

## An analysis of the attrition of drug candidates from four major pharmaceutical companies

Michael J. Waring<sup>1</sup>, John Arrowsmith<sup>2</sup>, Andrew R. Leach<sup>3</sup>, Paul D. Leeson<sup>3,4</sup>, Sam Mandrell<sup>2</sup>, Robert M. Owen<sup>5</sup>, Garry Pairaudeau<sup>1</sup>, William D. Pennie<sup>6,7</sup>, Stephen D. Pickett<sup>3</sup>, Jibo Wang<sup>8</sup>, Owen Wallace<sup>8,9</sup> and Alex Weir<sup>2</sup>

**Abstract** | The pharmaceutical industry remains under huge pressure to address the high attrition rates in drug development. Attempts to reduce the number of efficacy- and safety-related failures by analysing possible links to the physicochemical properties of small-molecule drug candidates have been inconclusive because of the limited size of data sets from individual companies. Here, we describe the compilation and analysis of combined data on the attrition of drug candidates from AstraZeneca, Eli Lilly and Company, GlaxoSmithKline and Pfizer. The analysis reaffirms that control of physicochemical properties during compound optimization is beneficial in identifying compounds of candidate drug quality and indicates for the first time a link between the physicochemical properties of compounds and clinical failure due to safety issues. The results also suggest that further control of physicochemical properties is unlikely to have a significant effect on attrition rates and that additional work is required to address safety-related failures. Further cross-company collaborations will be crucial to future progress in this area.

<sup>1</sup>AstraZeneca, Alderley Park, Cheshire SK10 4TG, UK.

<sup>2</sup>Thomson Reuters, 77 Hatton Garden, London EC1N 8JS, UK.

<sup>3</sup>GlaxoSmithKline, Stevenage, Hertfordshire SG1 2NY, UK.

<sup>4</sup>Present address: Paul Leeson Consulting, The Malt House, Main Street, Congerstone, Nuneaton, Warwickshire CV13 6LZ, UK.

<sup>5</sup>Pfizer, Cambridge, Cambridgeshire CB21 6GS, UK.

<sup>6</sup>Pfizer, Groton, Connecticut 06340, USA.

<sup>7</sup>Present address: Takeda Pharmaceuticals, Cambridge, Massachusetts 02139, USA.

<sup>8</sup>Eli Lilly, Indianapolis, Indiana 46285, USA.

<sup>9</sup>Present address: Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. Correspondence to M.J.W.

e-mail: [mike.j.waring@astrazeneca.com](mailto:mike.j.waring@astrazeneca.com)

doi:10.1038/nrd4609

Published online 19 June 2015

Reducing the high attrition rates in drug development continues to be a key challenge for the pharmaceutical industry. The number of failures of small-molecule drug candidates due to poor pharmacokinetic profiles seems to have diminished significantly in recent years, but this has been accompanied by a shift towards failures due to efficacy and safety issues, and overall attrition rates remain high<sup>1–3</sup>. However, ascribing single reasons for compound failure could be an oversimplification because there may be multiple factors contributing to the decision to terminate the development of a compound.

Over the past two decades, multiple studies have proposed benefits (in terms of reduced likelihood of attrition) from controlling the physicochemical properties of compounds, such as size, lipophilicity and polarity, based on analyses of data sets of approved oral small-molecule drugs and investigational compounds<sup>4–10</sup>. Beginning with Lipinski's rule of five study<sup>4</sup>, initial studies focused on a link between physicochemical properties and the likelihood of attrition owing to poor absorption, distribution, metabolism and excretion (ADME) characteristics<sup>4–6</sup>. More recently, attention has focused on the influence of physicochemical properties

on promiscuity<sup>5,8</sup> and *in vivo* toxicological outcomes<sup>7,10</sup>. It should, however, be noted that the manner in which analyses such as those in REFS 5,6 (and similar analyses) have been carried out has recently been criticized; in particular, critics have highlighted that the use of averaged or categorized (binned) data can exaggerate the strength of correlations<sup>11</sup>.

Evidence that compound-related issues contribute significantly to attrition due to lack of efficacy in Phase II was found in an analysis of 44 projects conducted at Pfizer from 2005 to 2009; in 43% of the Phase II projects, the biological concept was not tested adequately because the levels of the drug needed for the desired target exposure were not achieved<sup>12</sup>. Success in Phase II trials was associated with achieving high confidence in exposure, binding to the desired target and a pharmacological response. Of the 44 projects examined, only 14 achieved high confidence in all 3 of these factors; of those, 8 reached Phase III trials. Similar principles have also been described more recently by AstraZeneca<sup>13</sup>. It should be noted that the compounds forming the basis of these conclusions will have considerable overlap with those analysed in this article.

**Rule of five**

Lipinski's 'rule of five' was derived from an analysis of small-molecule oral drugs (Phase II and above) and identifies desirable limits for several key physicochemical parameters; specifically, molecular mass <500 Da, number of hydrogen-bond donors <5, number of hydrogen-bond acceptors <10 and calculated octanol–water partition coefficient <5. In the original study 90% of the data set obeyed at least three of these 'rules'.

**logP**

The logarithm of the octanol–water partition coefficient, which is a measure of a molecule's preference for aqueous or lipophilic environments, and can be used to rationalize the ability of molecules to cross biological membranes. logP is defined as the ratio of un-ionized drug distributed between the octanol and water phases at equilibrium. Larger values imply greater lipophilicity.

**Topological polar surface area**

(tPSA). A measure of the degree of polarity of a molecule; the PSA is calculated from the sum of surface contributions from polar fragments. PSA has been shown to correlate with the passive permeability of compounds through membranes. The summation method enables much more rapid calculation than other methods while returning practically identical results.

**Therapeutic index**

In a drug development setting, this is the quantitative ratio of the exposure level at the chosen safety end point divided by the exposure level at the chosen efficacy end point, typically the ratio of the highest exposure to the drug that results in no toxicity over that which produces the desired efficacy.

With regard to the influence of molecular properties on toxicological outcomes, perhaps the most significant work led to the "3/75 rule", derived from an analysis of exploratory or dose-finding toxicology studies of 245 compounds at Pfizer<sup>7</sup>. The key finding from this analysis was that compounds with a calculated logP <3 and a topological polar surface area (tPSA) >75 Å<sup>2</sup> were 2.5 times more likely to be non-toxic than toxic at 10 μM<sup>7</sup>. By contrast, those with a calculated logP >3 and a tPSA <75 Å<sup>2</sup> were 2.5 times more likely to be toxic than non-toxic at the same total exposure. Further support for the influence of compound lipophilicity on toxicology was reported in a study of >400 Eli Lilly and Company (Eli Lilly) compounds in rat toxicological studies<sup>14</sup>. In this analysis, there was a threefold enrichment in toxic compounds (in which toxic compounds were defined as those causing an adverse histological change or death) when logP was >3, but tPSA had little or no influence. A subsequent study from Pfizer has suggested that compounds that achieve efficacy at lower concentrations are more likely to progress through toxicology studies, and it further suggested that a predicted human efficacious total plasma concentration of <250 nM was indicative of a greater chance of success<sup>15</sup>.

