 n= 67

Lab H MF
V79-Hprt - S9 Day 6 n= 67
V79-Hprt - S9 Day 8-9 n= 67
V79-Hprt + S9 Day 6 n= 67
V79-Hprt + S9 Day 8-9 n= 67
2 replicate cultures

Lab I MF
V79-Hprt 286 experiments in all
Multiple solvents 2 replicates for each experiment, but not for every one,
Positive control data also

Lab J MF
V79-HPRT-S9 n= 67
V79-HPRT+S9 n= 64

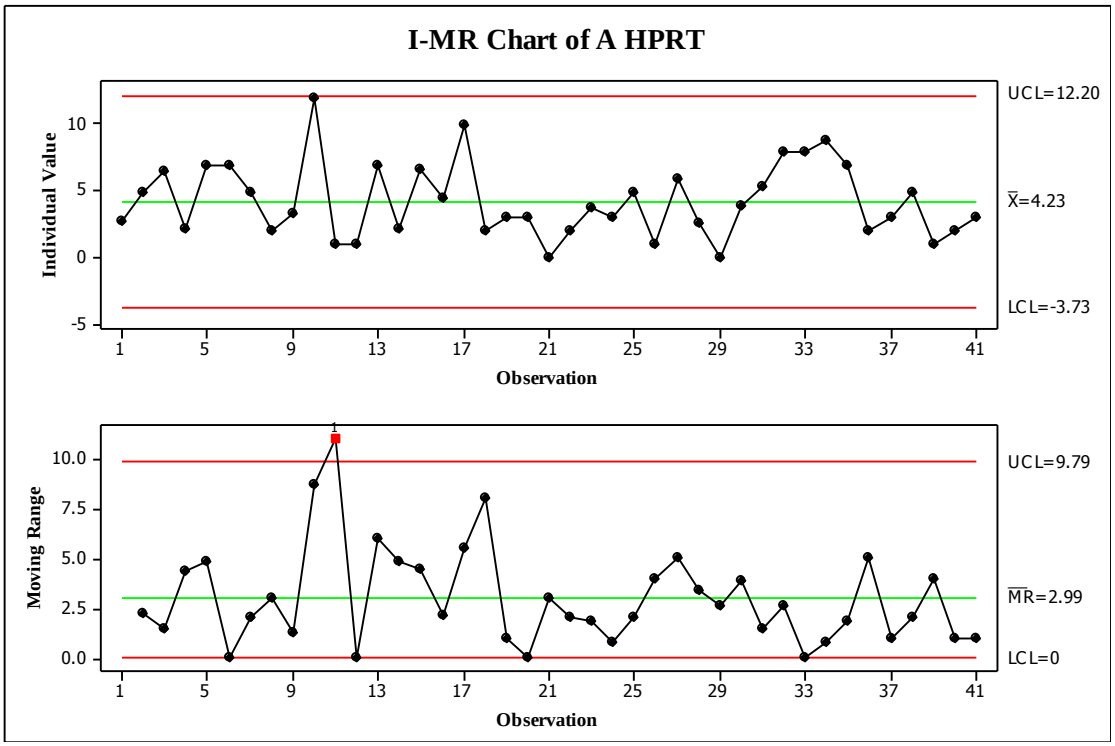
Provided September 2013

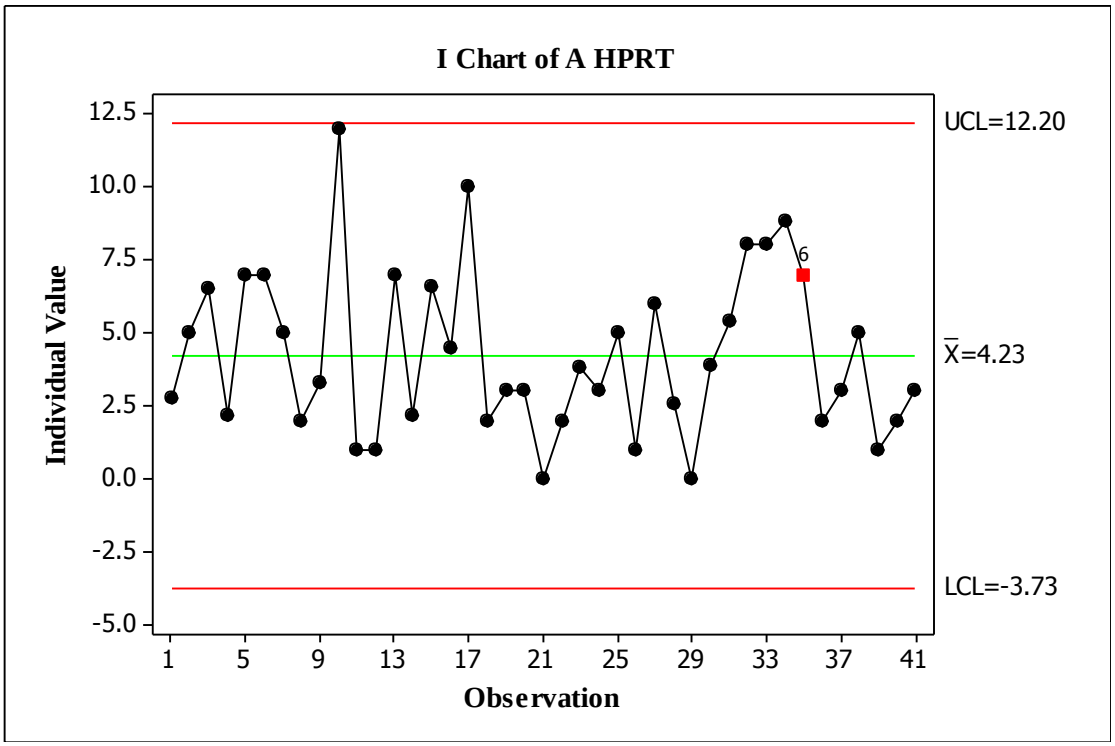
Lab B' MF
CHO – Hprt – S9 n= 22
CHO – Hprt + S9 n= 24

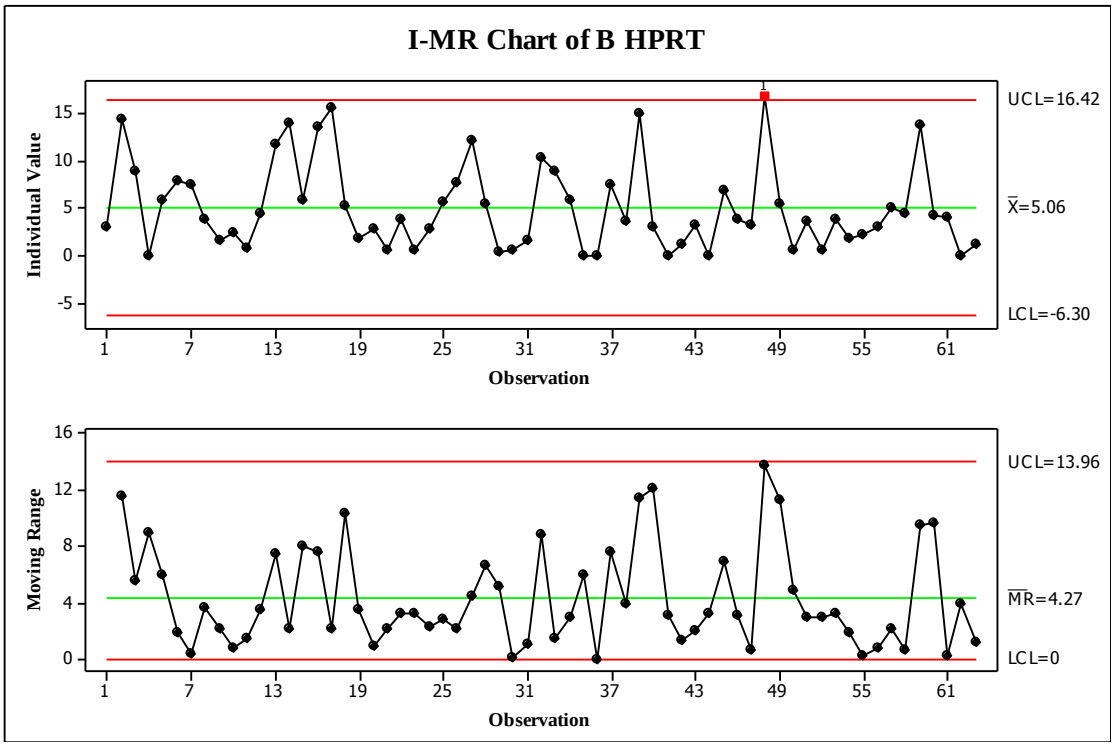
Lab K MF

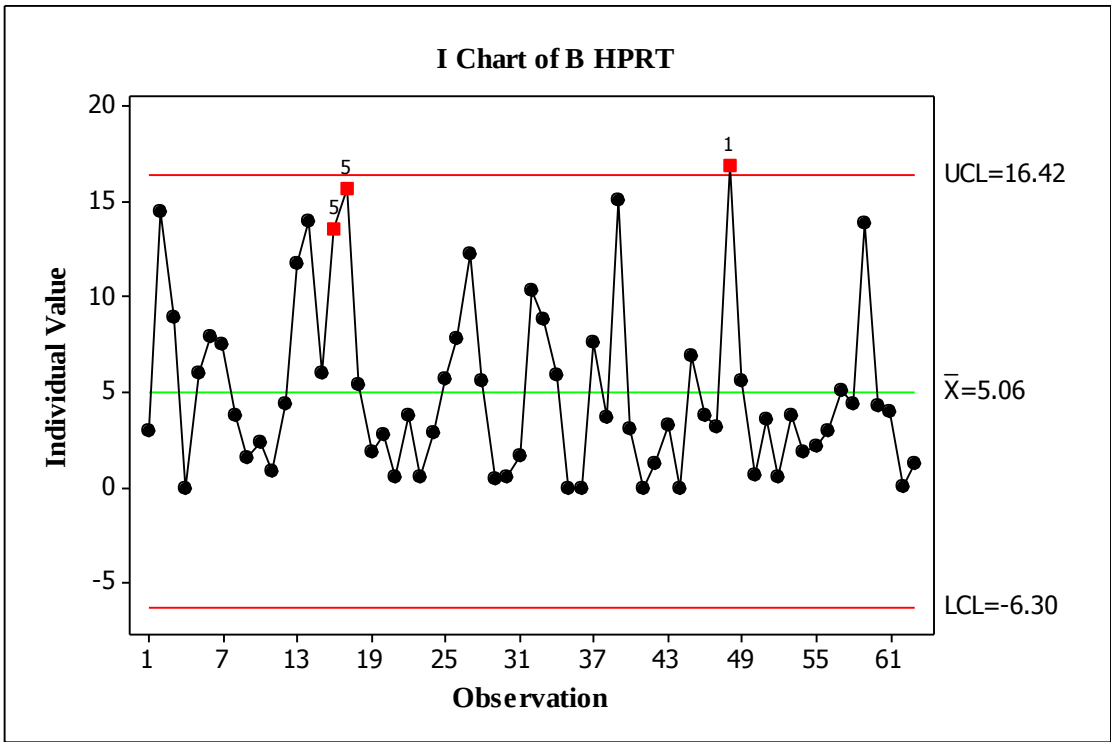
TK6-Hprt – S9	n= 40?
TK6-Hprt + S9	n= 20?

