

Interpreting pharmaceutical screening test results

Problem presented by

Richard Lyons, Rachel Russell

Pfizer Worldwide R&D

Executive Summary

In the pharmaceutical industry, compounds are tested for activity against a disease target. In a High Throughput Assay (HTA), each compound is tested at a single concentration. In an IC₅₀ test, a compound is tested at a range of concentrations. Sometimes there are discrepancies between the results of these tests and the Study Group was asked to model this. Rather than assume normally distributed errors in the percentage effect, the Study Group proposed a model in which there are also probabilities p and q of a test erroneously indicating very low or high activity. The parameters p and q can be estimated from the data, and then can be used in the design of testing procedures.

Version 1.1
May 27, 2011
iii+20 pages

Report author

David Allwright (Industrial Maths KTN)

Contributors

Joanna Giemza (University of Warsaw)
Paul Munday (Industrial Maths KTN)
Colin Please (University of Southampton)
Pawel Wilk (University of Warsaw)
Anatoly Zhigljavsky (University of Cardiff)

ESGI80 was jointly organised by

The University of Cardiff

The Knowledge Transfer Network for Industrial Mathematics

and was supported by

The Engineering and Physical Sciences Research Council

Contents

1	Introduction	1
1.1	Background and scope	1
2	Data	1
2.1	General description	1
2.2	HTA data	2
2.3	IC50 data	8
2.4	Comparison of HTA with IC50	8
3	Probabilistic error model	14
3.1	p,q -model	14
3.2	Estimating p and q	18
3.3	Use of this model	19

1 Introduction

1.1 Background and scope

- (1.1.1) In this report we shall first describe the available data in Section 2, and then propose a probabilistic model for the errors in these experiments in Section ???. Then in Section 3.2 we shall show how the parameters in such a model could be estimated from the data, and in Section 3.3 how such a model could be used to improve the screening efficiency.

2 Data

2.1 General description

- (2.1.1) Pfizer supplied 10 data files in `csv` format, each with information from screening tests performed against a particular target. (Actually 11 were supplied but `Prog_2011-03-09.csv` is identical to `Prog_antag_2011-03-10.csv`, and we shall refer to this data set by the latter name.) For brevity we sometimes refer to datasets simply by the prefix, *e.g.* `Prog_antag`, omitting the date. The files vary greatly in length: in Table 1 the first 2 columns are the file names and the number of lines in the file, and the other columns will be explained later.

Table 1: Summary information about data files.

File name	Numbers of		Standard compound	
	lines	compounds	nominal	commonest
<code>Prog_antag</code>	6843	95	PF-02327888-00	PF-04723078-00
<code>GABA</code>	11721	431	PF-01386346-00	PF-04459758-00
<code>iJAK</code>	35498	1122	PF-02384554-00	PF-02563601-00
<code>Integrase</code>	6168	146	PF-04269440	PF-04269440-00
<code>TrpM8</code>	124763	2466	PF-04674115-00	PF-04674115-00
<code>TrpA1antag</code>	169599	2200	PF-04667541-00	PF-04840154-00
<code>Nav1.7</code>	138238	3966	PF-00344649-01	PF-00344649-01
<code>Nav1.8</code>	188342	1831	PF-00344649-01	UK-483608-01
<code>VIF</code>	13234	160	PF-04854474-00	PF-04854474-00
<code>TrpA1_agonist</code>	190309	2198	PF-04840154-00	PF-04840154-00

Each screening test is for a particular compound at a particular concentration, and is measured by a percentage effect (PCE). This PCE is a normalized result obtained by mapping the experimental response linearly to a dimensionless number with the response to one control compound and concentration, called the minimum, mapped to 0, and the response to another control, called the maximum, mapped to 100. The third column in Table 1 is the number of compounds in each file.

- (2.1.2) In general there are 2 kinds of screening trial. In one kind, called a high throughput assay (HTA), a large number of compounds are tested, each at a single high concentration. In the other kind, called an IC50 series, a smaller number of compounds are tested, each at 10 or 11 concentrations, typically spaced geometrically over 5 orders of magnitude.
- (2.1.3) The data from individual experiments is controlled for quality in various ways to ensure that nothing has gone wrong with the experiment. The data supplied to us was selected as being cases where both HTA and IC50 data are available. That in itself is a filtering process, in that compounds are only tested in an IC50 test if they either have shown some kind of activity or are suspected, on pharmacological grounds, of being more likely to be active.

2.2 HTA data

- (2.2.1) The HTA tests generally involve testing a compound twice at the same concentration. Since we are interested in the repeatability and errors in experiments, we plotted for each file the percentage effects for each compound for which two percentage effects at the same concentration are available.
- (2.2.2) The results for the `Prog_antag` data are shown in Figure 1. It is clear that the HTA1 results are all at least about 50, and this is perhaps an example of the filtering that may have occurred.
- (2.2.3) It is also noticeable that some of the “percentage effect” values are greater than 100%: this simply reflects the fact that the controls used are not always the actual maximum and minimum possible. This occurs throughout, with PCE values being both greater than 100 and less than 0: it is simply a reminder that this normalization has taken place.
- (2.2.4) The corresponding plot for the `GABA` file is in Figure 2. This shows a spread of points along the diagonal, with larger errors in PCE when the values are mid-range than when they are at the top or bottom. But it also shows a cluster of points in the lower left corner, and another in the upper right. So these are cases where a compound is tested twice at the same concentration and one of the tests shows an activity close to 0 and the other close to 100.
- (2.2.5) The corresponding plots for each of the other files are also shown, in Figures 3–10. In general they show varying degrees of association between the different results. In the `Nav1.7` data, it looks as if some filtering partly based on whether the sum of the two HTA PCEs exceeds 60 has been applied.