Clearly, the benefits of establishing a link between important end points such as an *in vivo* toxicology outcome and simple descriptors such as logP and PSA (which can be easily calculated before synthesis) are highly attractive. However, there are several problems encountered in attempting to do so. First, the outcome of toxicology studies can be both subjective and context-dependent. In considering whether a toxicology finding is sufficient to halt progression of a compound to human subjects, several factors must be considered, including the following: the severity and reversibility of the finding; the margin between the concentration of the compound at which toxicity is observed compared with the concentration needed to show beneficial effects (also known as the therapeutic index); the risk–benefit ratio in the intended disease setting; the likelihood of such an effect translating to humans; and the ability to monitor and treat any subsequent consequences. As such, the decision to stop a drug candidate for toxicological reasons is a complex and nuanced judgement.

In addition, establishing meaningful correlations that might be expected to be generally applicable to future compounds requires large data sets, which are difficult to obtain given the relatively small number of compounds tested in such studies by any individual company. Toxicity analyses by individual companies could be inadvertently biased by the specific projects analysed and by different decision-making approaches to compound progression. Also, given the complexity of the process by which a compound can elicit a toxicological outcome *in vivo*, any correlations with simple physicochemical properties are likely to be weak. Moreover, if the toxicity is a consequence of the primary drug target mechanism or of a specific off-target pharmacology, then no correlations with physicochemical properties would be expected. This is in contrast to toxicology driven by the general promiscuity of a compound — a feature that is more

likely to be correlated with physicochemical properties. Taken together, these considerations make it challenging to establish causality and increase the likelihood of observing chance correlations, which do not ultimately prove to be predictive<sup>16</sup>.

We have compiled attrition data for oral, small-molecule drug candidates from four large pharmaceutical companies — AstraZeneca, Eli Lilly, GlaxoSmithKline and Pfizer — nominated for development in the period between 2000 and 2010. It was anticipated that pooling data would help minimize any issues of inter-company variability and that the larger data set would increase the chances of observing meaningful relationships. Here, the nature of this compiled data set and an analysis of the causes of attrition, with a focus on toxicology and safety, are described. The links between these observations and the physicochemical properties of the drug candidates are assessed. This assessment is presented in as open and objective a manner as possible, with the aim of addressing the criticisms levelled at previous analyses<sup>11</sup>. In particular, the full range of the distributions in the data is shown alongside the mean values, and standard deviations (s.d.) are quoted. As well as statistical tests for differences in the means, the magnitude of the differences are considered to assess the utility of the observations (see Supplementary information S1 (box) for further details).

**Characteristics of the data set**

Overall, the four companies identified a total of 812 oral development compounds nominated during the period 2000–2010 inclusive, along with information regarding their fate at the time of compilation (2012). Specifically, this information included the highest phase of development reached at the time of data collection, whether the compound was still progressing and, if not, the top-level reason for its failure. In addition, the broad disease area and target class were captured. Some of these categories, such as reasons for failure and target class, needed to be ascribed using an element of subjective judgement. In an attempt to achieve the greatest possible consistency between companies in the data compilation, a set of principles and definitions were pre-agreed within the consortium prior to data compilation (for further details, see Supplementary information S1 (box)).

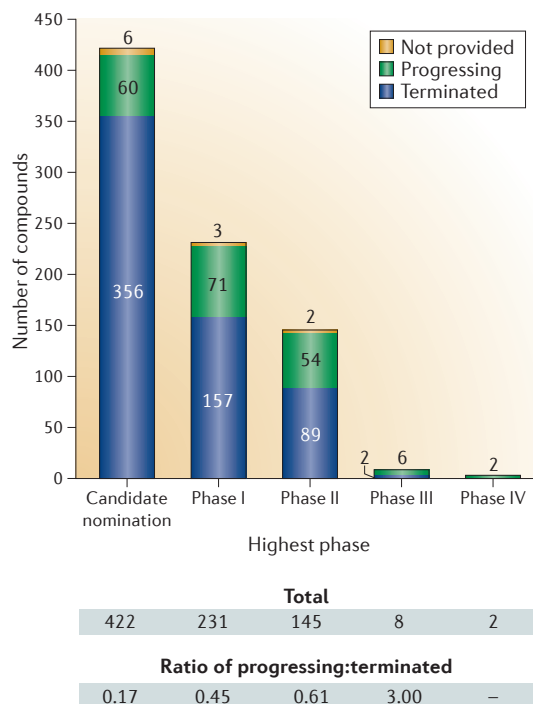
A set of physicochemical descriptors was independently calculated for these compounds by each company using ChemAxon 5.9 (see Further information) to provide a consistent set of descriptors. This set of descriptors included the following: lipophilicity (calculated logP and calculated logD<sub>7.4</sub>); pK<sub>a</sub> (acidic and basic); size descriptors (molecular mass and heavy atom count); hydrogen bonding descriptors (tPSA, acceptor count, donor count); and crude descriptions of substructure (aliphatic atom count, aliphatic ring count, carbocyclic ring count, heteroaliphatic ring count, aromatic atom count, aromatic ring count, carboaromatic ring count, heteroaromatic ring count, chiral centre count, stereodouble bond count, fraction of sp<sup>3</sup> atoms and rotatable bond count). Owing to the confidential nature of these compounds, the actual structures were not

shared, nor were the specific intended pharmacological targets or any known off-target pharmacologies or interactions. The data sets were provided to a third-party broker (Thomson Reuters), which combined the data and encrypted sensitive fields, such as disease area, in accordance with a prior agreement. All analyses were carried out using *JMP Statistical Discovery* software, version 11.0.0, from SAS Institute (see Further information).

Of the 812 compounds, 808 had a highest phase described, of which 422 had not progressed into clinical studies, 231 were in Phase I, 145 had progressed to Phase II, 8 to Phase III and 2 to Phase IV (FIG. 1). Hence, even with such a large combined data set, a sufficient sample of Phase III and IV compounds was still not obtained and subsequent analyses were restricted to earlier-phase compounds. The proportion of compounds still progressing was seen to increase steadily through the phases. These figures do not necessarily reflect the overall rates of attrition and, indeed, do not illustrate a rate consistent with those reported previously<sup>1,2</sup>. There are several reasons for this; notably, compounds progressing to the market are not included in the analysis. However, the figures should be representative of the early phases.

A total of 605 compounds had been terminated, and a primary reason for the termination had been provided for all of these (FIG. 2a; TABLE 1). Non-clinical toxicology was by far the highest cause of attrition, accounting for 240 (40%) of the failures. The second highest cause of attrition was rationalization of the company portfolio (124 compounds; 20%). Breaking the compounds down to those nominated in the first half of the decade (2000–2005) and those in the second half (2006–2010) revealed that the incidence of “rationalization failures” has increased significantly in the second half of the decade, which may be linked to an increased degree of organizational change during this period (FIG. 2b; TABLE 1). Rationalization failures were also prominent (second or third most common) when the data set was broken down by development phase, indicating that such failures had an important role independent of the phase reached. Although a broad range of specific reasons could account for compounds being classified as rationalization failures, including a disease area no longer appearing commercially viable or a mechanism being shown to have efficacy or safety issues, these results imply that substantial resources are invested in research and development across the industry into compounds that are ultimately simply not desired or cannot be progressed for other reasons (for example, agreed divestiture as part of a merger or acquisition). In addition, these results suggest that frequent strategy changes are a significant contributor to lack of research and development success.