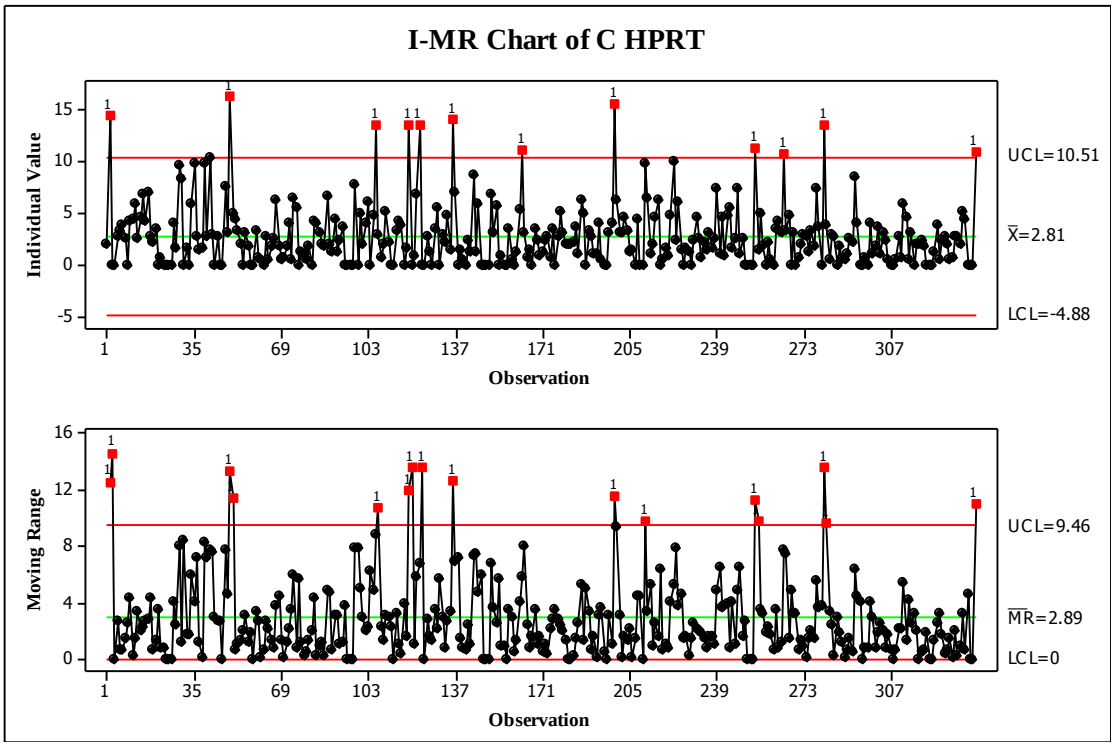
Appendix of I charts for individual laboratories

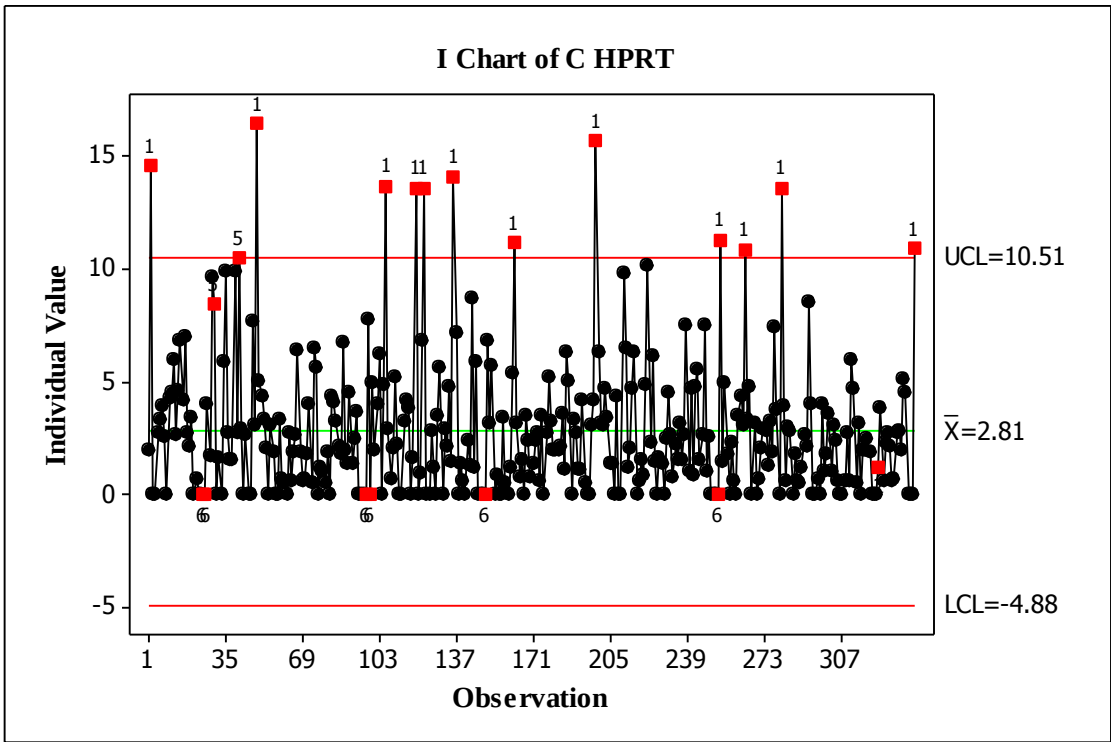

I-MR Chart of A HPRT
Individual Value
UCL=12.20
X̄=4.23
LCL=-3.73
Observation
Moving Range
UCL=9.79
MR̄=2.99
LCL=0
Observation