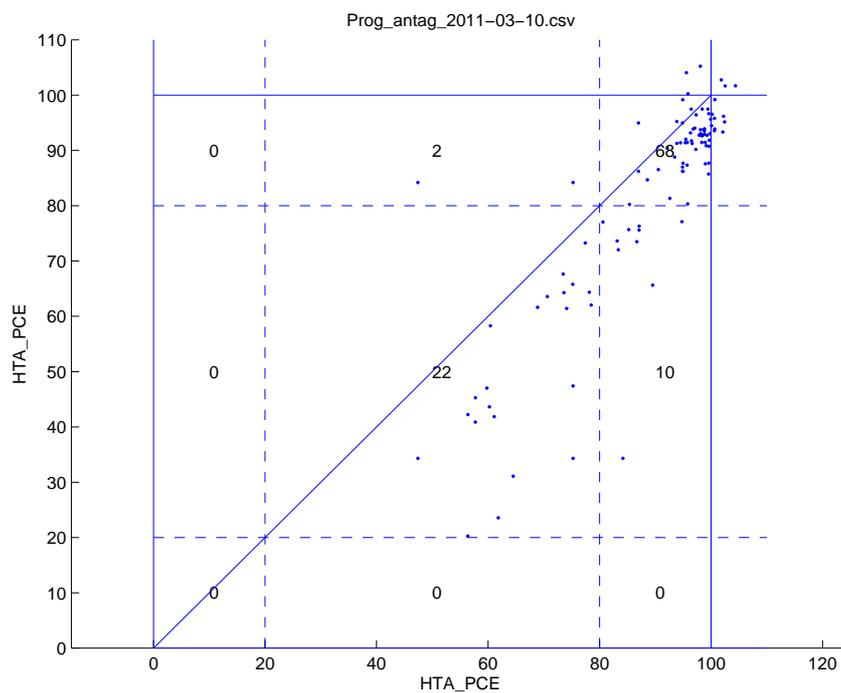


Figure 1: Prog_antag. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

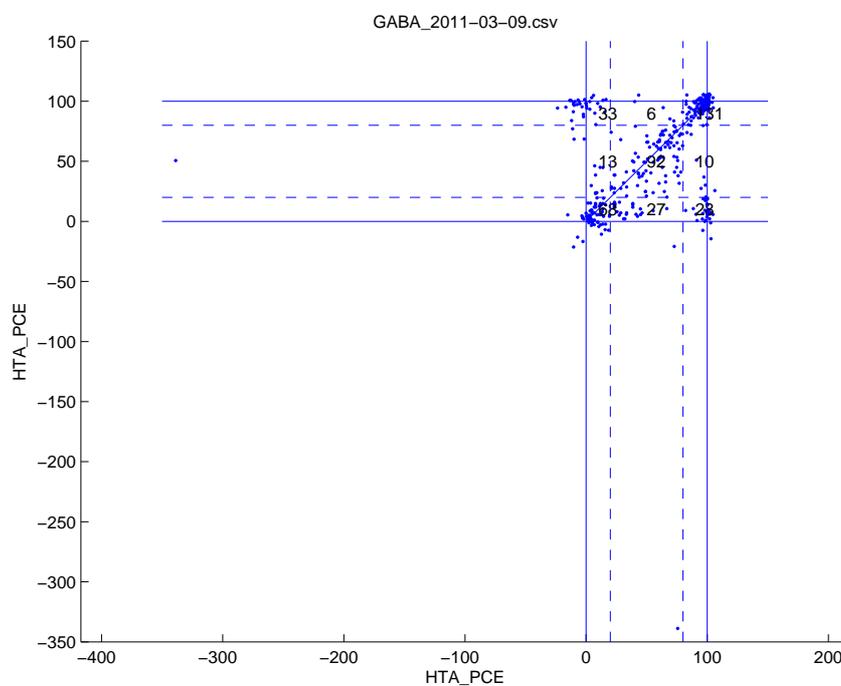


Figure 2: GABA. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

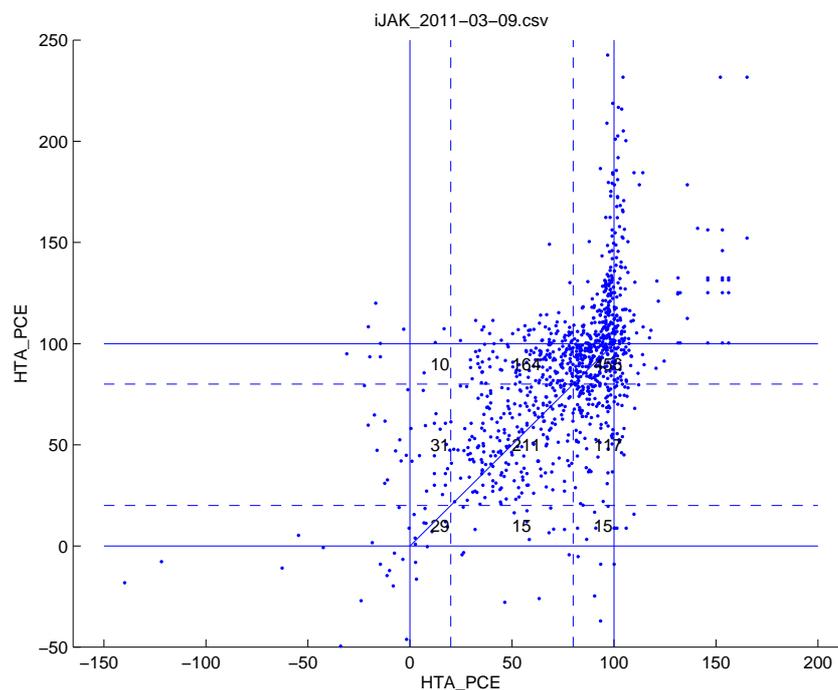


Figure 3: iJAK. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

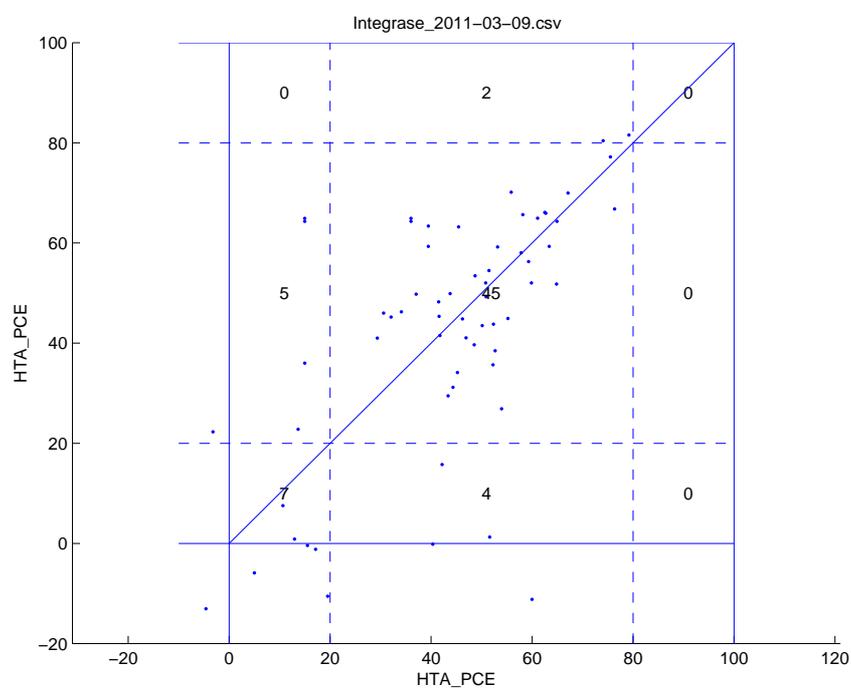


Figure 4: Integrase. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

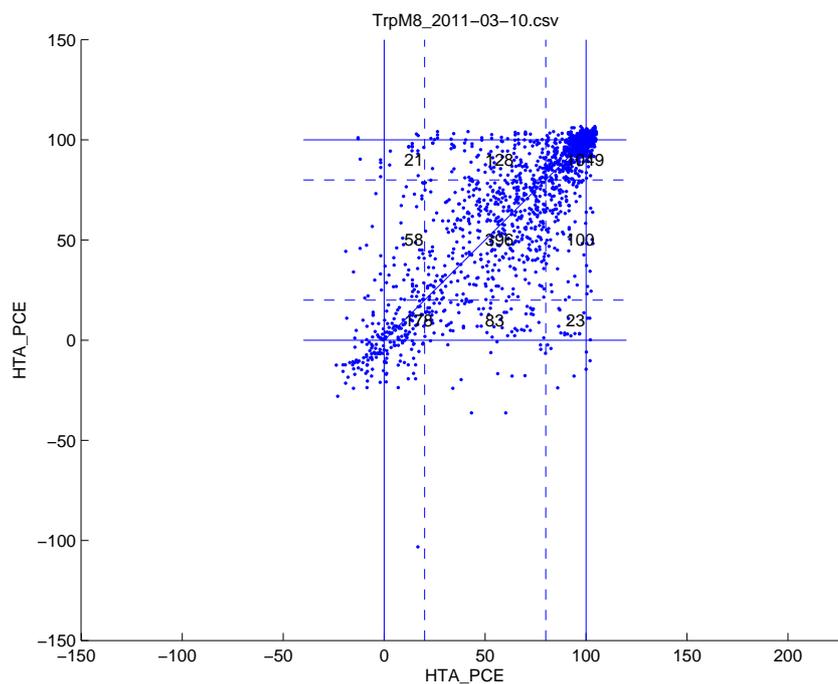


Figure 5: TrpM8. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

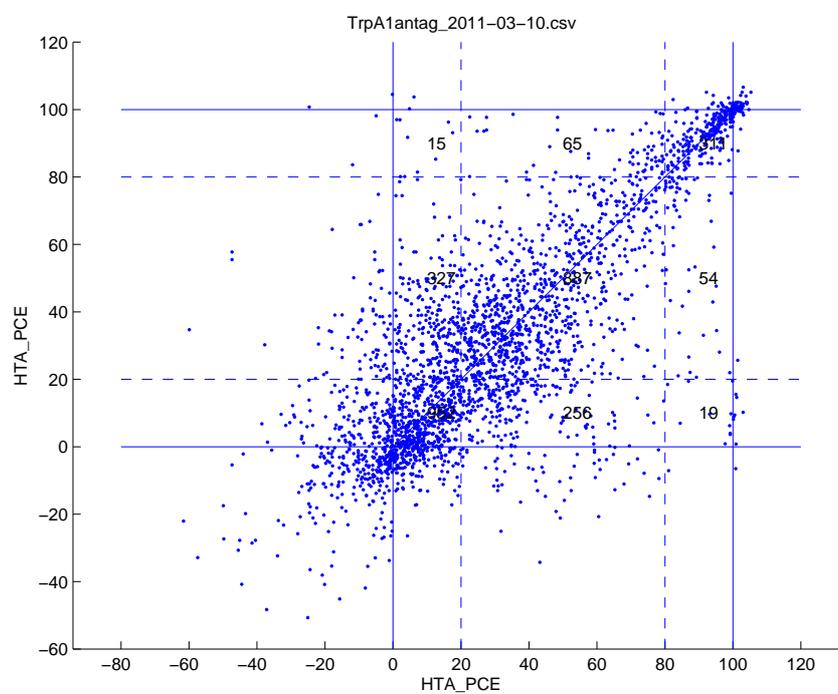