It is also clear from this analysis that non-clinical toxicology failures represent one of the key challenges still facing the industry, the incidence of which has remained relatively constant between the two halves of the decade analysed. Breaking this down by phase revealed that the proportion of non-clinical toxicology terminations was most prominent in the preclinical phase, as would be expected (FIG. 2c; TABLE 1). Clinical



**Figure 1 | Composition of the data set by phase and progression status.** The data set is composed of information on 812 oral small-molecule drug candidates from AstraZeneca, Eli Lilly and Company, GlaxoSmithKline and Pfizer. For further details, see the main text and Supplementary information S1 (box).

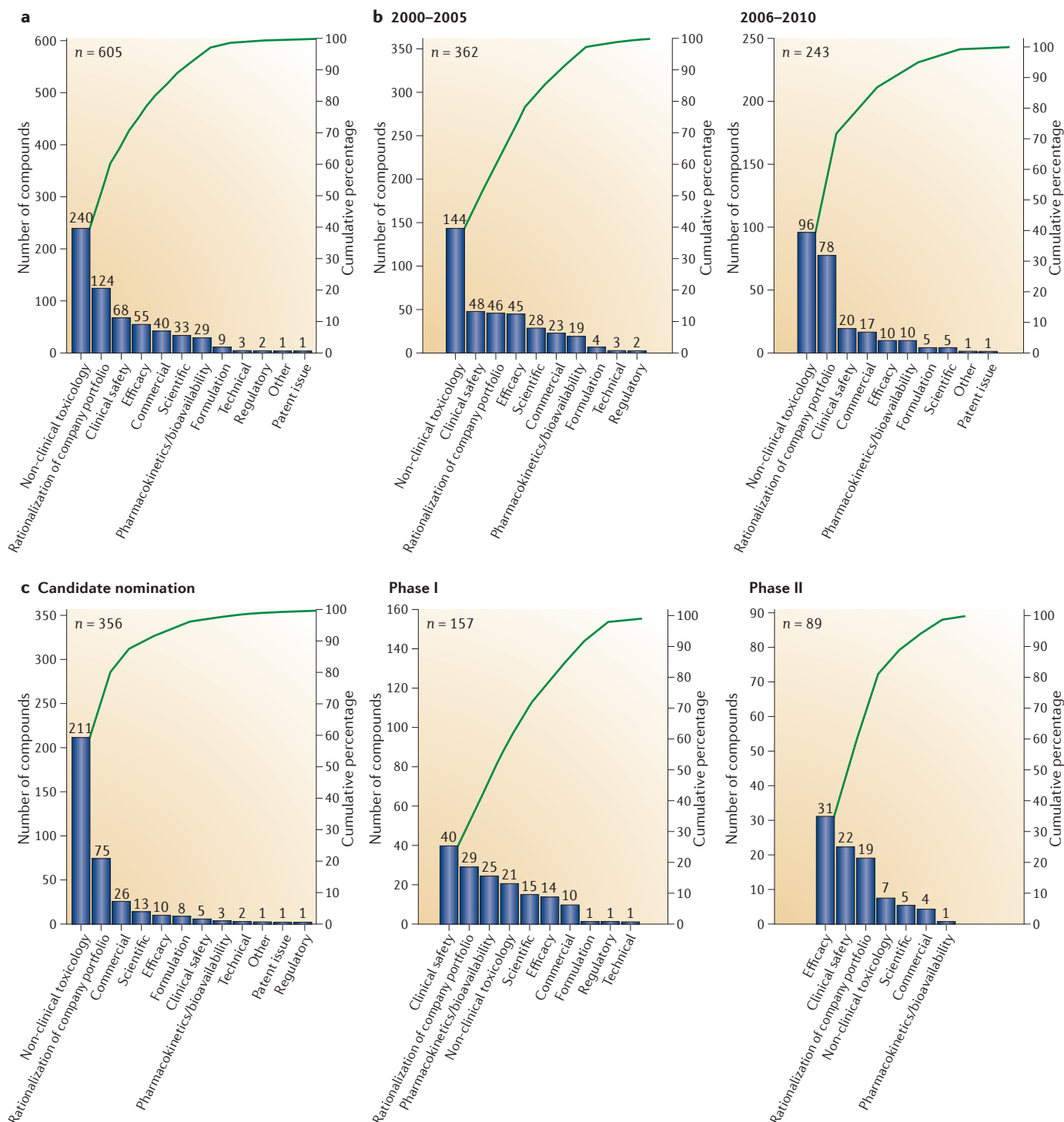
safety was most prominent in Phase I, and lack of efficacy most prominent in Phase II, but a significant number of safety failures were still apparent in Phase II. These data suggest that although minimizing safety-related attrition has been a significant area of investment across the industry, it remains a key area for improvement.

### Physicochemical descriptors

The compounds had a mean molecular mass of 443 Da (s.d. = 110 Da), a mean calculated logP of 3.2 (s.d. = 2.0) and a mean calculated logD<sub>7.4</sub> of 2.2 (s.d. = 2.1). Hence, the majority of compounds in the data set fall within the expected and desirable ranges of these properties, with 75% of compounds having a molecular mass <499 Da and 75% having a logD<sub>7.4</sub> <3.5 (for full distributions of all the descriptors, see Supplementary information S1 (box)). Violations of the two most important Lipinski rule-of-five criteria<sup>4</sup> — molecular mass >500 Da (24% of candidates) and calculated logP >5 (15% of candidates) — are notable, but only 7.6% of the candidates violated both rules. Significantly, for the analysis presented below, few compounds lie outside the accepted ranges (those of launched drugs, for example), and therefore finding statistically significant relationships between the causes of development attrition and physicochemical properties is difficult because of less overall property variation. The majority (53%) of compounds in the set were neutral; 18% were acidic

### logD<sub>7.4</sub>

The logarithm of the octanol–water distribution coefficient, accounting for ionization at pH 7.4. logD<sub>7.4</sub> is equal to the logP for un-ionized (neutral) compounds but lower than logP for acids or bases because the majority of the ionized form partitions into the aqueous phase.



**Figure 2 | Reasons for failure.** **a** | Primary cause of failure for terminated compounds. **b** | Differences in the cause of failure for the first half (2000–2005) and second half (2006–2010) of the decade. **c** | Differences in the cause of failure in preclinical, Phase I and Phase II development.

(having calculated acidic  $pK_a$  of  $<7.4$ ), 21% were basic (basic  $pK_a >7.4$ ) and 8% were zwitterionic (FIG. 3; see Supplementary information S1 (box)).