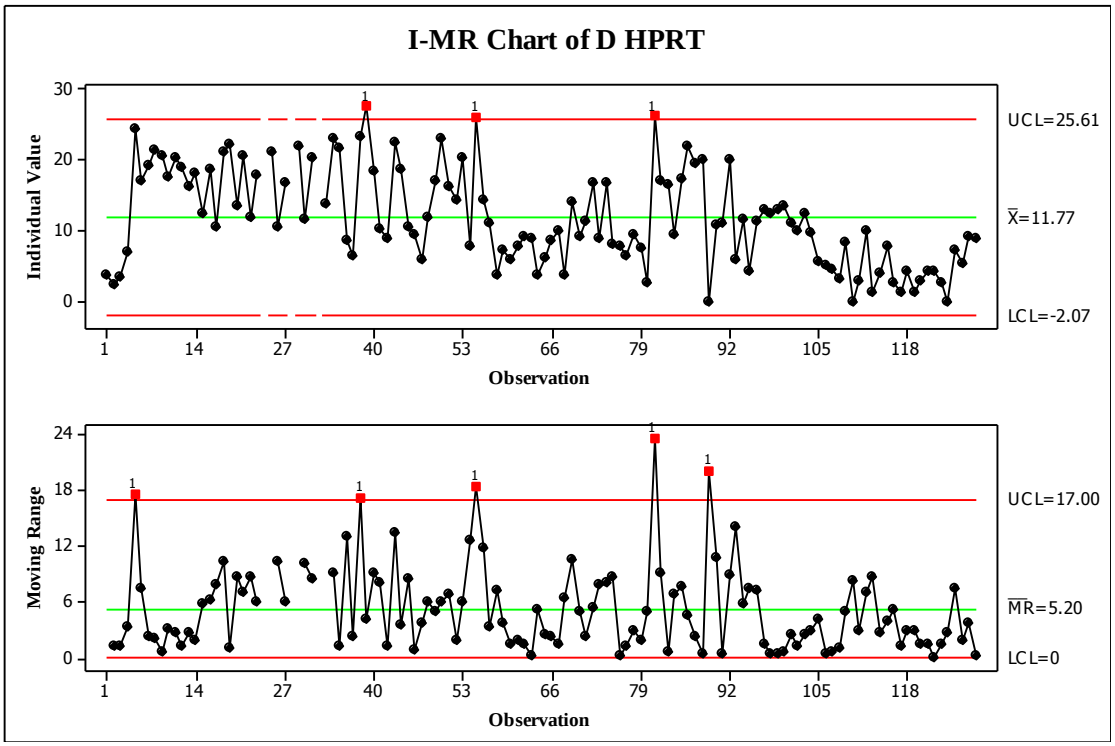

I Chart of A HPRT
Individual Value
UCL=12.20
X̄=4.23
LCL=-3.73
Observation

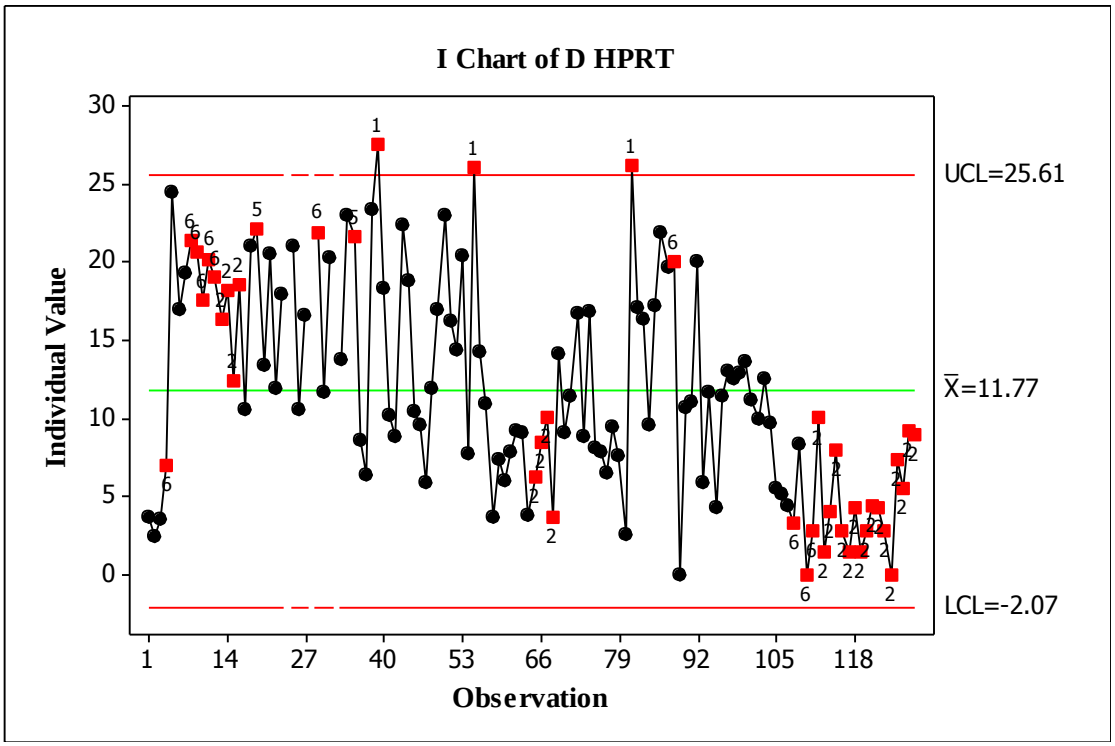

I-MR Chart of B HPRT
Individual Value
UCL=16.42
X̄=5.06
LCL=-6.30
Observation
Moving Range
UCL=13.96
MR̄=4.27
LCL=0
Observation

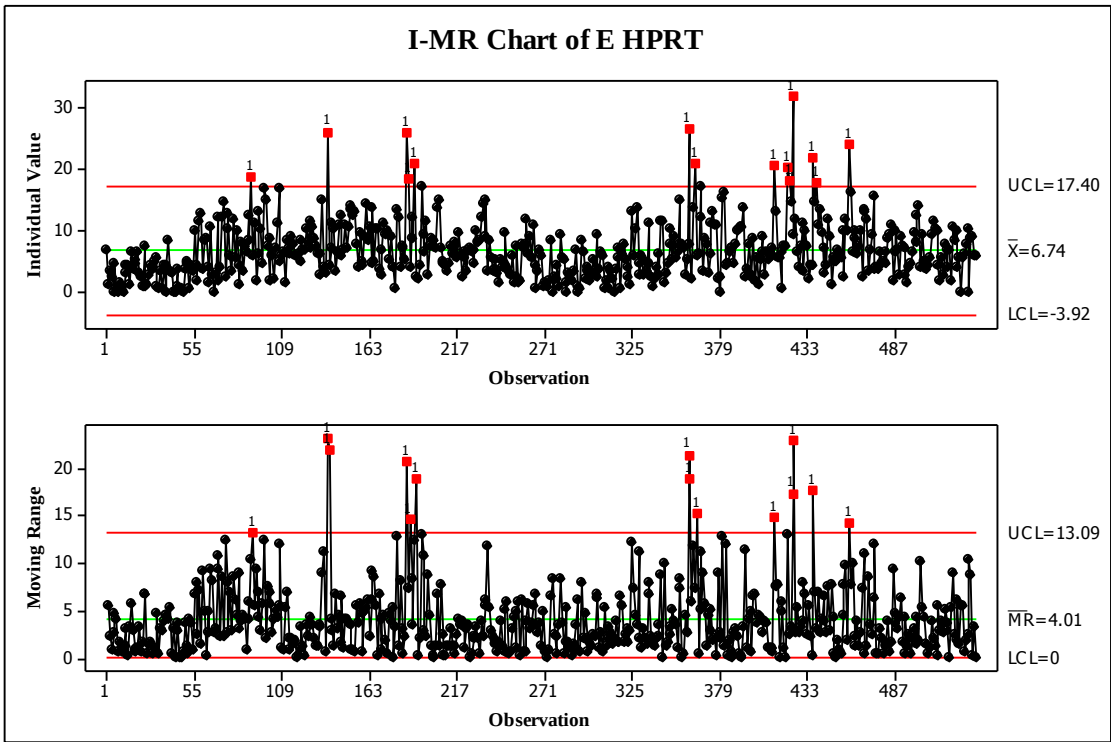

I Chart of B HPRT
Individual Value
UCL=16.42
X̄=5.06
LCL=-6.30
Observation

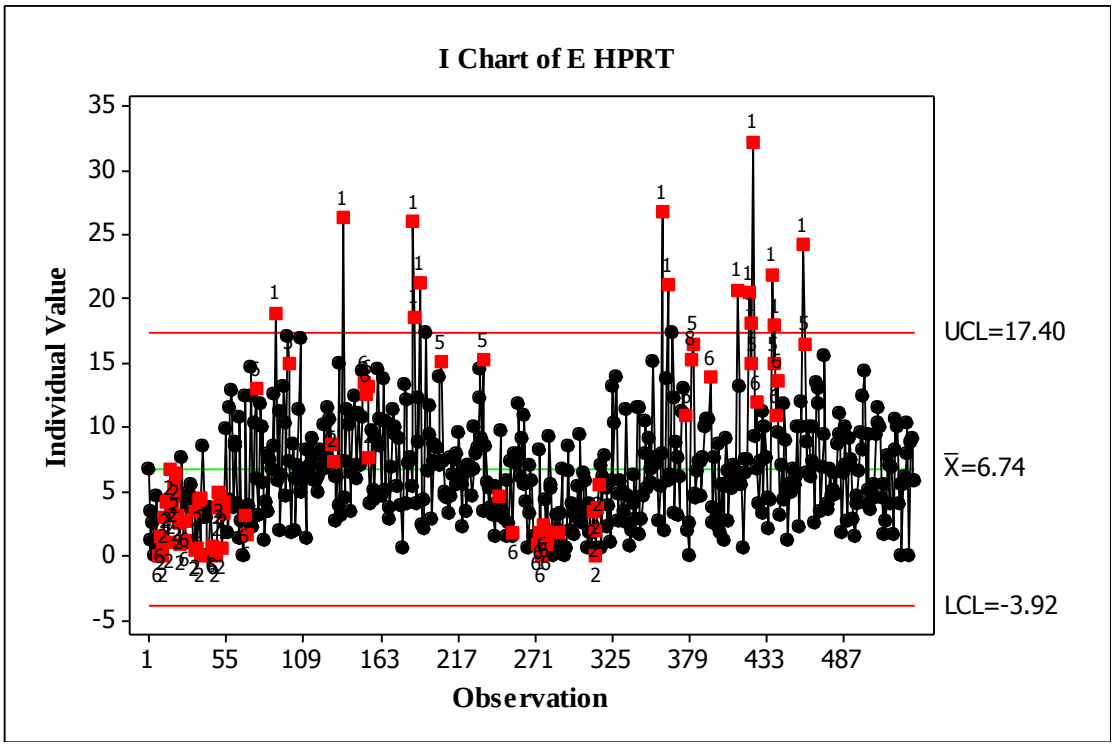

I-MR Chart of C HPRT
Individual Value
UCL=10.51
X̄=2.81
LCL=-4.88
Observation
Moving Range
UCL=9.46
MR̄=2.89
LCL=0
Observation

