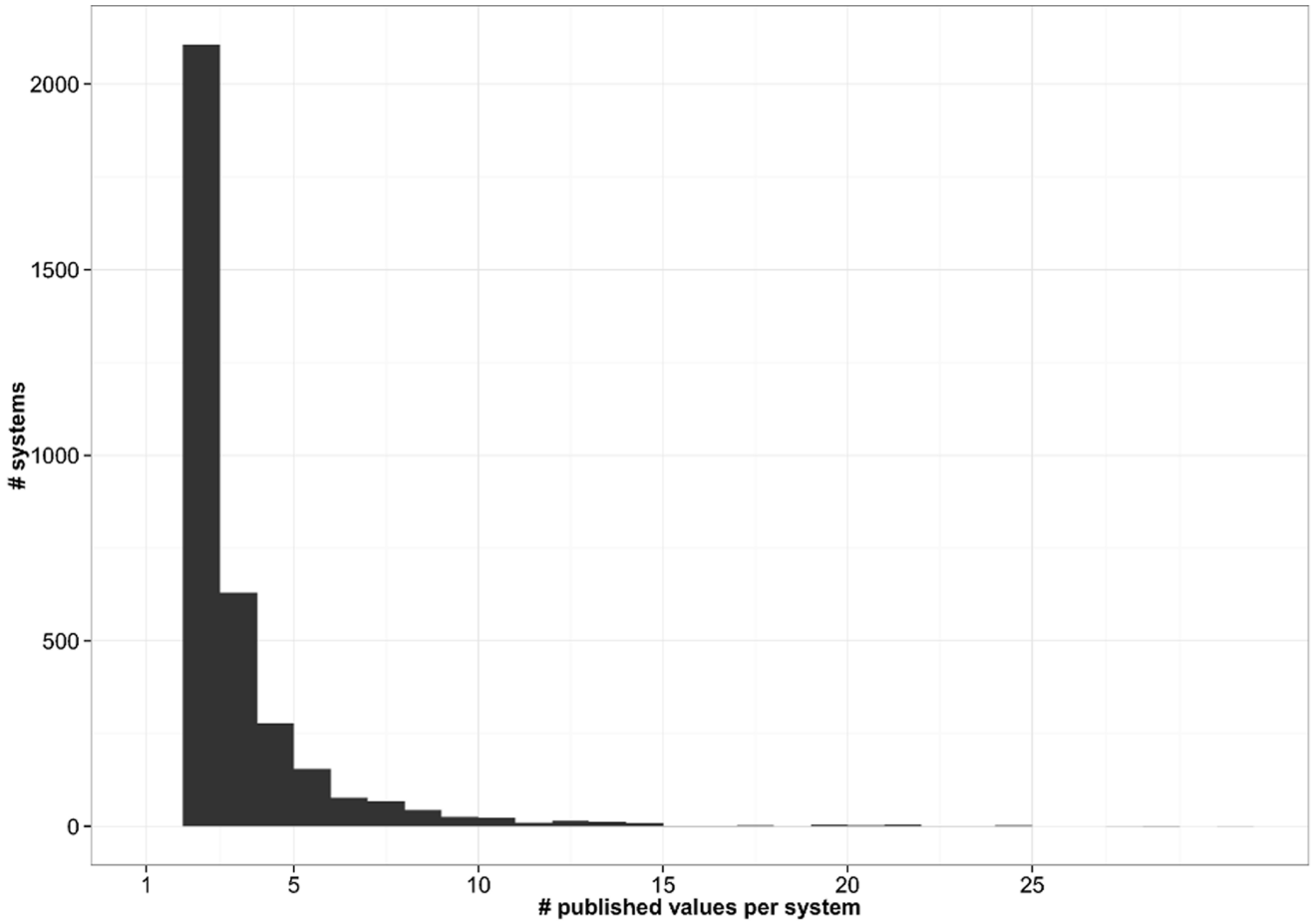


**Figure 3. Number of published independent values per protein-ligand system.**
doi:10.1371/journal.pone.0061007.g003

$\Delta pIC_{50}$'s and $\Delta pK_i$'s with an upper threshold of 2.0 are shown in Figure 5.

The standard deviations of the $\Delta pIC_{50}$ data is constantly 21–26% larger than the standard deviation of the ΔpKi data. After dividing by $\sqrt{2}$, the σ for the Gaussian distribution fitted to all $\Delta pK_i$ values <2.5 then becomes 0.47 (a bit lower than the σ value of 0.54 previously calculated for heterogeneous $pK_i$ data from ChEMBL version 12 data without upper threshold for ΔpKi data. [13] Since σ, MUE, and $M_{ed}UE$ are proportional to each other in Gaussian distributions, we can estimate σ, MUE and MedUE for the $IC_{50}$ data to be 21–26% larger than the same metrics for $pK_i$ data, yielding $\sigma_{pIC50} = 0.68$, $MUE_{pIC50} = 0.55$ and $M_{ed}UE_{pIC50} = 0.43$ (when using a factor of +25% for converting $pK_i$ data to $pIC_{50}$ data).

In order to test the alternative approach of directly obtaining quality metrics from the data, we calculated the quality metrics from the $\Delta pIC_{50}$ data with an upper threshold of $\Delta pIC_{50} = 2.5$. Here, $\sigma_{pIC50} = 0.68$, $MUE_{pIC50} = 0.54$ and $M_{ed}UE_{pIC50} = 0.43$ are obtained. These values are very similar to the values obtained from comparing fitted Gaussian distributions and indicate that the

**Table 2.** Errors found for samples of pairs of measurements with specific differences in measured $pIC_{50}$.

| $\Delta pIC_{50}$ | # invalid pairs out of 10 | Error types found |
|---|---|---|
| From 4.7 to 7.8 | 9 | unit error, receptor subtype error, stereochemistry error, cellular assay error |
| 3.2 | 10 | unit error, cellular assay error, target error, value error |
| 2.5 | 8 | unit error, receptor subtype error, value error |
| 1.5 | 6 (+2 dubious) | unit error, cellular assay error, receptor subtype error, value error |
| 1.1 | 1 (+2 dubious) | cellular assay error, receptor subtype error |
| 0.05 | 1 (+1 dubious) | value error, different assay conditions |
| 0.02 | 0 (+4 dubious) | original paper retracted, data cited from third source which is not available any more, receptor subtype error |

doi:10.1371/journal.pone.0061007.t002