The distributions of these descriptors were compared to those of marketed drugs and to those of compounds published in patent applications by the four

companies during a similar period<sup>17</sup>. The patented compounds were considered to be representative of the compounds synthesized during the period leading up to the nomination of the compounds in our data set. As such, the progression from patented compounds to drug candidates to launched drugs should represent a

Table 1 | Populations of the primary cause of failure categories for terminated compounds\*

Termination reason	Overall	Period		Phase		
		2000–2005	2006–2010	Candidate nomination	Phase I	Phase II
Clinical safety	68 (11%)	48 (13%)	20 (8%)	5 (1%)	40 (25%)	22 (25%)
Commercial	40 (7%)	23 (6%)	17 (7%)	26 (7%)	10 (6%)	4 (4%)
Efficacy	55 (9%)	45 (11%)	10 (4%)	10 (3%)	14 (9%)	31 (35%)
Formulation	9 (1%)	4 (1%)	5 (2%)	8 (2%)	1 (0.6%)	0
Non-clinical toxicology	240 (40%)	144 (40%)	96 (40%)	211 (59%)	21 (13%)	7 (8%)
Patent issue	1 (0.2%)	0	1 (0.4%)	1 (0.3%)	0	0
Pharmacokinetics or bioavailability	29 (5%)	19 (5%)	10 (4%)	3 (0.8%)	25 (16%)	1 (1%)
Rationalization of company portfolio	124 (21%)	46 (13%)	78 (32%)	75 (21%)	29 (18%)	19 (21%)
Regulatory	2 (0.3%)	2 (0.6%)	0	1 (0.3%)	1 (0.6%)	0
Scientific	33 (5%)	28 (8%)	5 (2%)	13 (4%)	15 (10%)	5 (6%)
Technical	3 (1%)	3 (1%)	0	2 (0.6%)	1 (0.6%)	0
Other	1 (0.2%)	0	1 (0.4%)	1 (0.3%)	0	0
Total	605	362	243	356	157	89

\*Table entries for each column indicate the total number and the percentage in parentheses.

meaningful increase in compound quality, and observations of changes in property distributions between these groups may be suggestive of a physicochemical driver for reducing attrition. Compared with the drug candidates in our data set, the molecular masses of the patented compounds have a similar mean value (443 Da for the drug candidates compared with 450 Da for patented compounds), whereas the mean molecular mass of launched drugs was significantly lower (396 Da), as shown in FIG. 4 (which also includes the s.d. values). The mean calculated logP and calculated logD<sub>7.4</sub> values of compounds in these three sets show significant downward trends from compounds in patent applications to drug candidates to launched drugs (calculated logP = 3.6, 3.2, 2.8, and calculated logD<sub>7.4</sub> = 2.8, 2.2, 1.8, for patented compounds, drug candidates and launched drugs, respectively; FIG. 4). This suggests that there is greater attrition for more lipophilic compounds in both discovery and development. It is also noteworthy that the mean fraction of sp<sup>3</sup> atoms and the mean aromatic ring count change in a small but significant and consistent manner; the mean fraction of sp<sup>3</sup> atoms increases (0.34, 0.37 and 0.42 for patented compounds, drug candidates and launched drugs, respectively) and the mean aromatic ring count decreases (2.9, 2.6 and 2.1 for patented compounds, drug candidates and launched drugs, respectively). It has been suggested that decreased planarity of molecules, represented by the fraction of sp<sup>3</sup> atoms, may lead to improved chances of clinical success<sup>18</sup>, although this is subject to some controversy<sup>11</sup>. Accordingly, it is worth noting that, despite these trends, the distributions overlap considerably, as highlighted by the s.d. values, and the observed changes may be too small to infer any useful guidance with respect to compound design.

### Target class

As part of the data collection, a broad target class relating to the primary molecular target was assigned to each compound. Although there is no completely unambiguous process by which to do this, a consistent system was agreed prior to assignment. These categories included G protein-coupled receptors (GPCRs), nuclear hormone receptors (NHRs) and ion channels. Kinases and proteases were categorized separately to other enzymes and a separate category was included for epigenetic targets, which could also contain enzyme targets such as histone deacetylases. In addition, a miscellaneous category was included along with an option to state whether the molecular target of a drug was unknown. GPCRs represented the most populous target class (34% of the compounds), with enzymes (17%), kinases (13%) and NHRs (9%) being the next most frequent (FIG. 5).

Analysis of these subsets revealed some differences in physicochemical properties between the target classes (see Supplementary information S1 (box)). Ion channel compounds had significantly lower molecular mass distributions (mean = 333 Da); ion channel and protease compounds tended to be less lipophilic (mean calculated logP = 2.1 and 1.8, respectively; mean calculated logD<sub>7.4</sub> = 1.2 for both classes), and NHRs were significantly more lipophilic (mean calculated logP = 4.8; mean calculated logD<sub>7.4</sub> = 3.1). Nevertheless, the majority of compounds populated the expected region of property space regardless of target class. These observations are consistent with similar target-class analyses in the literature<sup>17,19,20</sup>.

### Analysis of toxicology failures

This data set provides a means of investigating the links between physicochemical properties and non-clinical toxicology outcomes. Despite the complexities

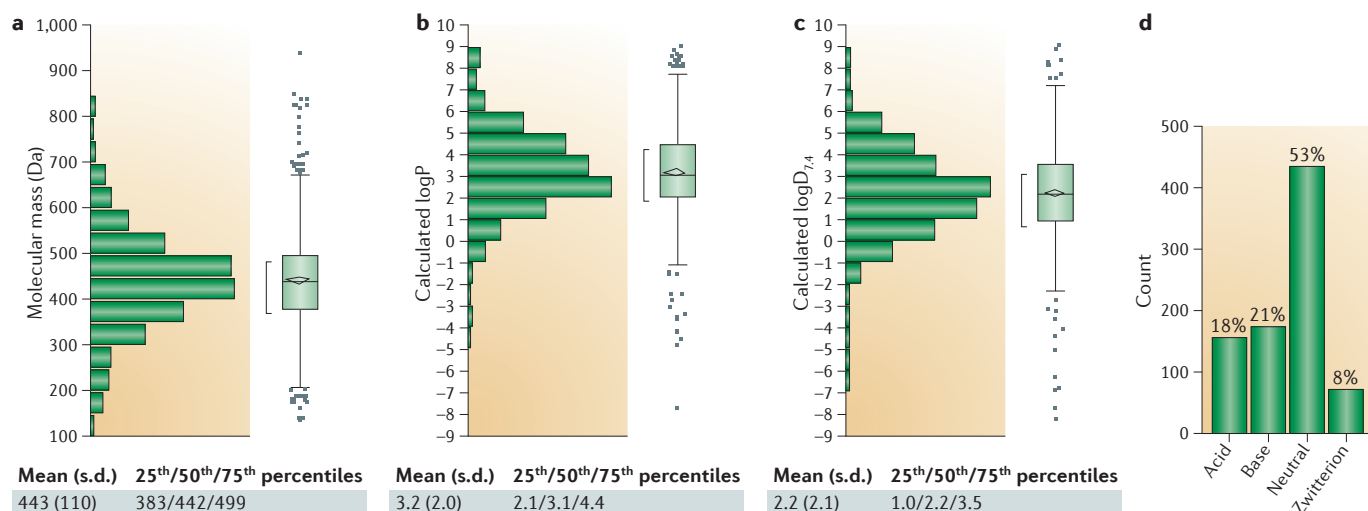


Figure 3 | Distributions of selected calculated properties for the data set. **a** | Molecular mass (Da). **b** | Calculated logP. **c** | Calculated logD<sub>7.4</sub>. **d** | Acid–base properties. For further details, see the main text and Supplementary information S1 (box).

and subjectivity of such an end point noted above, the ultimate desirable outcome of these studies is captured; that is, progression to the clinic. Hence, a meaningful comparison can be made between those compounds that were terminated owing to non-clinical toxicology and those that progressed to the clinic. For the purposes of this comparison, compounds that were still progressing preclinically or those that were terminated for other reasons were removed from the analysis.