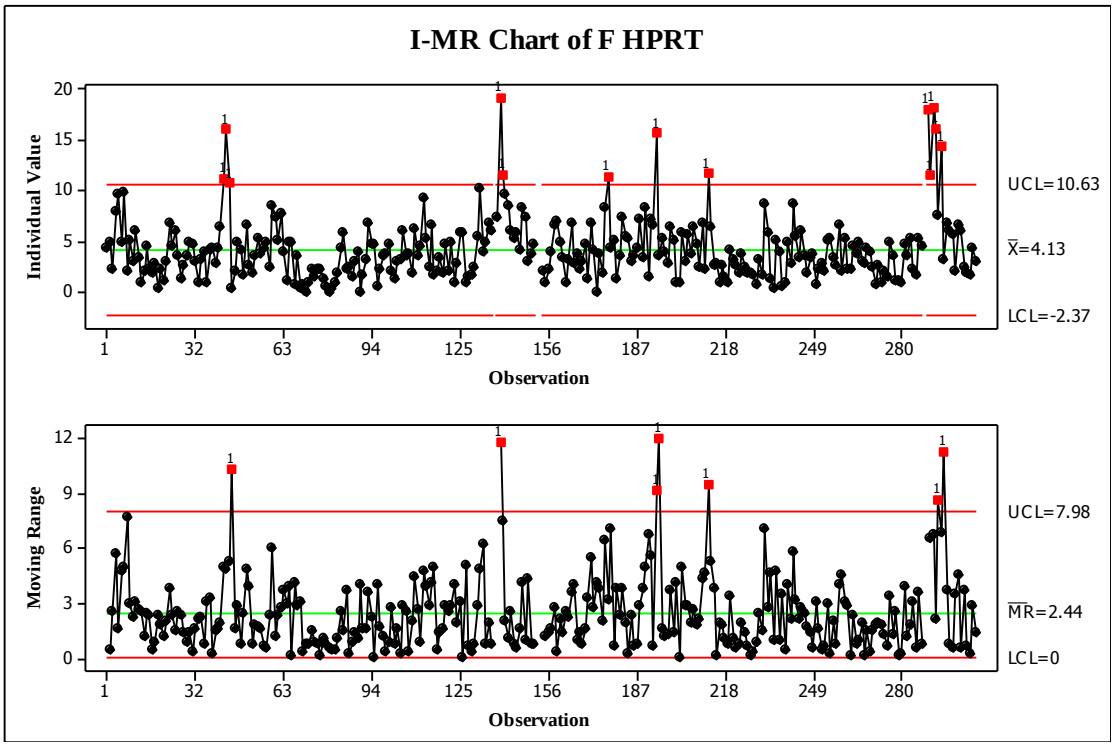

I Chart of C HPRT
Individual Value
UCL=10.51
X̄=2.81
LCL=-4.88
Observation

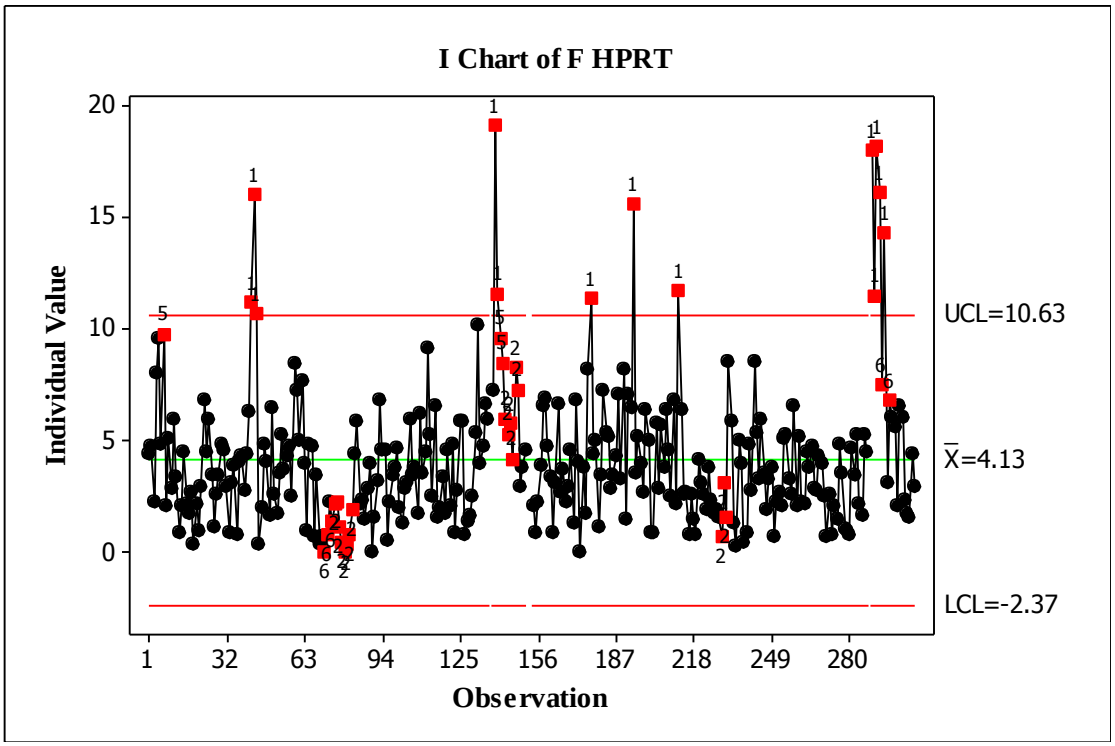

I-MR Chart of D HPRT
Individual Value
UCL=25.61
X̄=11.77
LCL=-2.07
Observation
Moving Range
UCL=17.00
MR̄=5.20
LCL=0
Observation

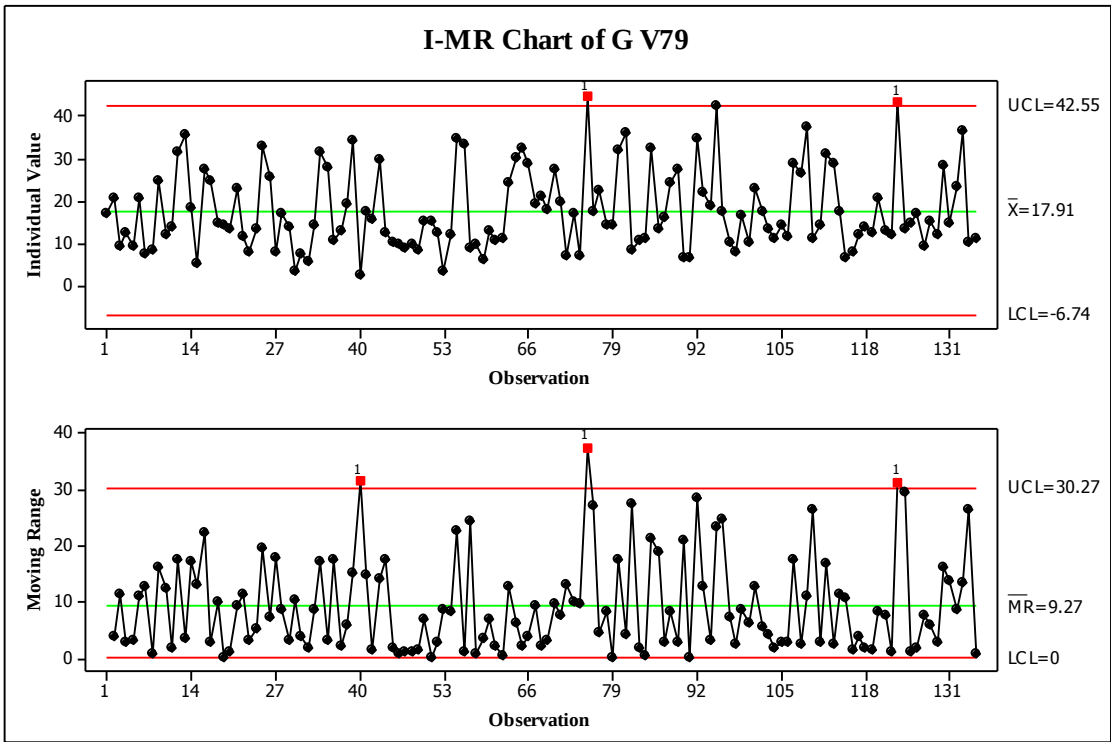

I Chart of D HPRT
Individual Value
UCL=25.61
X̄=11.77
LCL=-2.07
Observation

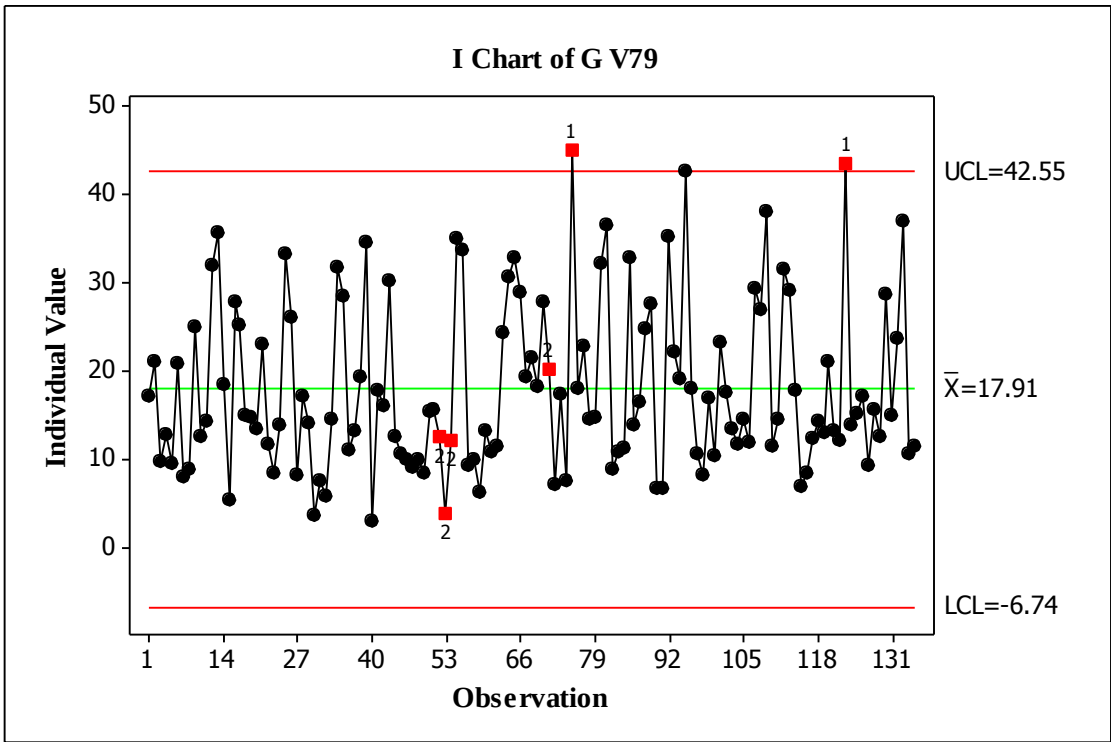

I-MR Chart of E HPRT
Individual Value
Observation
UCL=17.40
X=6.74
LCL=-3.92
Moving Range
Observation
UCL=13.09
MR=4.01
LCL=0