Figure 6: TrpA1antag. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

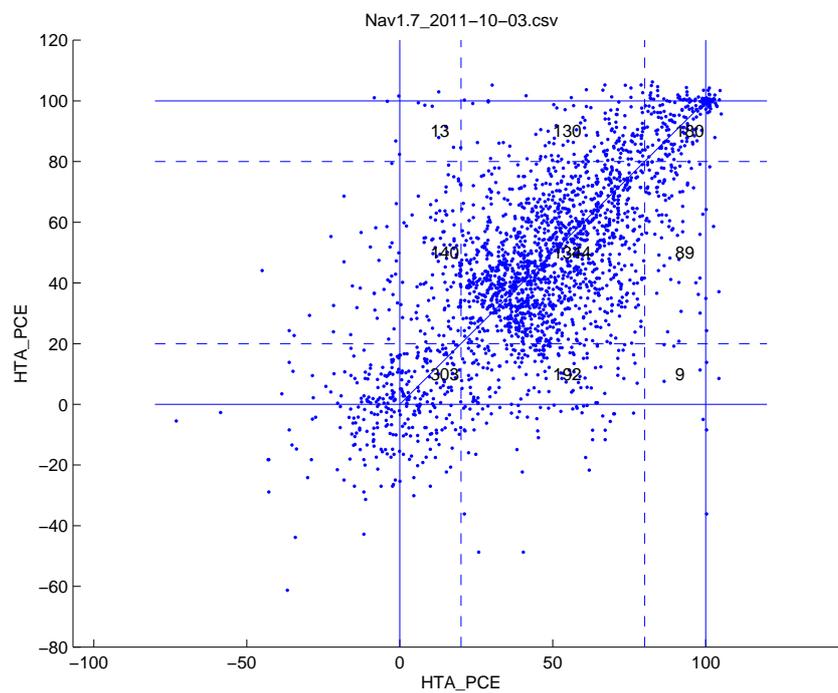


Figure 7: Nav1.7. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

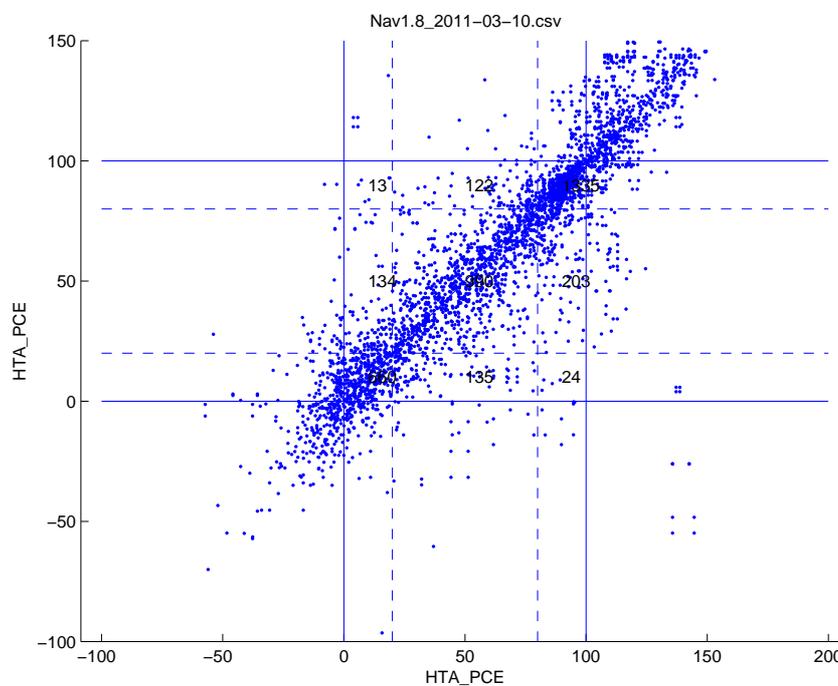


Figure 8: Nav1.8. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

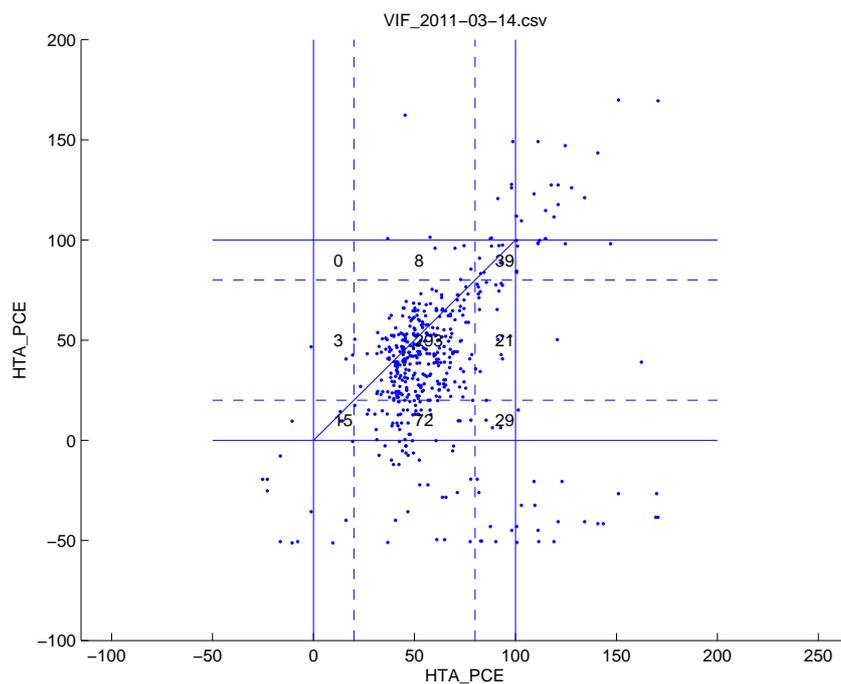


Figure 9: VIF. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

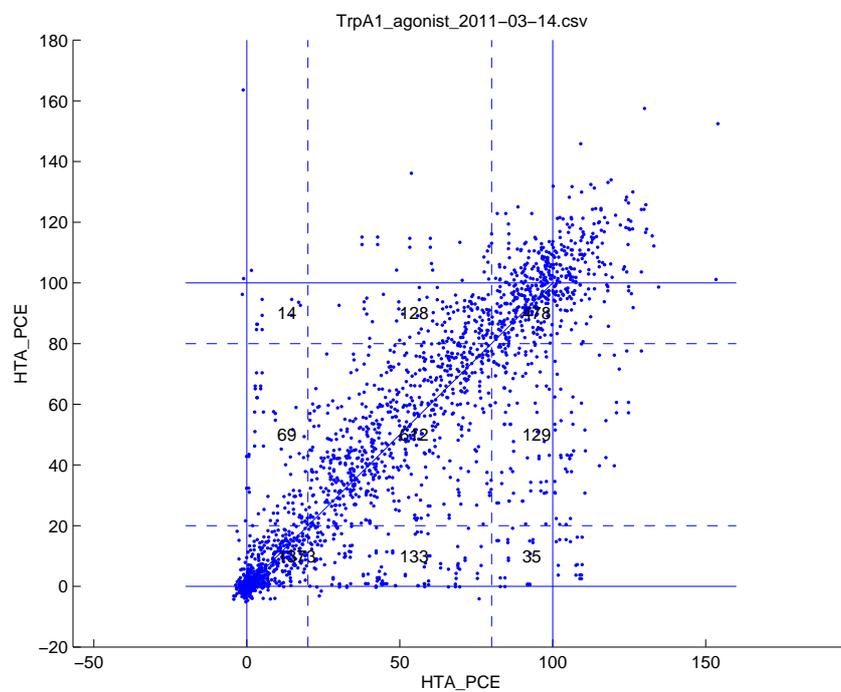


Figure 10: TrpA1_agonist. Percentage effects in 2 HTA tests at the same concentration. Each point represents a different compound.