Comparing physicochemical property distributions between these two categories revealed no significant differences in their molecular mass, calculated logP, calculated logD<sub>7.4</sub> or tPSA distributions, as assessed by one-way analysis and logistic regression (FIG. 6a). The only structural descriptor showing statistically significant differences in the averages between the two categories (Tukey–Kramer HSD (honestly significant difference) method) was the carboaliphatic ring count, but even this did not show a sufficient effect to be meaningful (see Supplementary information S1 (box) for more details). Although these results appear inconsistent with previous publications linking these parameters with promiscuity<sup>5,7,14</sup> and with *in vivo* toxicological outcomes<sup>10,14</sup>, these differences may be attributable to the different end points used in the analyses. Some differences may also arise from different methods used to calculate the descriptors between studies; readers are encouraged to cross reference the original papers for details.

In addition, application of the 3/75 rule<sup>7</sup> to the data set showed no meaningful differences in toxicology attrition among any of the four categories (the chance of progressing ranged from 0.60 to 0.69, with the highest chance being in the calculated logP >3, tPSA >75 Å<sup>2</sup> category; FIG. 6b). This is consistent with a report suggesting that previous analyses may not be predictive of different compound sets<sup>16</sup>.

Examining the attrition of compounds from different ionization classes revealed no significant differences between acids (defined as those compounds with an acidic

pK<sub>a</sub> < 7.4), bases (basic pK<sub>a</sub> > 7.4) or neutral compounds (FIG. 6c). There was a marked difference in the zwitterionic class, however, with a lower proportion of compounds in this class being terminated for preclinical toxicology (22% compared with 37% for the others). This difference was statistically significant.

There were no major differences across the target classes in their rates of preclinical toxicology attrition, with all classes showing between 31% and 42% attrition (FIG. 6d). The only category outside this range was the ‘unknown’ class, which showed lower attrition (17%), but as this was a very small sample size (*n* = 6), nothing meaningful should be attributed to this observation. It is possible that there are differences in the levels of toxicology that could be tolerated between the classes. For example, kinase inhibitors as a class may have more toxicity than GPCR modulators owing to lower selectivity and associated pharmacology, but may be developed in disease areas in which more significant side effects may be tolerated, such as oncology.

### Clinical safety failures

In a similar manner to that described for preclinical toxicology, clinical safety attrition can be assessed by comparing the compounds failing to show clinical safety in Phase I with compounds successfully reaching Phase II. There are still a significant number of clinical safety failures in subsequent phases; however, compared with Phase I safety failures, failures in Phase II owing to safety issues may be more subtle and idiosyncratic in nature.

This analysis revealed a statistically significant difference between the mean calculated logP values of the compounds failing owing to clinical safety (3.8, s.d. = 1.6) compared with those progressing (3.1, s.d. = 2.1) (FIG. 7a). A similar difference was present, but did not reach statistical significance, in the mean calculated logD<sub>7.4</sub> values (2.7, s.d. = 1.8 and 2.1, s.d. = 2.1 for failed and progressing compounds, respectively). There were no obvious differences in the distributions of the other parameters evaluated.

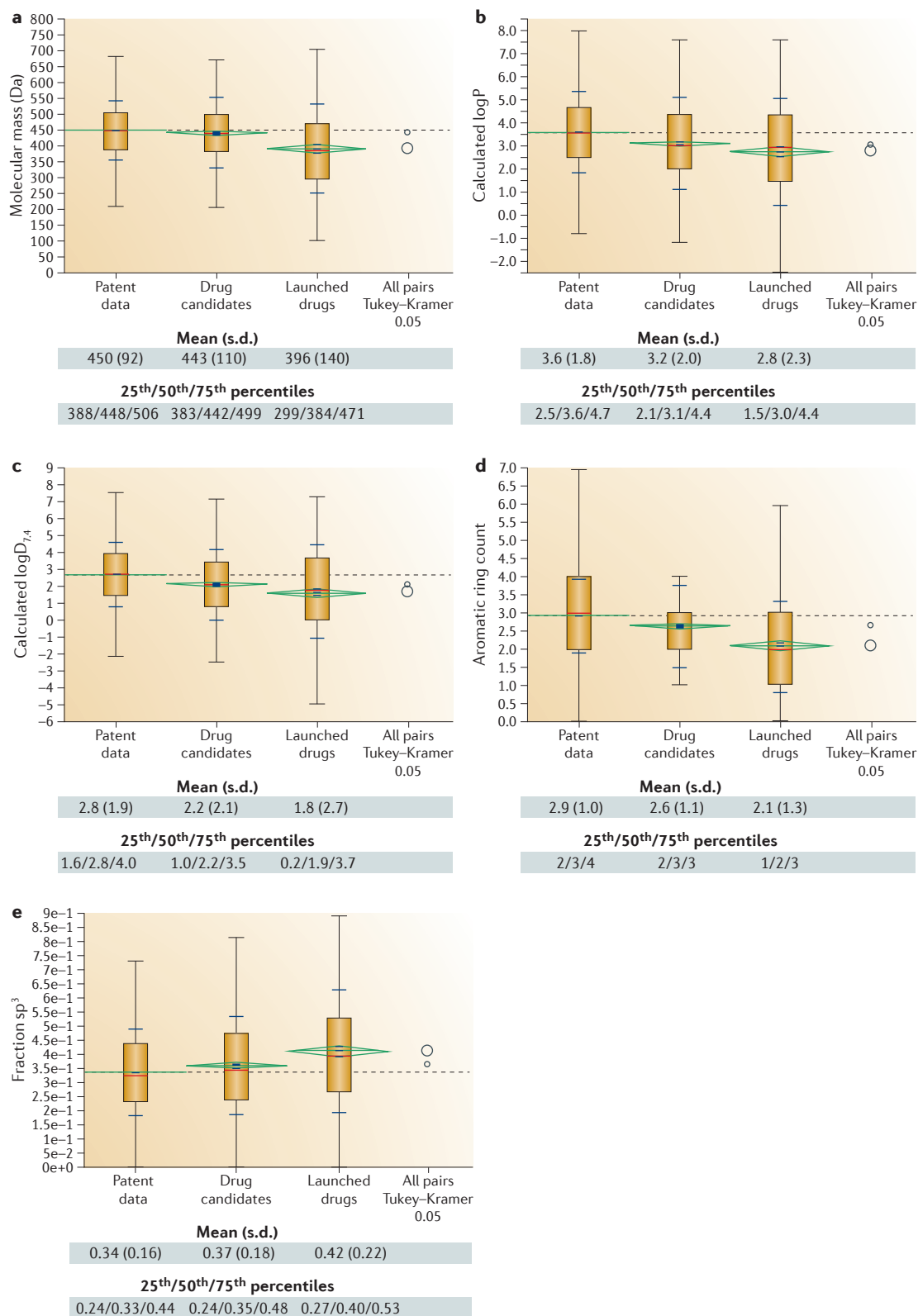


Figure 4 | Comparison of selected physicochemical property distributions for the data set compared with compounds in patent applications and launched drugs. **a** | Molecular mass (Da). **b** | Calculated logP. **c** | Calculated logD<sub>7.4</sub>. **d** | Aromatic ring count. **e** | Fraction of sp<sup>3</sup> atoms. The drug candidate set studied here was shown to have statistically significantly different mean values from both the patent set and the launched drug set for calculated logP, calculated logD<sub>7.4</sub>, aromatic ring count and the fraction of sp<sup>3</sup> atoms. The mean molecular mass of the drug candidate set was statistically significantly different from the launched drug set but not from the patent data set. For an explanation of the depiction of the plots, see Supplementary information S1 (box).