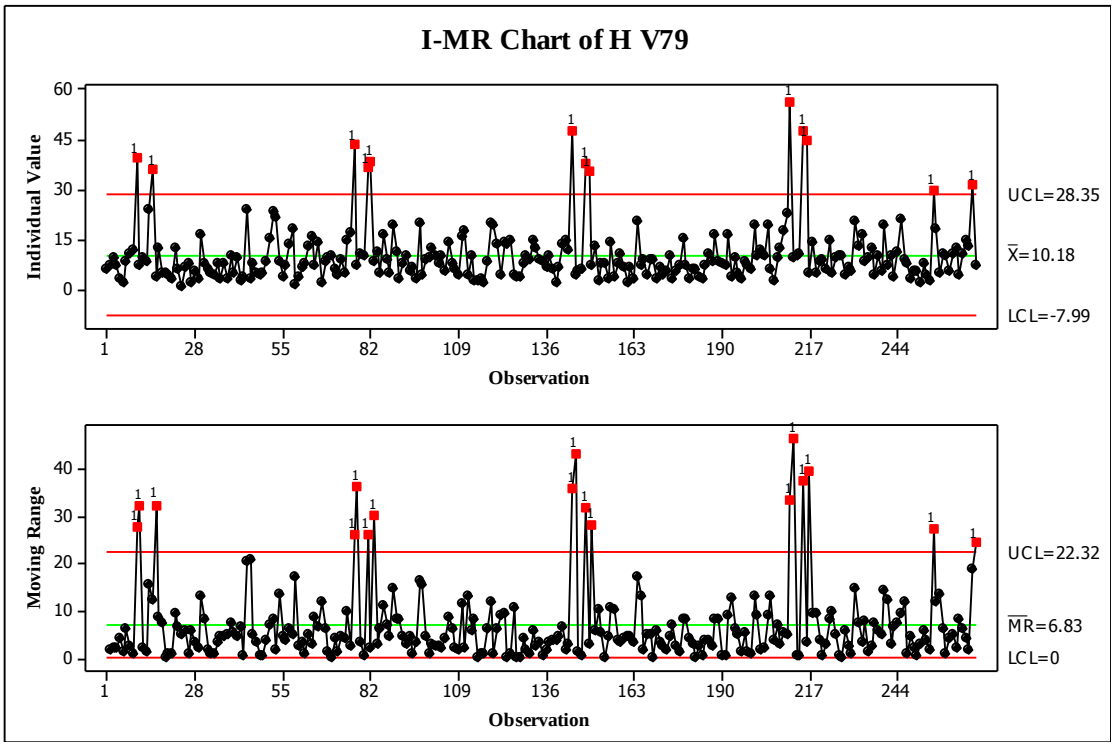

I Chart of E HPRT
Individual Value
Observation
UCL=17.40
X=6.74
LCL=-3.92

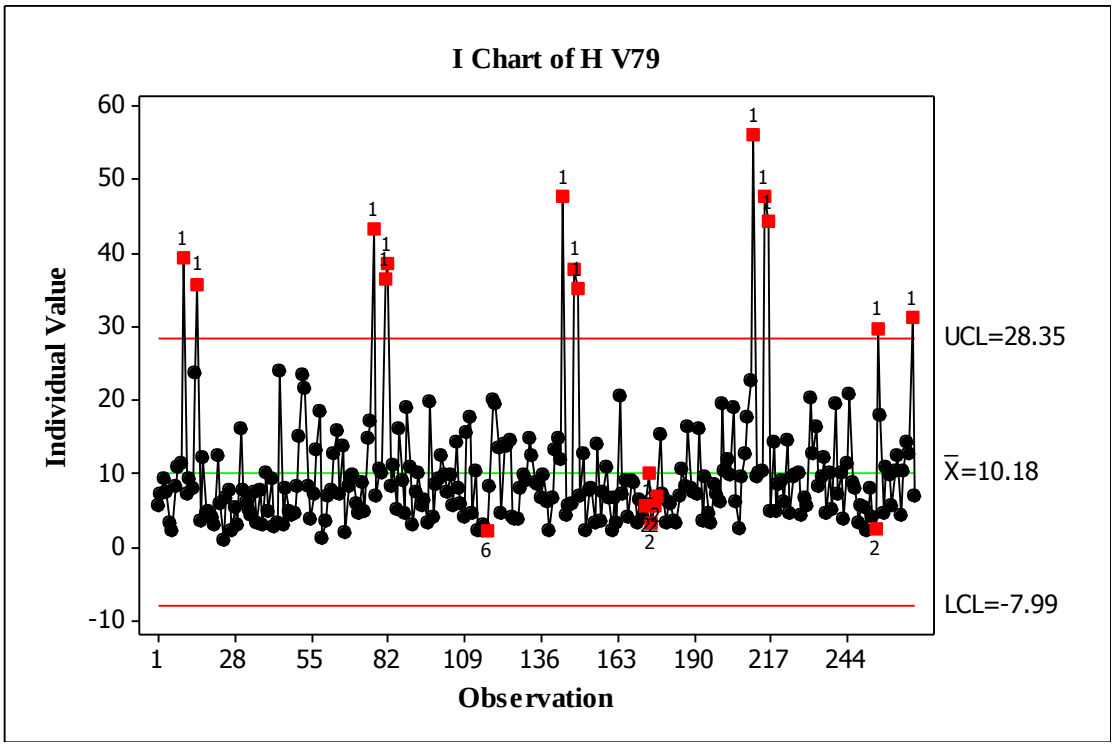

I-MR Chart of F HPRT
Individual Value
Moving Range
Observation
UCL=10.63
X=4.13
LCL=-2.37
UCL=7.98
MR=2.44
LCL=0

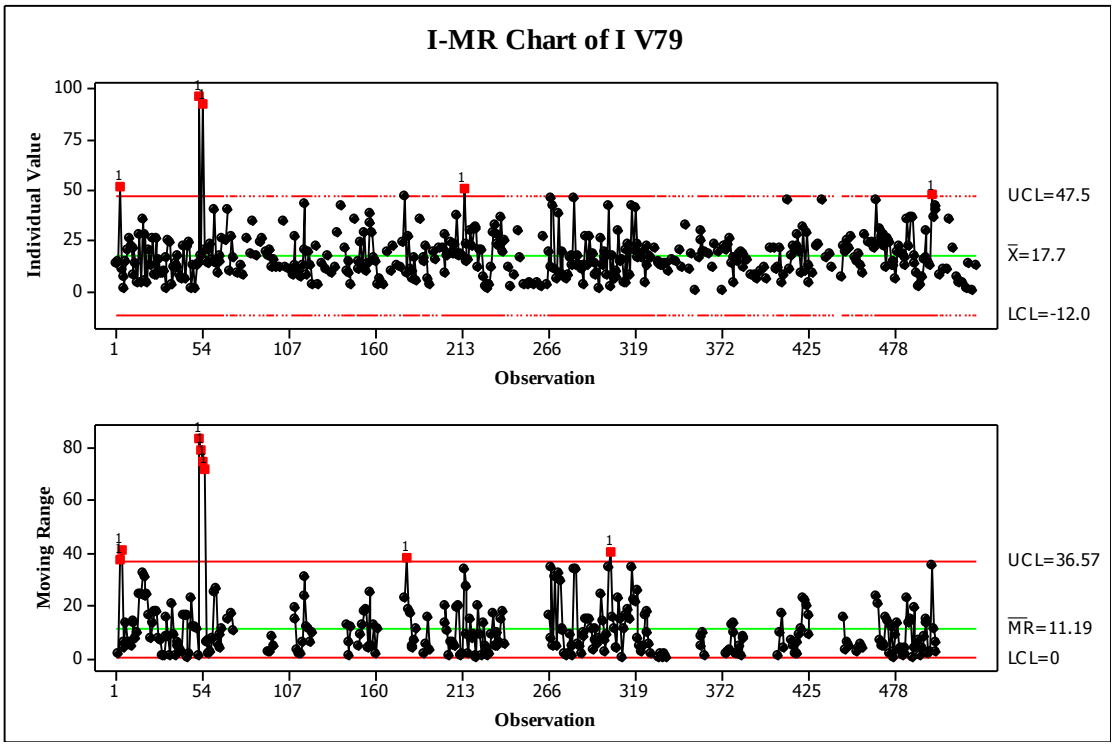

I Chart of F HPRT
Individual Value
Observation
UCL=10.63
X=4.13
LCL=-2.37

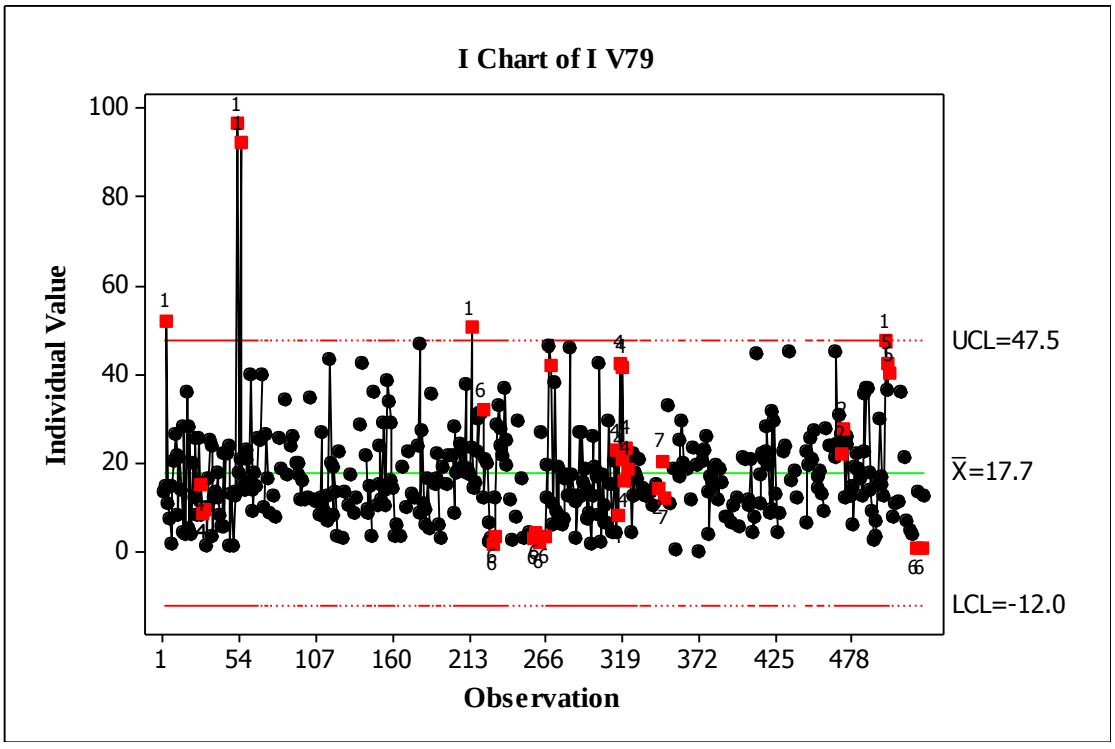

I-MR Chart of G V79
Individual Value
UCL=42.55
X̄=17.91
LCL=-6.74
Observation
Moving Range
UCL=30.27
MR̄=9.27
LCL=0
Observation