2.3 IC50 data

- (2.3.1) The IC50 data is subject to more quality controls than the HTA data. In particular a “standard” compound is used on each plate, and if the results for the standard do not produce a satisfactory dose-response curve, then all the values from that plate are discarded. The fact that the standard compound is tested multiple times and on different plates is of course a good source of information about experimental error.
- (2.3.2) The identifier of the standard compound in each file was given to us by Pfizer, but was not always the identifier of the compound that occurred most often in the data. When this happened we in fact took the standard to be the compound whose identifier had the most IC50 points. In Table 1 the final 2 columns give the nominal standard compound, and then the compound that had the most IC50 data points.
- (2.3.3) When the dose-response curves were plotted for each file, there were clearly anomalies. From discussion with Pfizer, we learned that in the IC50 PCE values for `Prog_antag` and `Integrase`, x should be replaced by $100 - x$.
- (2.3.4) From further looking at the data it appeared that this also applied to the `Nav1.8` data, and that it was applied only *partly* to the `Integrase` data: we did *not* replace the IC50 PCE values in lines 702–1281 of the `csv` file nor lines 2372–2391 (which are both within the data for the standard compound in those tests.) Ideally, we would like confirmation from Pfizer that these choices are correct, and advice on any similar data-cleaning that may need to be applied.
- (2.3.5) The resulting plots of dose-response for the standard compound in each data file are presented in Figures 11–20. In each of these, the horizontal axis is the base-10 logarithm of the concentration. If the dose-response model is taken as
- $$y = d - \frac{a - d}{1 + (x/c)^b} \quad (1)$$
- then when y is plotted against $\log x$ a tanh profile results (translated and scaled suitably) and this is visible in all cases.
- (2.3.6) Clearly there is a wide variety of different kinds of error, but the general form is more-or-less visible throughout. For the `VIF` data the saturation to 100% has not been reached even at the highest concentration used.

2.4 Comparison of HTA with IC50

- (2.4.1) The percentage effect can also be compared between the HTA tests and the IC50 tests if one of the IC50 concentrations is at, or close to, the HTA concentration. This generally does happen and therefore allows us to produce scatter plots of the IC50 PCE against the HTA PCE for each file.

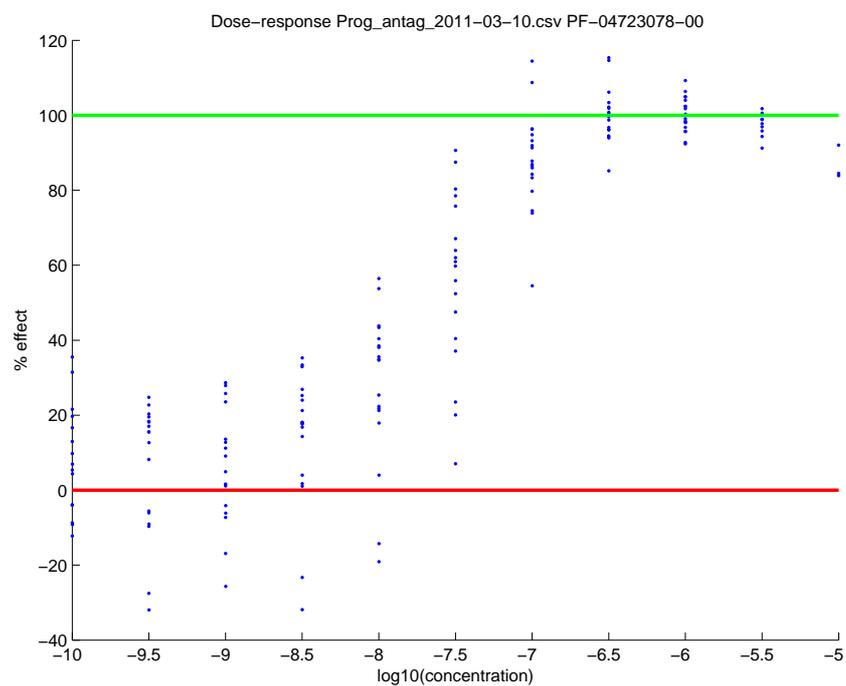


Figure 11: Prog_antag. IC50 test results for the standard compound.

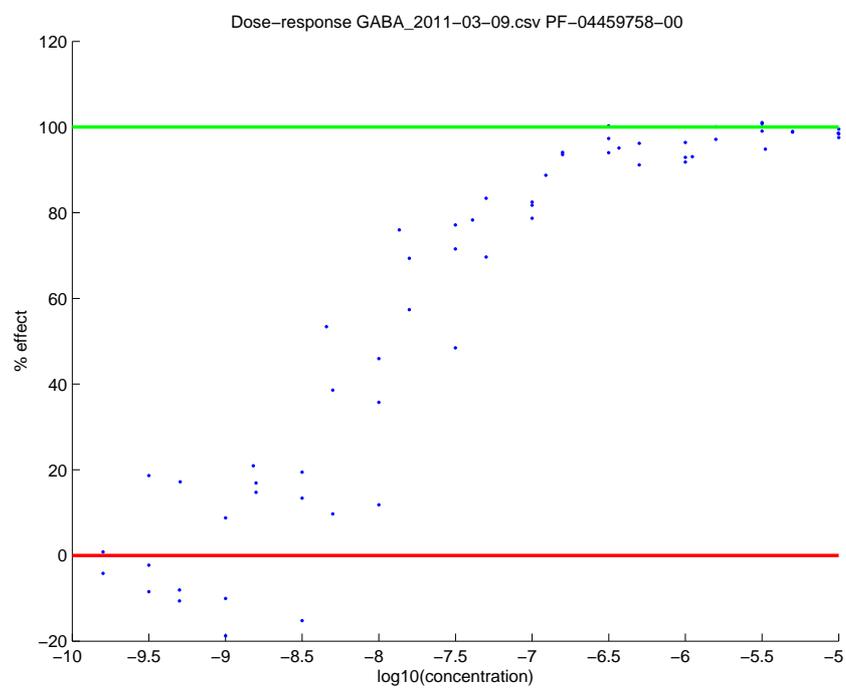


Figure 12: GABA. IC50 test results for the standard compound.

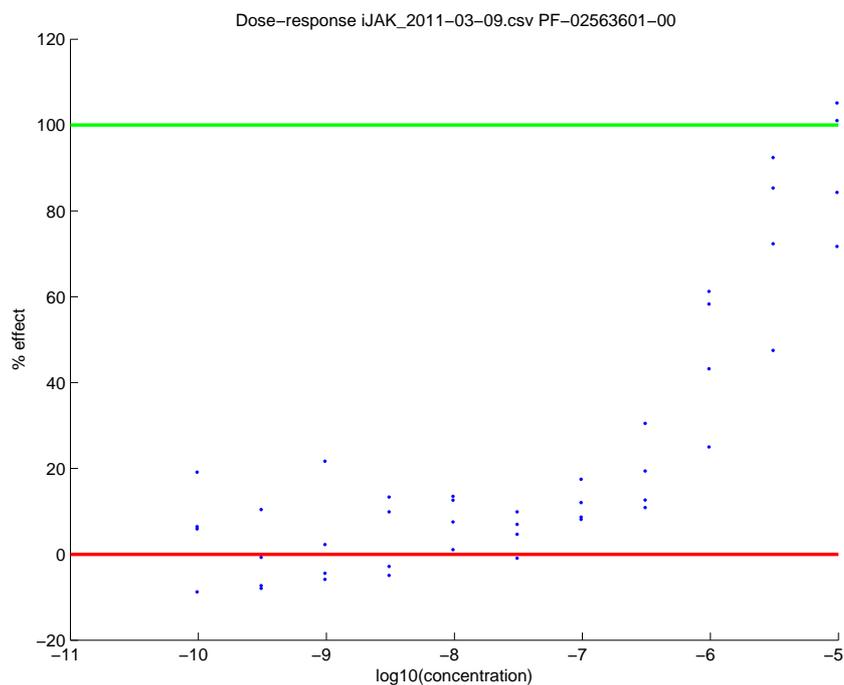


Figure 13: iJAK. IC50 test results for the standard compound.