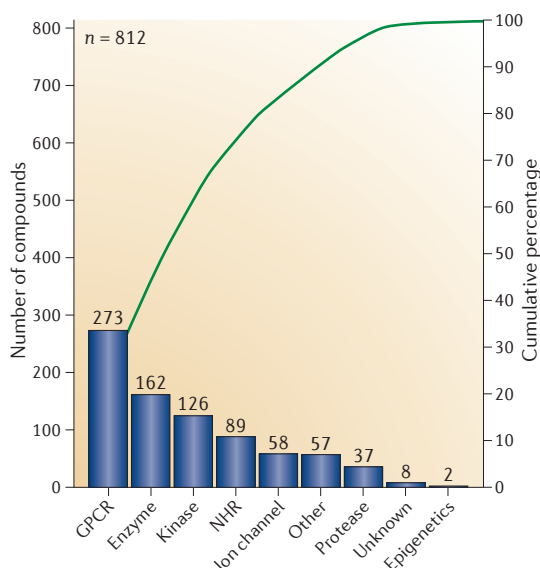


Figure 5 | **Distributions of different target classes within the data set.** GPCR, G protein-coupled receptor; NHR, nuclear hormone receptor.

The reason for this difference is not clear; it can be speculated that the difference may be due to the higher likelihood of promiscuity or polypharmacology of more lipophilic compounds, but it is noteworthy that the same trend is not evident in the preclinical toxicology data. Nevertheless, this is one of the few significant physicochemical property findings in this study. It helps to account for the fact that the mean lipophilicity of marketed drugs is lower than that observed in drug candidates and reinforces the need to continue to control lipophilicity. However, it is not sufficiently discriminatory to suggest that further control of lipophilicity would have a significant impact on success. Examination of how the probabilities of observing clinical safety failures change with calculated logP and calculated  $\log D_{7.4}$  by logistic regression showed that there is no useful difference over the relevant ranges; for example, the probability of failure for a compound with calculated logP of 3.5 in this data set is 22%, whereas it is 19% for a calculated logP of 2.5, compared with an overall probability of 22% (see Supplementary information S1 (box)).

As with preclinical toxicology, acidic, basic and neutral compounds showed similar levels of clinical safety attrition, with zwitterions appearing to fare better (11% failing compared with 23% for the others). In this case, these differences did not reach significance (FIG. 7b).

Some variation in clinical safety attrition was observed across the target classes, with ion channel compounds showing the lowest rates of failure (10%). This observation failed to reach significance for this data set (FIG. 7c).

### Human pharmacokinetics

The primary purpose of a Phase I clinical trial is to assess human pharmacokinetics. Despite considerable progress being made to better predict human pharmacokinetics from preclinical parameters and a notable shift in the

overall attrition in the clinic from pharmacokinetics to safety<sup>3</sup>, pharmacokinetics and bioavailability remained the third most common cause of attrition in Phase I for compounds in our data set (16% of the failures). Comparing compounds failing for pharmacokinetic reasons with those progressing to Phase II enables an objective assessment of the link between poor human pharmacokinetics and physicochemical properties. Comparing the property distributions of the set of calculated parameters revealed no difference in any of their distributions between these two groups (FIG. 8a).

In contrast to safety and toxicology attrition, zwitterions exhibited a statistically significant higher rate of failure attributed to pharmacokinetics (29% compared with 13% for the others) (FIG. 8b). Acidic compounds also showed a statistically significant higher rate of failure (20%) than the basic and neutral compounds (11%). This might be attributed to greater challenges associated with predicting human pharmacokinetic parameters from preclinical data for these types of compounds.

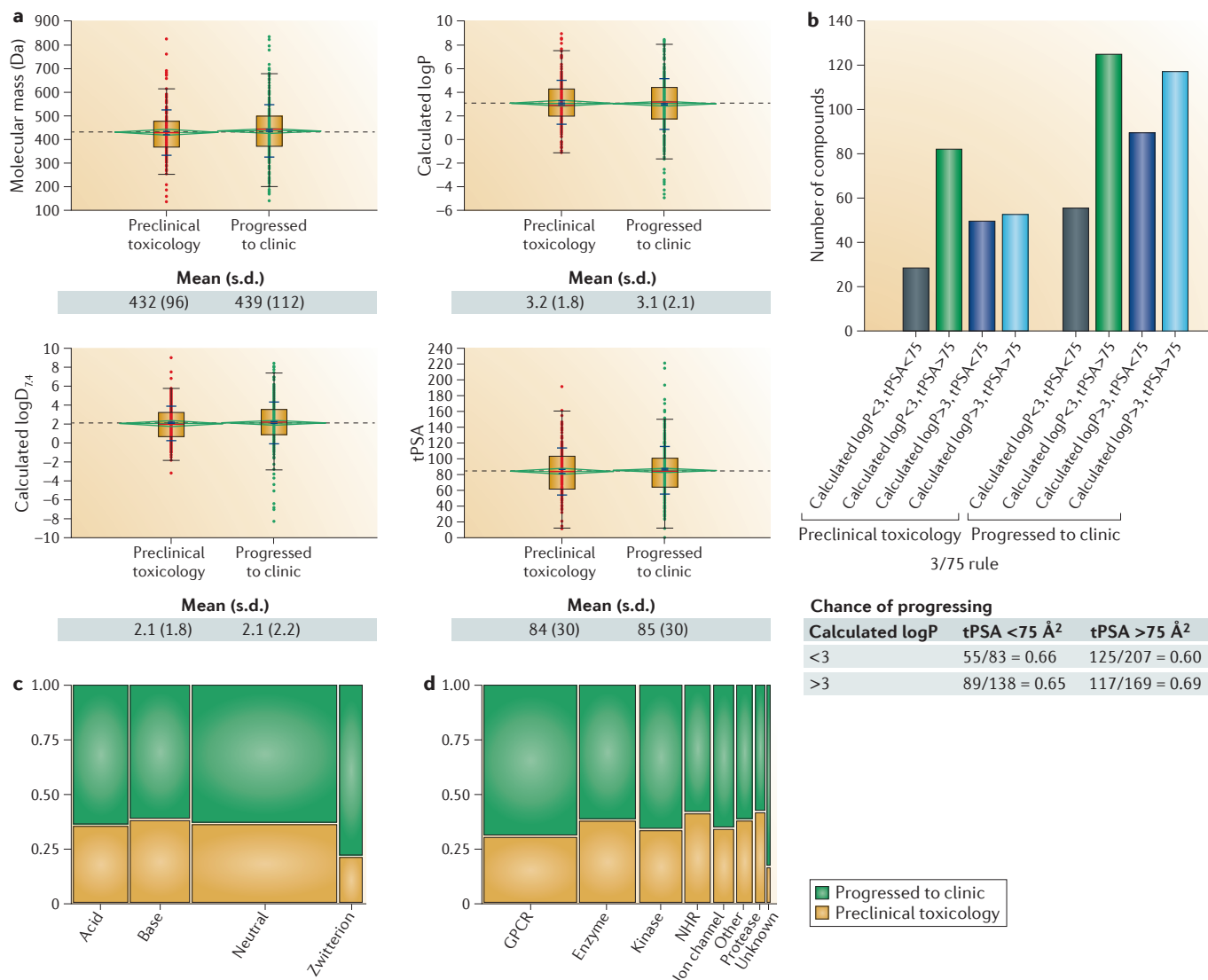
Statistically significant differences in pharmacokinetics attrition were observed across the target classes (FIG. 8c). In this case, GPCR modulators fared worst (excluding the small number in the “unknown” class), with 23% of compounds failing for pharmacokinetics. By contrast, NHR and protease compounds showed no failures. The reason for this is unclear. It cannot be related to the target class per se and if it were related to the compound properties associated with these classes, the trends should have been apparent in the physicochemical property analysis. It is possible that there are causative chemical aspects to this trend that are not captured by the crude descriptors in the data set.