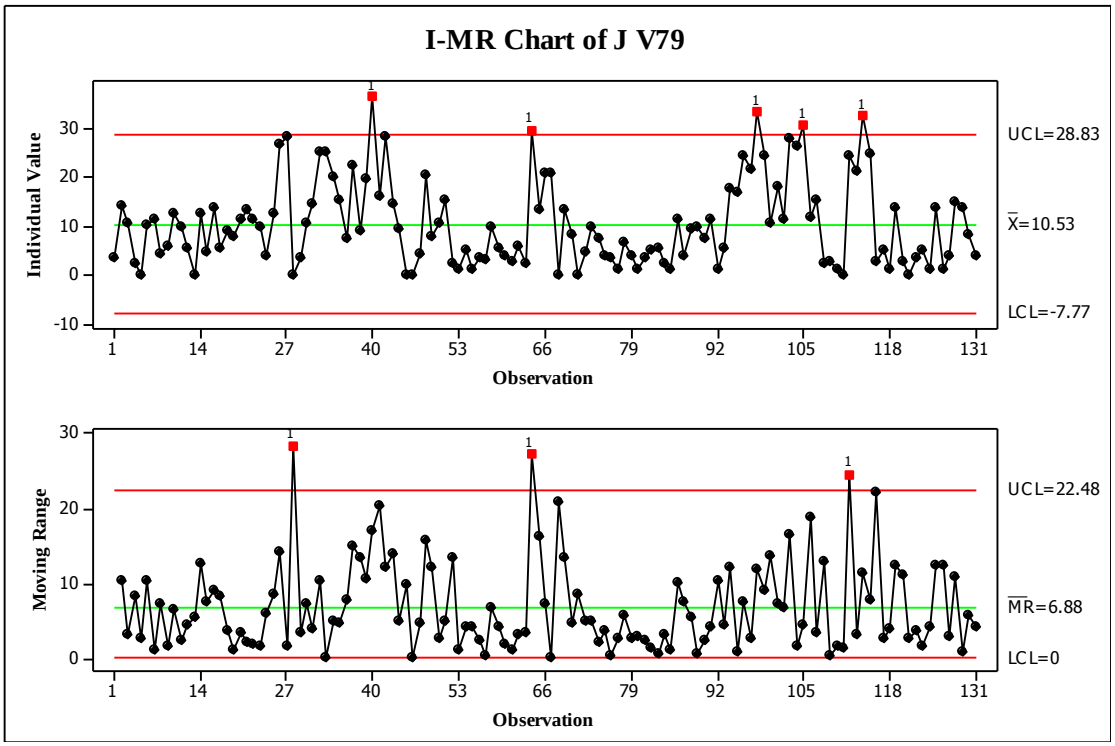

I Chart of G V79
Individual Value
UCL=42.55
X̄=17.91
LCL=-6.74
Observation

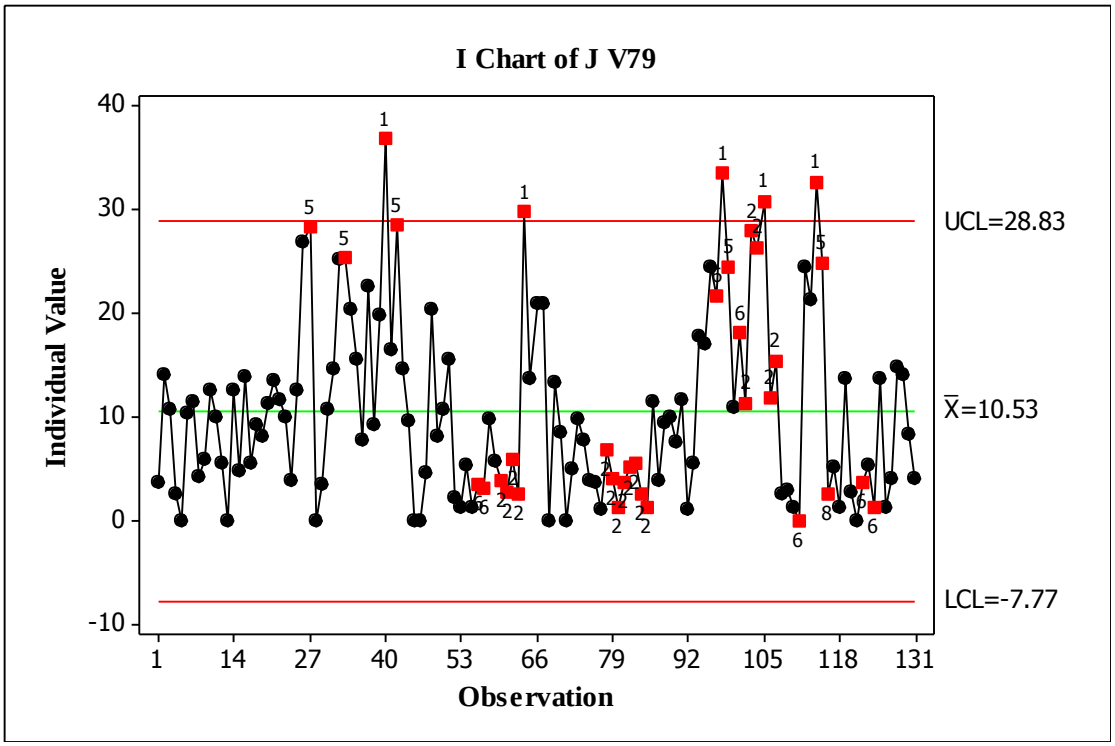

I-MR Chart of H V79
Individual Value
60
45
30
15
0
1
28
55
82
109
136
163
190
217
244
Observation
UCL=28.35
X=10.18
LCL=-7.99
Moving Range
40
30
20
10
0
UCL=22.32
MR=6.83
LCL=0
Observation

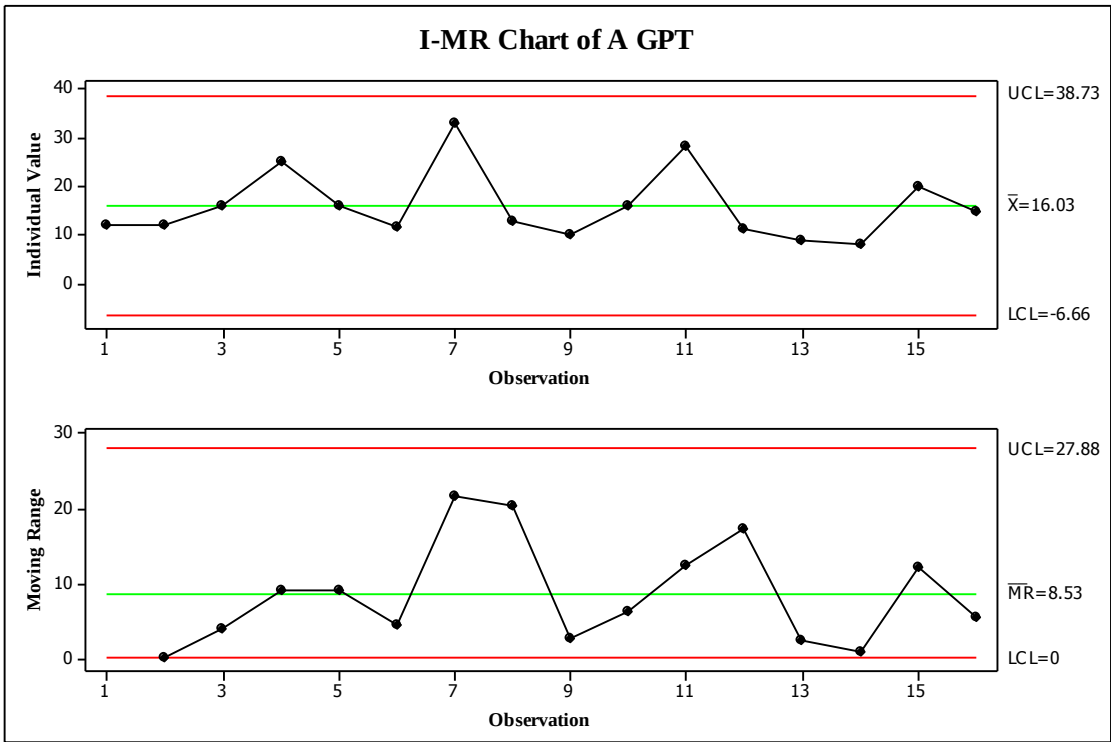

I Chart of H V79
Individual Value
60
50
40
30
20
10
0
-10
1
28
55
82
109
136
163
190
217
244
Observation
UCL=28.35
X=10.18
LCL=-7.99