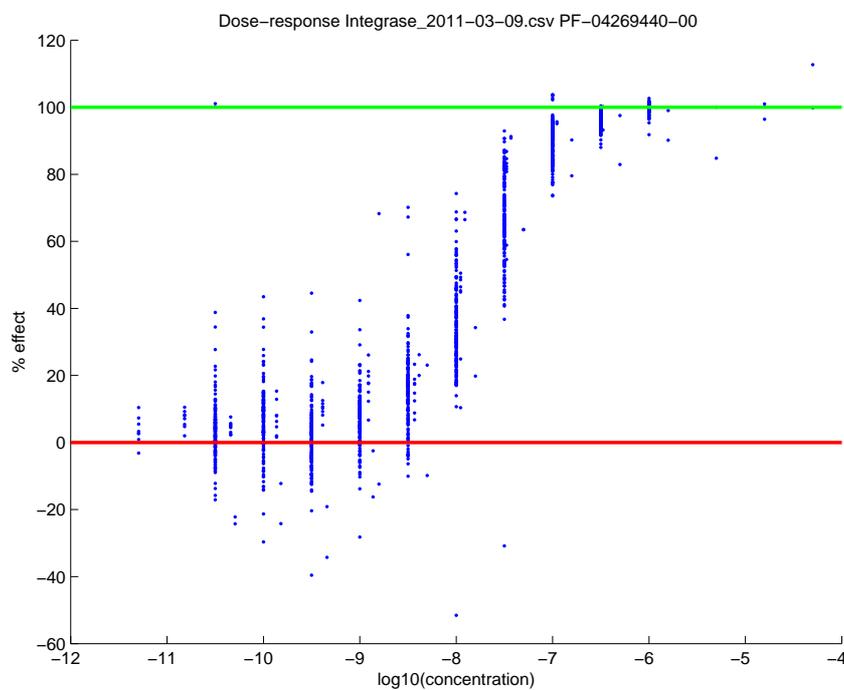


Figure 14: Integrase. IC50 test results for the standard compound.

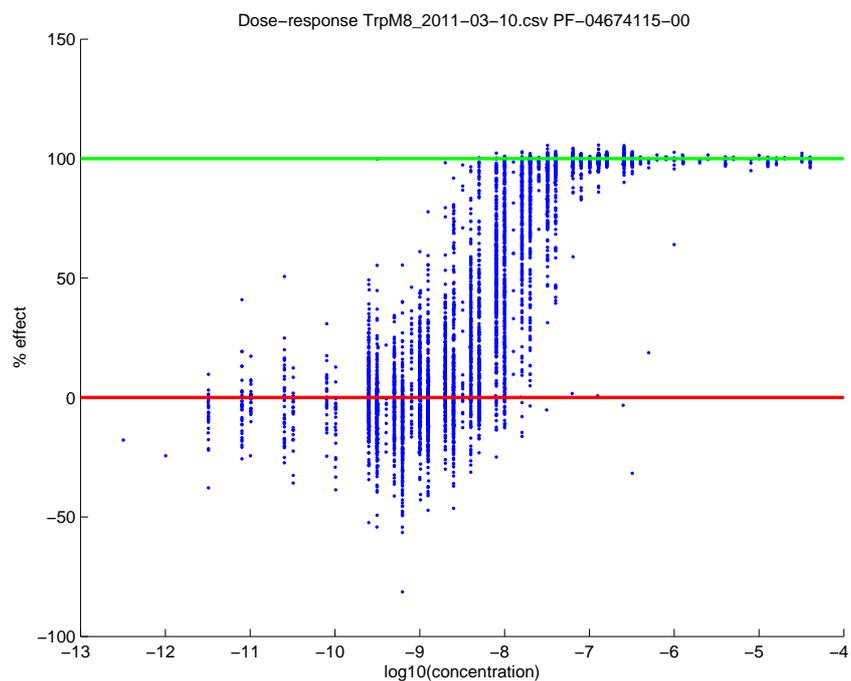


Figure 15: TrpM8. IC50 test results for the standard compound.

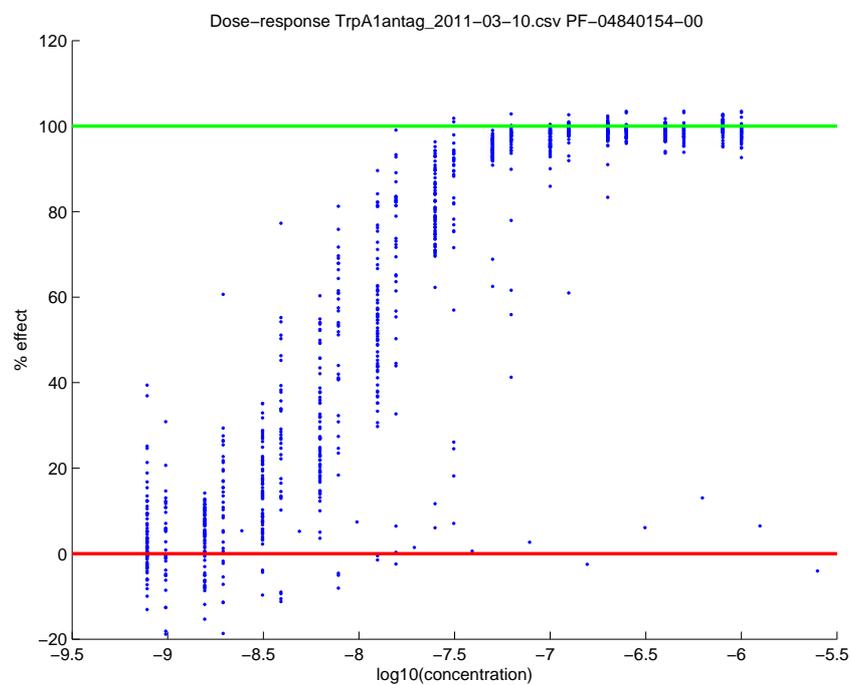


Figure 16: TrpA1antag. IC50 test results for the standard compound.

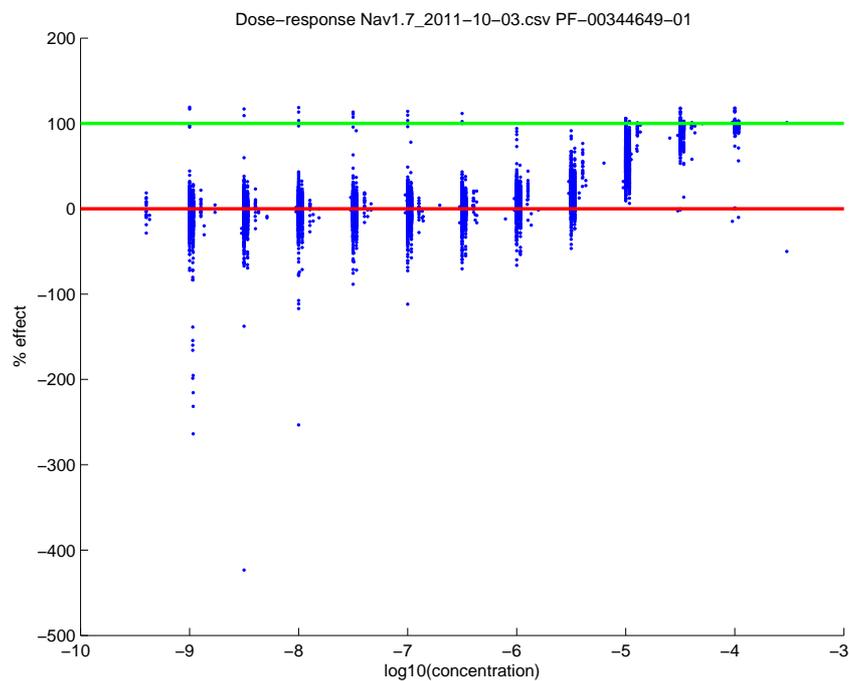


Figure 17: Nav1.7. IC50 test results for the standard compound.