### Discussion

It was hoped that this substantially larger and more diverse data set (compared with previous studies of this type) could be used to identify meaningful correlations between physicochemical properties and compound attrition, particularly toxicity-based attrition. Any such observations would have been useful in providing guidance in the design of new compounds with an increased probability of surviving the development phase. However, beyond reinforcing the already established general trends concerning factors such as lipophilicity, this did not prove generally to be the case. The observation that compounds failing owing to clinical safety in Phase I were significantly more lipophilic (as assessed by calculated LogP) than those progressing into Phase II is new, and may be a factor accounting for the differences in lipophilicity between candidates and drugs.

The observations concerning zwitterions are potentially interesting; if the issues causing the pharmacokinetic failures of these compounds could be addressed then they may deserve greater attention, as they appear inherently less liable to toxicology and safety failure. Of course, many targets may not be compatible with compounds of this type. It is also possible that zwitterions with acceptable pharmacokinetic profiles are no longer differentiated from other compounds with respect to their safety profile. Overall, the results of this



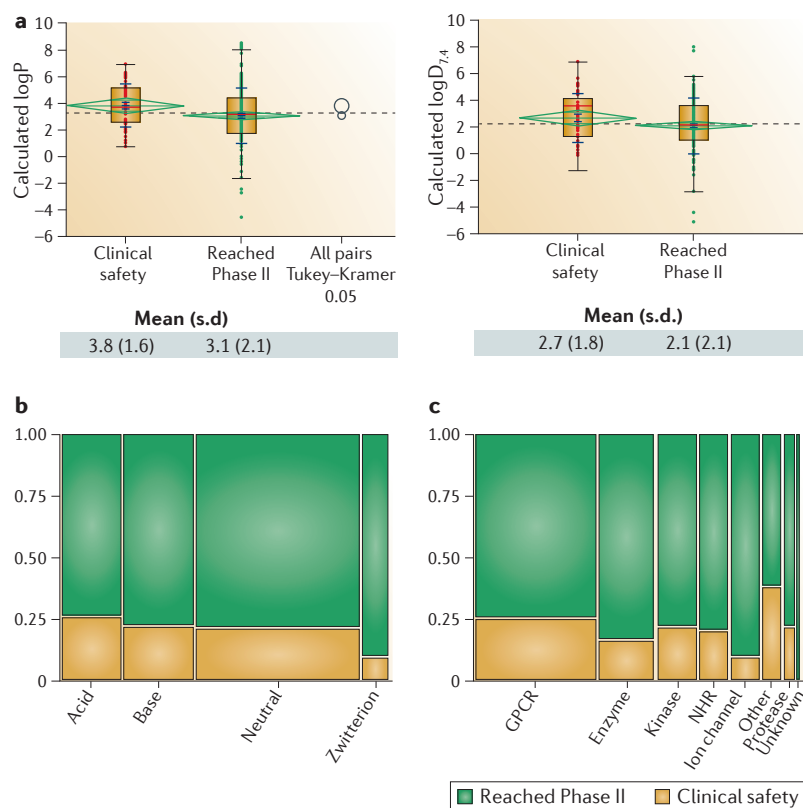


**Figure 6 | Analysis of potential links between physicochemical properties and non-clinical toxicology outcomes.**  
**a** | Comparisons of selected property distributions of the compounds progressing to Phase I ( $n = 386$ ) versus those failing in preclinical toxicology studies ( $n = 211$ ). None of the differences in the mean values were statistically significant.  
**b** | Predictive power of the Pfizer 3/75 rule for the compounds progressing to Phase I versus those failing in preclinical toxicology studies.  
**c** | Ionization classes of the compounds progressing to Phase I versus those failing in preclinical toxicology studies. Probability of observing preclinical toxicology = 0.22 for zwitterions, 0.37 for the combined other classes;  $\chi^2 = 5.0$ ; probability  $\chi^2 > 0.026$ .  
**d** | Target class distributions of the compounds progressing to Phase I versus those failing in preclinical toxicology studies. GPCR, G protein-coupled receptor; NHR, nuclear hormone receptor, tPSA, topological polar surface area.

analysis do suggest more complex means will be required to gain a better understanding of toxicity attrition post-candidate nomination.

Although these results do not contradict analyses already described in the literature, they do suggest that rules derived from one set of compounds may not necessarily be applicable to another. For example, the analysis that led to the 3/75 rule primarily focused on toxicity studies for compounds before candidate nomination that were toxic at a fixed total exposure<sup>7</sup>. In addition, no distinctions in the severity of the toxicity findings were made, and no attempt was made to incorporate the therapeutic index into the analysis. By contrast, this

combined data set contains compounds post-candidate nomination and is focused on the question of compound progression, which is ultimately a question of demonstrating a sufficient therapeutic index. As a result, differences between the conclusions from previous analyses and those described here may be expected. A further caveat is that none of these analyses considered mechanism-related toxicity. It is likely that a reasonable proportion of toxicology and safety failures are attributable to inhibition of the primary target or related proteins for which selectivity cannot be achieved. In these cases, little could be done to avoid the consequences once such a target is selected. In this regard, it is noteworthy that none of the target classes



**Figure 7 | Analysis of potential links between physicochemical properties and non-clinical safety failures.** **a** | Comparisons of calculated  $\log P$  and calculated  $\log D_{7.4}$  distributions of the compounds progressing to Phase II ( $n = 159$ ) versus those failing for clinical safety in Phase I ( $n = 45$ ). The set of compounds failing for clinical safety were shown to have statistically significantly different mean calculated  $\log P$  values from those progressing. The difference in the calculated  $\log D_{7.4}$  values was not statistically significantly different between the two sets. **b** | Ionization classes of the compounds progressing to Phase II versus those failing for clinical safety in Phase I. Probability of observing clinical safety failure in Phase I = 0.11 for zwitterions, 0.23 for the combined other classes;  $\chi^2 = 1.9$ , probability  $> \chi^2 = 0.17$ . **c** | Target class distributions of the compounds progressing to Phase II versus those failing to show clinical safety in Phase I. Probability of observing clinical safety failure in Phase I = 0.10 for ion channels, 0.23 for the combined other classes;  $\chi^2$  (ion channel versus others in class) = 2.2, probability  $> \chi^2 = 0.14$ . GPCR, G protein-coupled receptor; NHR, nuclear hormone receptor.

appear to show differences in toxicity-related attrition. Perhaps more could be done to understand the negative consequences of pharmacological intervention with a particular target at the target validation stage in addition to the possible beneficial effects on disease.