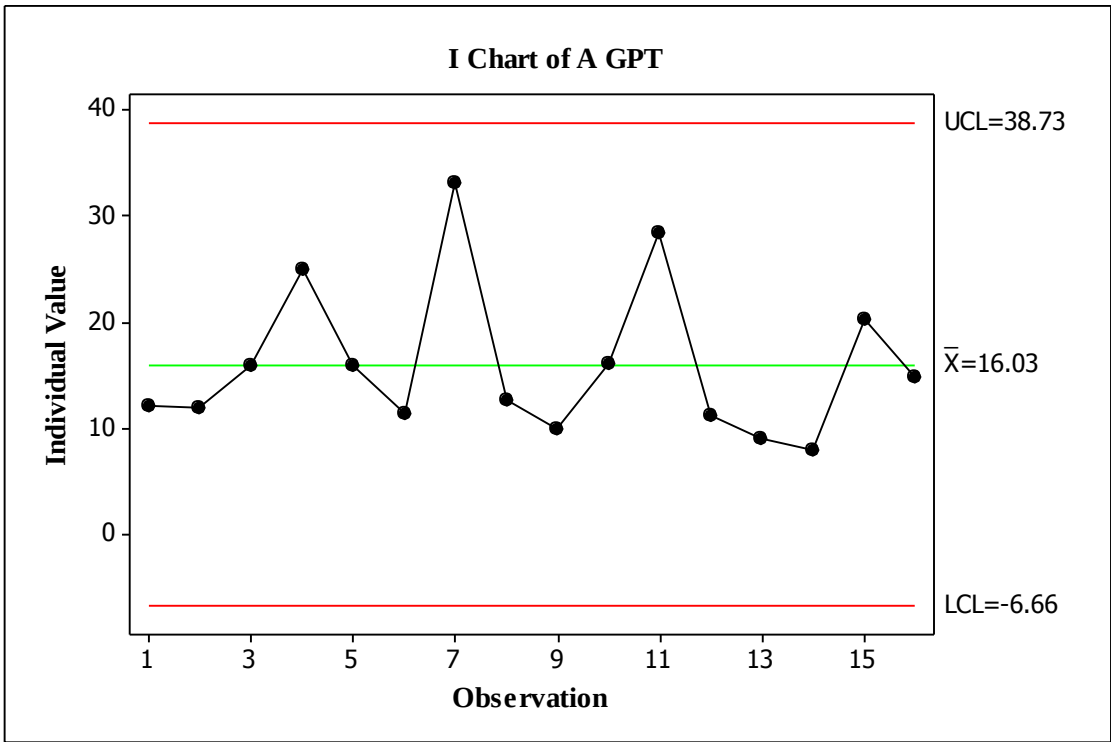

I-MR Chart of I V79
Individual Value
100
75
50
25
0
UCL=47.5
X=17.7
LCL=-12.0
1
54
107
160
213
266
319
372
425
478
Observation
Moving Range
80
60
40
20
0
UCL=36.57
MR=11.19
LCL=0
Observation

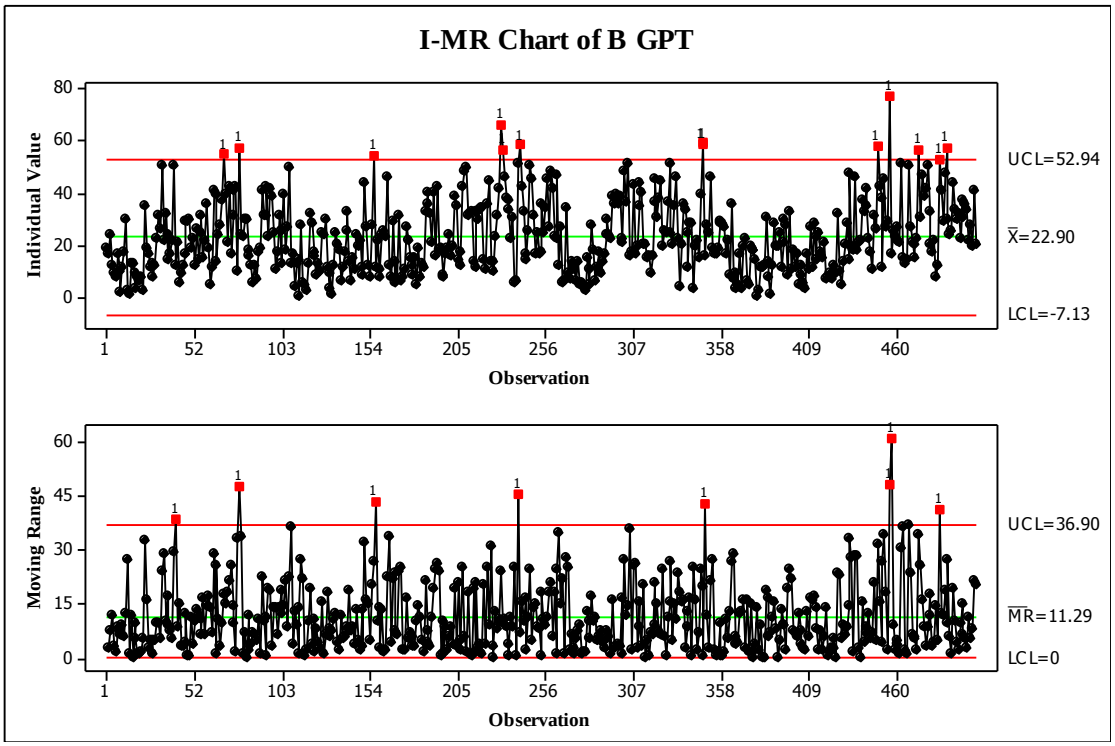

I Chart of I V79
Individual Value
100
80
60
40
20
0
UCL=47.5
X=17.7
LCL=-12.0
1
54
107
160
213
266
319
372
425
478
Observation

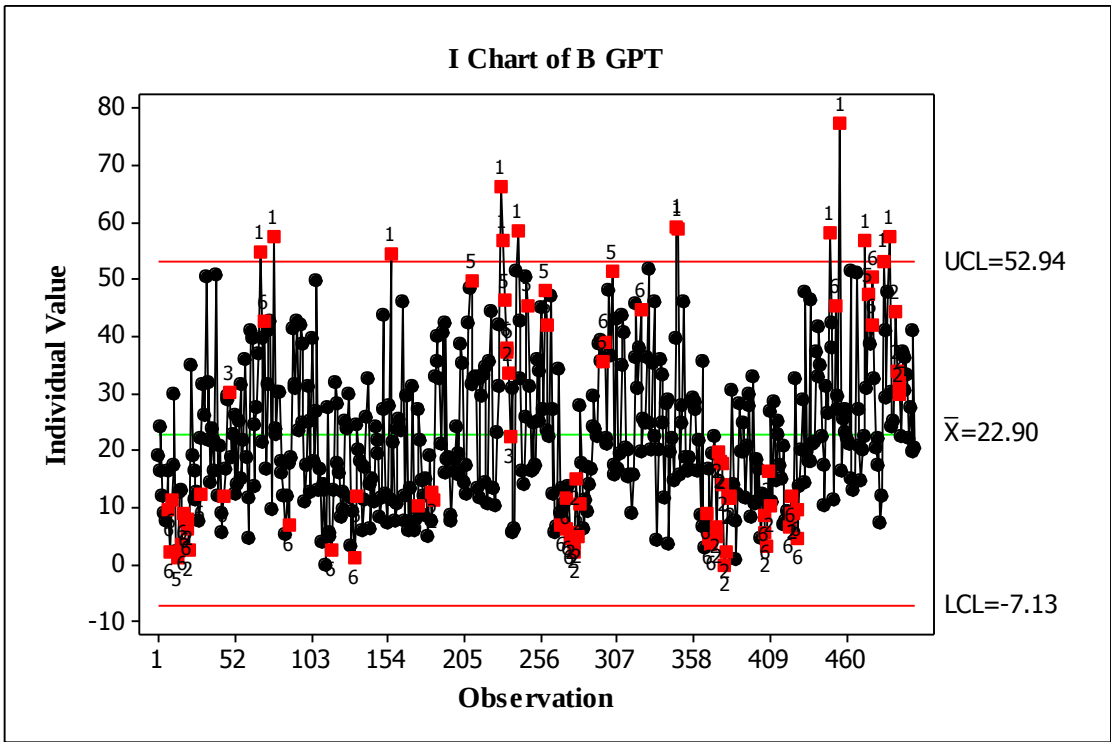

I-MR Chart of J V79
Individual Value
UCL=28.83
X̄=10.53
LCL=-7.77
Observation
Moving Range
UCL=22.48
MR̄=6.88
LCL=0
Observation