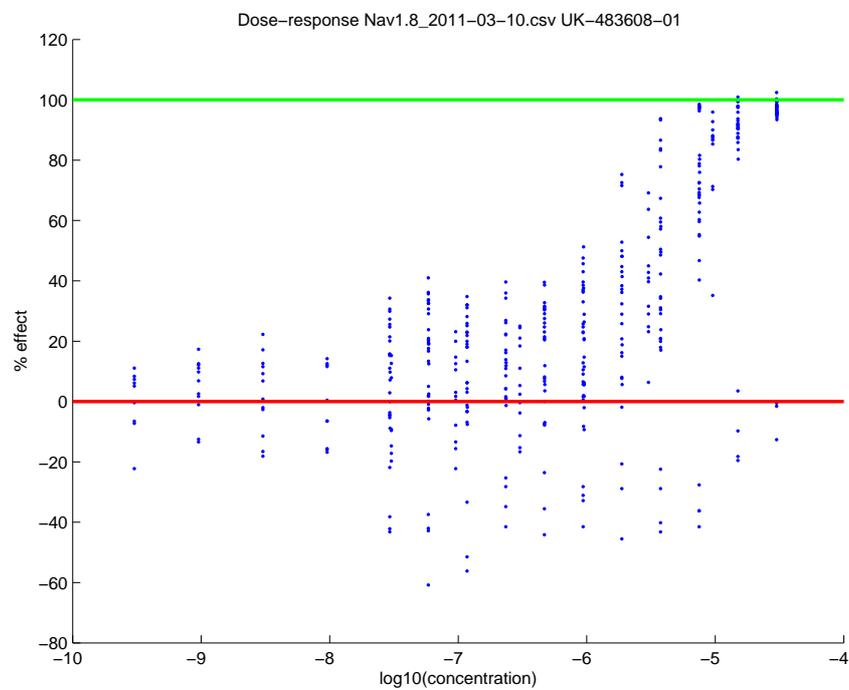


Figure 18: Nav1.8. IC50 test results for the standard compound.

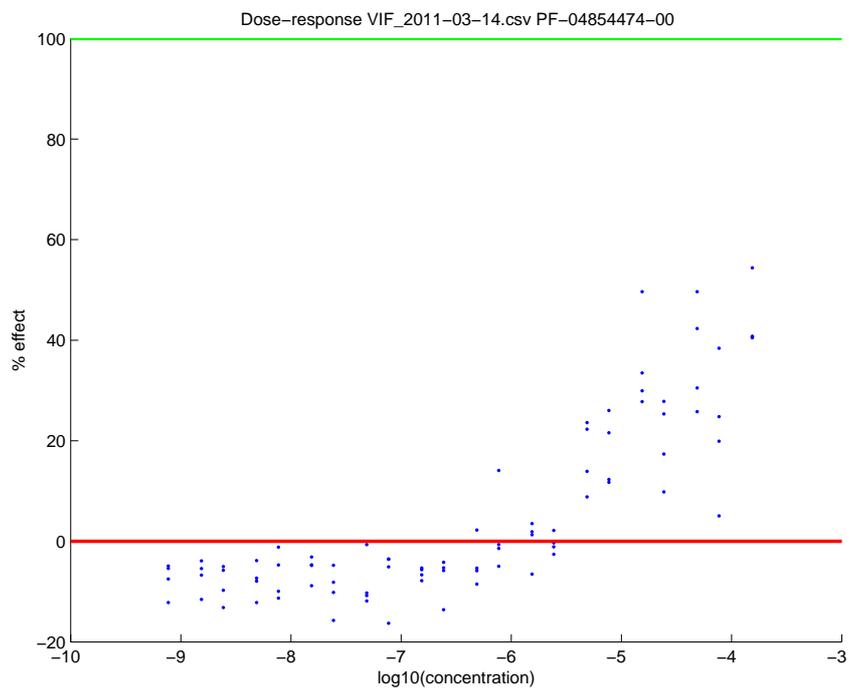


Figure 19: VIF. IC₅₀ test results for the standard compound.

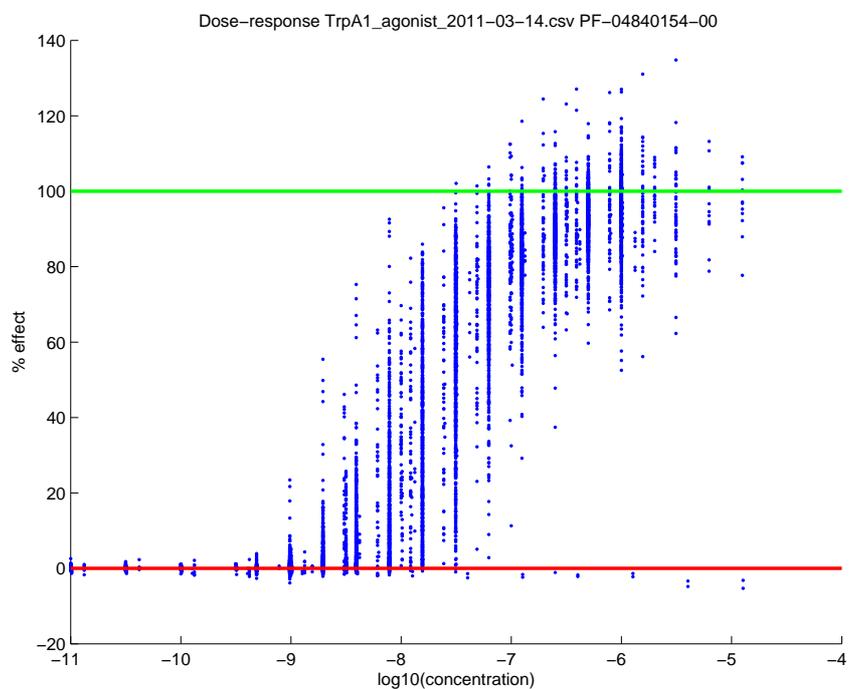


Figure 20: TrpA1_agonist. IC₅₀ test results for the standard compound.

In these plots, the red points are for compounds where the HTA concentration and IC50 concentration agreed exactly (to within the square root of machine precision), the green points are where they agreed within 1%, and the yellow within 10%. These ratios were evaluated logarithmically, *i.e.* we plot yellow if $0.01 < |\log_e(x_1/x_2)| < 0.1$.

- (2.4.2) These plots show many features related to what we have seen before. Figure 28 in particular shows a strong tendency for the points to fall below the diagonal, but without falling particularly into the lower right corner. This is most naturally interpreted as compounds whose HTA effect was large due to measurement noise, and which then gave a lower but truer value in the IC50 tests.

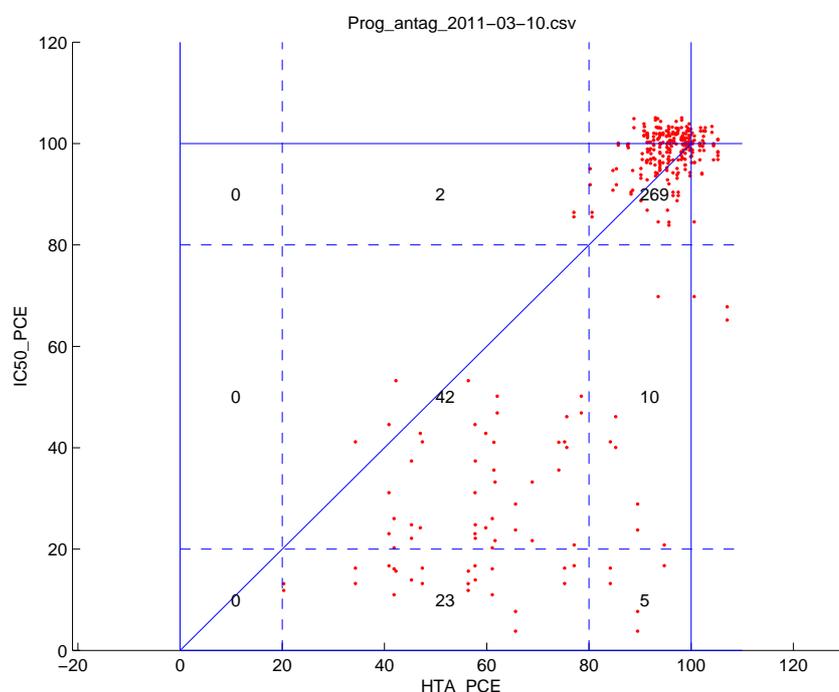


Figure 21: Prog_antag. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

3 Probabilistic error model

3.1 p,q -model

- (3.1.1) In view of the cases where experiments seem to give very low effects when we would expect high, and vice versa, it is natural to propose a model where the errors in percentage effect do not simply have a Gaussian distribution, but also have a probability p of an error that results in a very

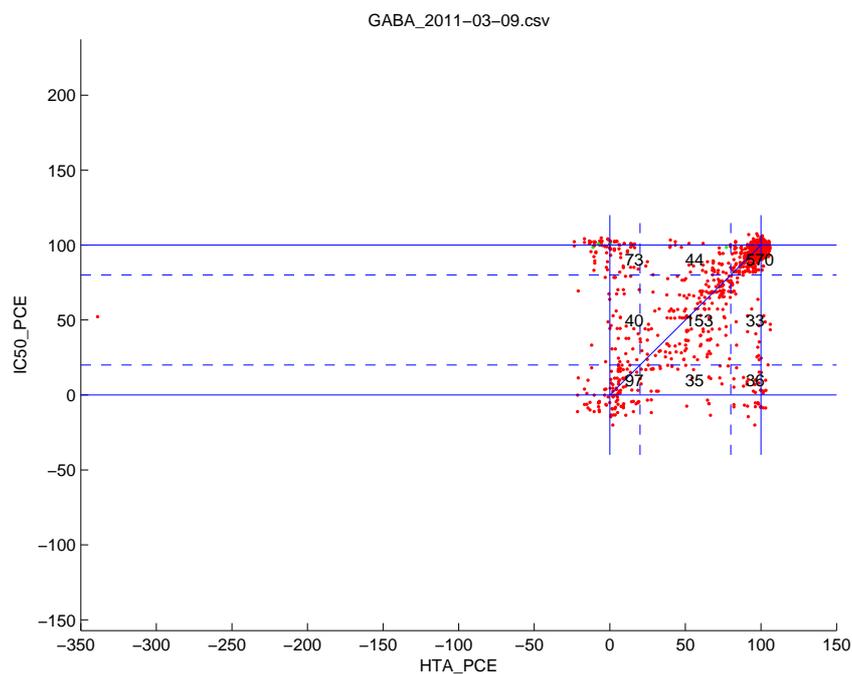


Figure 22: GABA. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

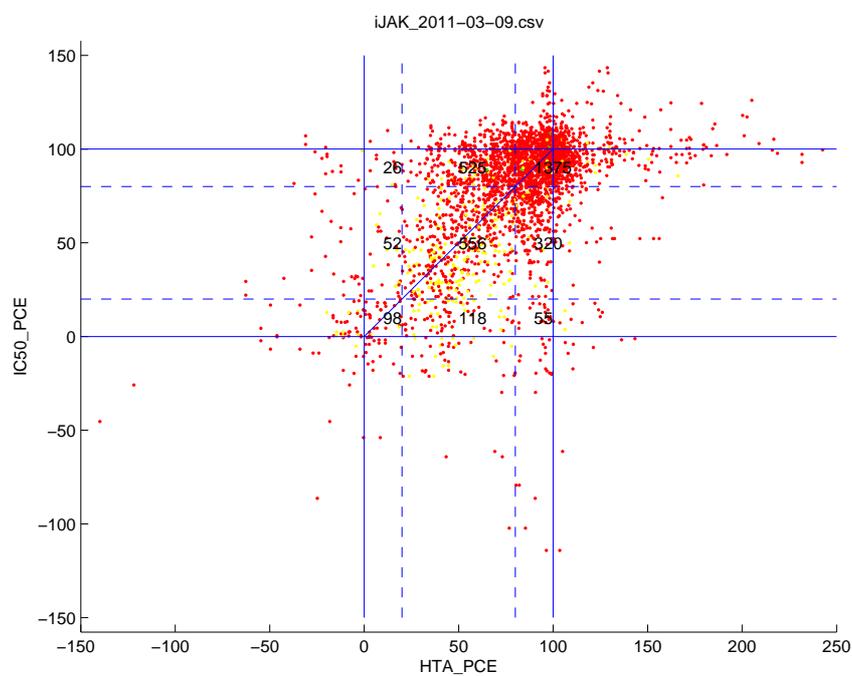


Figure 23: iJAK. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

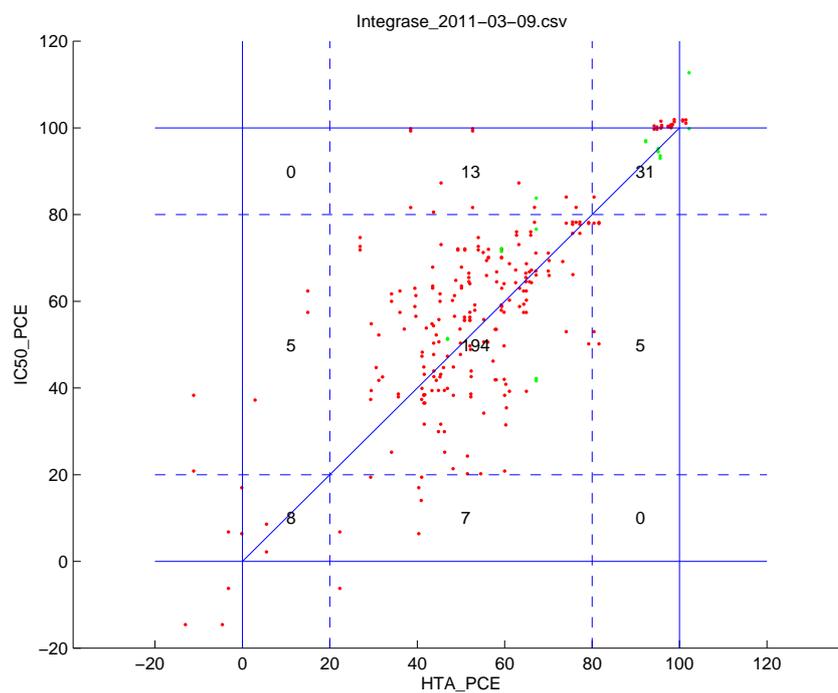


Figure 24: Integrase. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

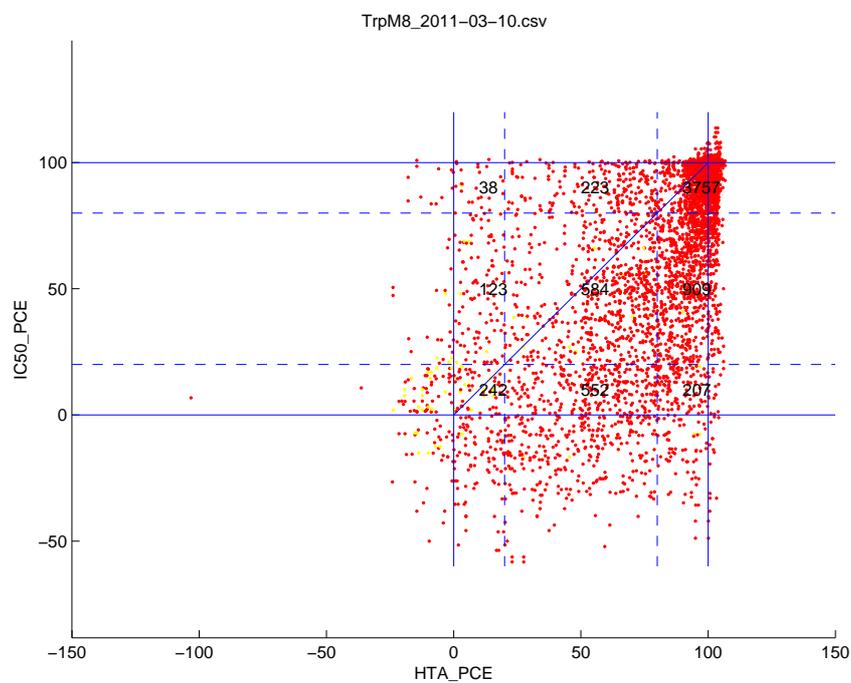


Figure 25: TrpM8. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

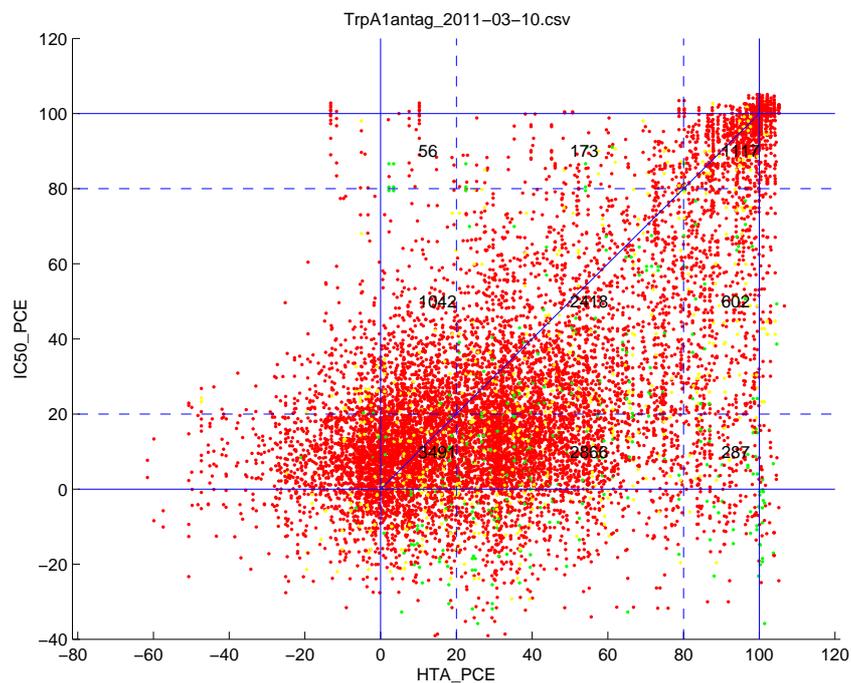


Figure 26: TrpA1antag. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

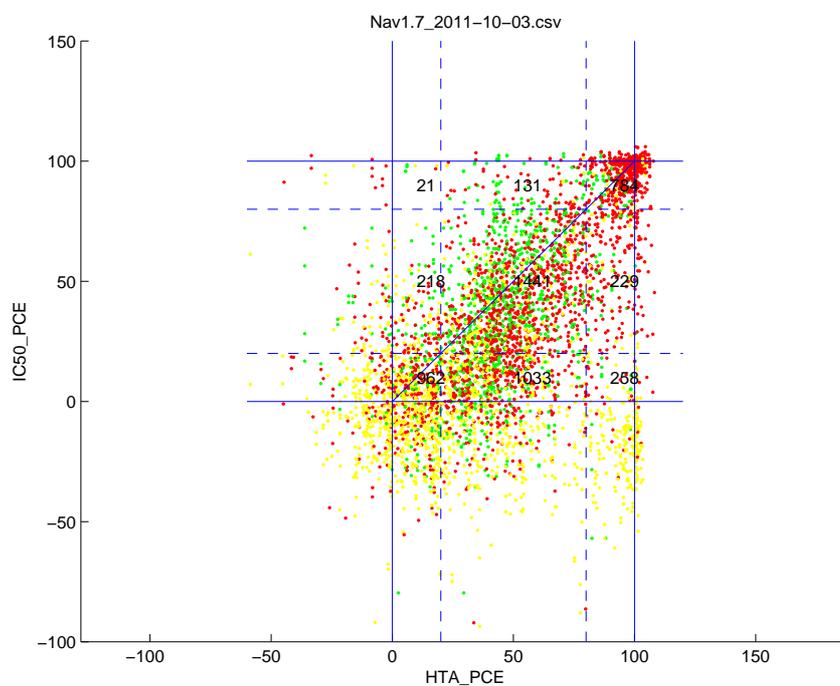


Figure 27: Nav1.7. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

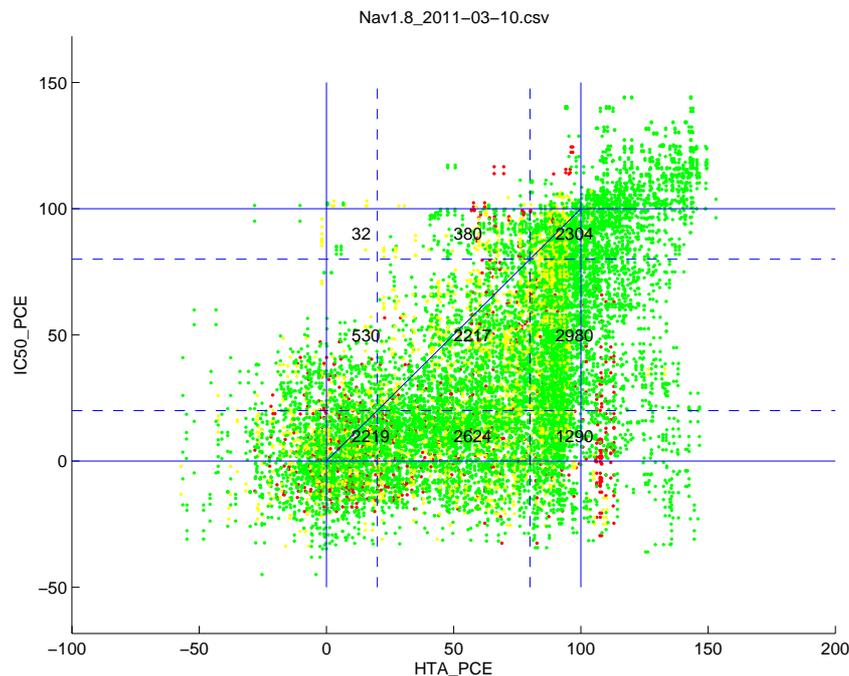


Figure 28: Nav1.8. IC50 percentage effect plotted against HTA percentage effect at the same concentration.

low value — a false negative, and q of an error that results in a very high value — a false positive. We expect these both to be small, so that with probability $1 - p - q$ a genuine result is obtained, but which will still be subject to some error naturally.

- (3.1.2) In discussions, Pfizer said that there can be particular experimental failures that result in both of these: a blocked tip resulting in no effect, or carry-over resulting in possible introduction of an unwanted active compound. The probabilities of these kind of events would depend on the kind of screening that was being carried out and the nature of the reagents: it could not be assumed to be constant for all trials.
- (3.1.3) We shall here work in terms of a model in which the false negatives and positives occur independently with probabilities p and q for each individual experiment.

3.2 Estimating p and q

- (3.2.1) To estimate p , the probability of a false negative, we need enough trials in which we know there should be a strong activity and we need the proportion of null effects in those trials. One possibility is to use the HTA and IC50 results at the same concentration. Since the quality control on the IC50 tests is greater, we could assume that if the HTA result is low

and the IC50 high, then the HTA result is a false negative of the kind we are looking for. So the number of points near the top left corner as a proportion of points near the top edge would be an estimate of p .

- (3.2.2) Another possibility from the data we have here would be to use the IC50 tests on the standard compound for each file, and see how often the IC50 tests at high concentrations, where we can see from the general shape of the dose-response curve that there should be strong activity, in fact show low activity.