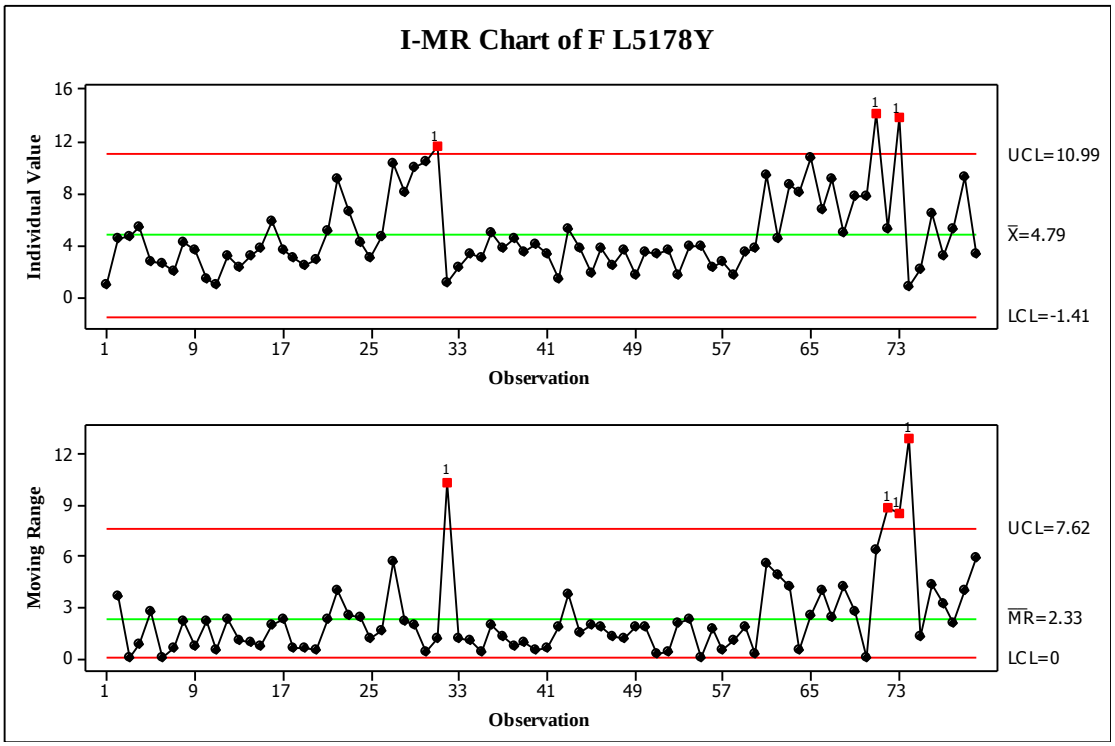

I Chart of J V79
Individual Value
UCL=28.83
X̄=10.53
LCL=-7.77
Observation

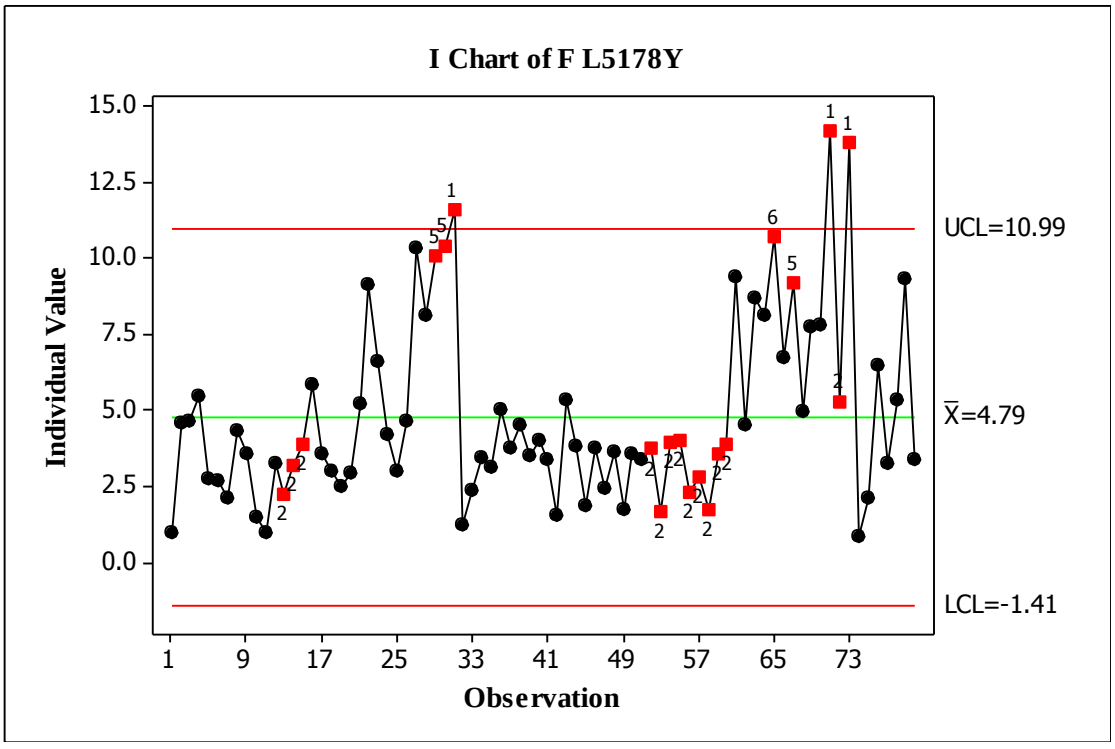

I-MR Chart of A GPT
Individual Value
40
30
20
10
0
UCL=38.73
X=16.03
LCL=-6.66
1
3
5
7
9
11
13
15
Observation
Moving Range
30
20
10
0
UCL=27.88
MR=8.53
LCL=0
1
3
5
7
9
11
13
15
Observation

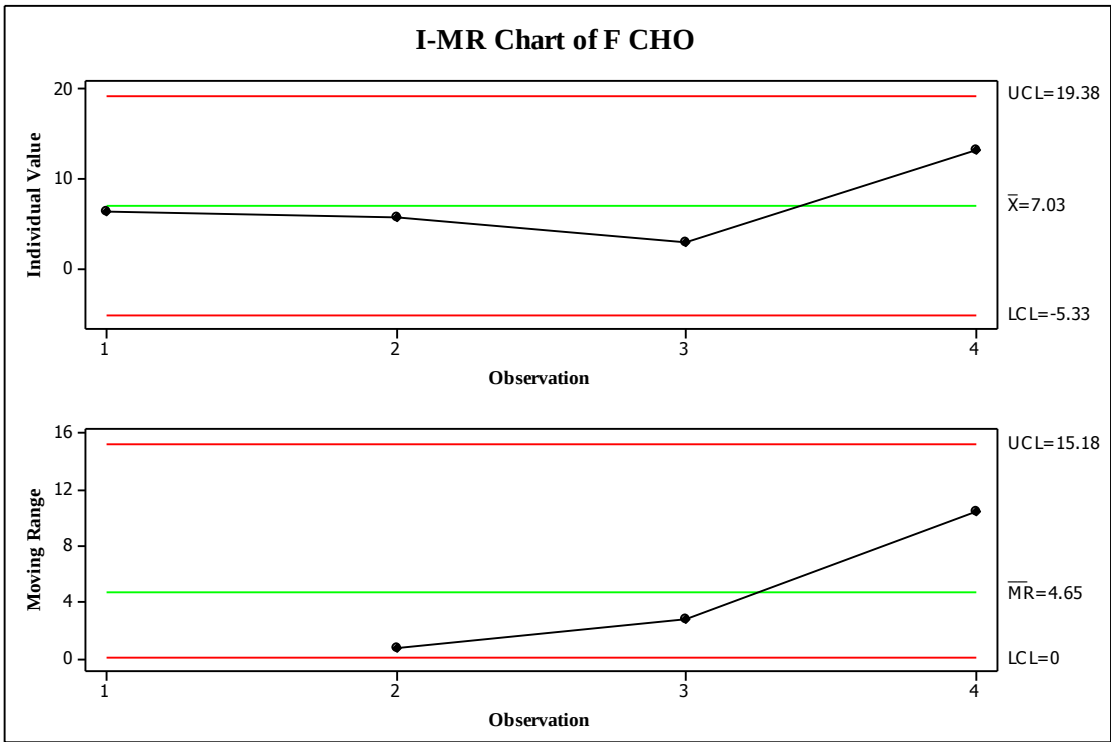

I Chart of A GPT
Individual Value
40
30
20
10
0
UCL=38.73
X=16.03
LCL=-6.66
1
3
5
7
9
11
13
15
Observation

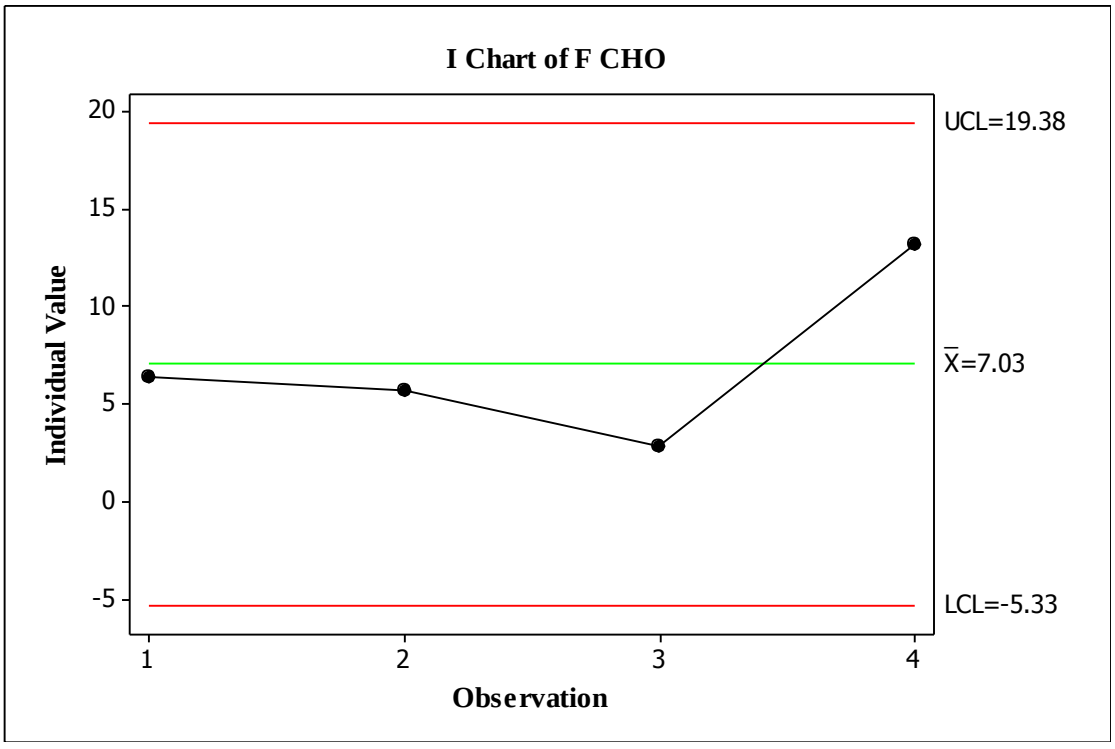

I-MR Chart of B GPT
Individual Value
Moving Range
Observation
UCL=52.94
X=22.90
LCL=-7.13
UCL=36.90
MR=11.29
LCL=0


I Chart of B GPT
Individual Value
Observation
UCL=52.94
X=22.90
LCL=-7.13


I-MR Chart of F L5178Y
Individual Value
Moving Range
Observation
UCL=10.99
X̄=4.79
LCL=-1.41
UCL=7.62
MR̄=2.33
LCL=0


I Chart of F L5178Y
Individual Value
Observation
UCL=10.99
X̄=4.79
LCL=-1.41


I-MR Chart of F CHO
Individual Value
20
10
0
UCL=19.38
X=7.03
LCL=-5.33
1
2
3
4
Observation
Moving Range
16
12
8
4
0
UCL=15.18
MR=4.65
LCL=0
1
2
3
4
Observation


I Chart of F CHO
Individual Value
20
15
10
5
0
-5
UCL=19.38
X=7.03
LCL=-5.33
1
2
3
4
Observation