- (3.2.3) Both of these methods have obvious problems related to the pre-filtering of the data that has naturally occurred. For instance, the quality control applied to the IC50 tests discards the data from a plate if the results for the standard compound cannot be fitted reasonably well by the model. This undoubtedly means that we are seeing the IC50 data through a selection window meaning that some of the points we would like to have seen to estimate p are not there. This is why the HTA-IC50 comparison is preferable, since (as far as we know) the data has not been filtered on agreement of those figures. For instance if we consider the data shown in Figure 2 then this gives an estimate of p as about 0.2.
- (3.2.4) For estimating q , the probability of a false positive, we can use similar methods to estimate the number of occurrences, either from the HTA-IC50 plots, or from the false positives visible at the lower end of the dose-response curves. Those will be subject to filtering considerations similar to those that apply to the estimation of p . However, there is also the much more serious consideration that the true population from which those are a sample is much larger: there are many HTA tests that correctly measure no effect, and those have not come through into the data that we have available here, because the compound has been ruled out and has not gone through to the IC50 testing. However, assuming that Pfizer has full records of the tests prior to those we have available here, the value of q can be estimated. It will be essentially the number of points near the lower right corner of the HTA-IC50 plot divided by the corresponding number of trials that gave low results. We therefore expect that q will be much less than p .

3.3 Use of this model

- (3.3.1) Imagine we are screening a large number of compounds n , say of order 10^6 , and we suspect that they may contain a small number perhaps $n_A = O(10^2)$ active compounds that we wish to identify. If we do k HTA tests, and we select the compounds that pass at least one of these for further trial, how large should k be to ensure that we catch most of the active compounds with high probability?

- (3.3.2) The probability that an active compound is missed completely is p^k , so the expected number of active compounds missed is $p^k n_A$. The actual number of active compounds missed will be binomial, but approximately Poisson, with this mean. So we can choose k to control the number of missed active compounds as we wish. In fact since n_A is small, we do not need to take k particularly large to get a low probability of missing any of them.
- (3.3.3) A further refinement would be to consider that tests on samples on the same assay plate may have correlated errors that are not just independent with probabilities p and q .