A practical aspect of this study needs to be highlighted. As indicated in the introduction, our goal was to establish a larger data set of compounds than was possible from any one company alone. Owing to the practical challenges arising from the confidential nature of these data, it was decided that calculated properties would be used as the basis for the analysis in this project. This has the advantage that such properties are easily accessible and are easy to implement in a predictive sense, but they can be very crude descriptors of chemical structures and, in some cases, can introduce uncertainties owing to prediction errors<sup>21</sup>. As a consequence, it should also

be remembered that different methods of calculation for parameters such as  $\log P$  and  $\log D_{7.4}$  can result in markedly different values for individual compounds. The relationship of calculated lipophilicity values for new compounds to literature analyses and comparisons between analyses should be performed with caution when the method of calculation differs. Experimental determination of  $\log D_{7.4}$ , wherever possible, is instructive in this regard.

With regard to toxicology-related attrition, a tremendous amount of investment in *in vitro* toxicology screening has been, and continues to be, made within the industry. This provides an alternative means to tackle the problem of subsequent toxicology-related attrition, but it is apparent from the relatively similar incidence of preclinical toxicology failure in the pre- and post-2005 periods that these initiatives may not have made an impact in this timeframe. However, because the underlying mechanism of toxicity-related attrition was not included, it is also possible that although the incidence of toxicology failures has remained the same, the cause of the failures may have changed over the years. For example, the industry has invested heavily in screening for genotoxicity, inhibition of the hERG channel and general promiscuity, and has gained a better understanding of the importance of physicochemical properties in the past 10 years<sup>5,8,9,22</sup>. Although it is impossible to know with any certainty, we speculate that this investment may have improved our ability to remove compounds that are more likely to fail owing to general toxicity (toxicity driven by compound properties, such as phospholipidosis or general promiscuity) or owing to known mechanisms (such as hERG inhibition or genotoxicity) before they reach either *in vivo* toxicology studies or the clinic. As a result, the continued presence of toxicology failures despite these improvements may point to a shift towards attrition due to new specific toxicity mechanisms or those mechanisms that are harder to predict from *in vitro* data. These data clearly highlight the need for further work concerning minimizing safety-related attrition and suggest that this will be a significant challenge for the industry in the coming years. It is likely that novel approaches to develop more predictive assays in the major toxicity arenas are needed to address this critical problem. In the same spirit as this study, collaborative approaches to data sharing, some of which are already initiated<sup>23</sup> (see Further information), may have a substantial impact on identifying areas for improvements.

The control of lipophilicity and molecular mass as a means to optimize preclinical ADME properties<sup>8,9,24</sup> and certain common toxicological outcomes, such as hERG inhibition<sup>25</sup> and phospholipidosis<sup>26</sup>, is well established. These principles provide a clear means of increasing the chances of finding viable development candidates. This notion is supported by the fact that the compounds selected in this study lie within the expected desirable range for such parameters and that the expected trends are observed between compounds in patent applications, development candidates and marketed drugs. Although there is room for further improvement in the physicochemical property distributions of clinical

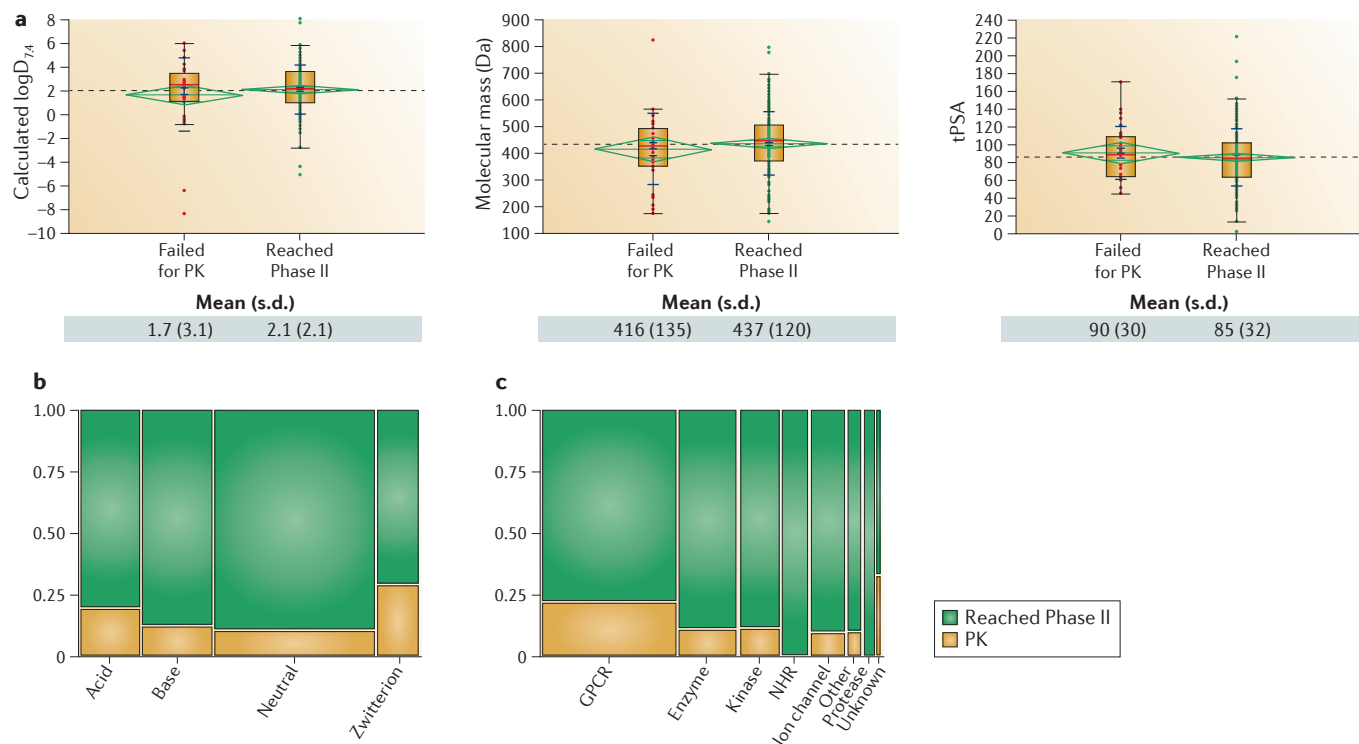


Figure 8 | Analysis of potential links between physicochemical properties and failures due to poor pharmacokinetics.

**a** | Comparisons of selected calculated property distributions of the compounds progressing to Phase II ( $n = 159$ ) versus those failing owing to poor pharmacokinetics (PK) in Phase I ( $n = 28$ ). None of the differences in the mean values were statistically significant. **b** | Ionization classes of the compounds progressing to Phase II versus those failing owing to poor PK in Phase I. Probability of observing PK failure in Phase I = 0.29 for zwitterions, 0.13 for the combined other classes;  $\chi^2 = 3.7$ , probability  $> \chi^2 = 0.054$ . **c** | Target class distributions of the compounds progressing to Phase II versus those failing for PK in Phase I. Probability of observing PK failure in Phase I = 0.23 for G protein-coupled receptors (GPCRs), 0.0 for nuclear hormone receptors (NHRs),  $\chi^2$  (GPCRs versus other classes) = 5.7, probability  $> \chi^2 = 0.017$ ;  $\chi^2$  (NHRs versus other classes) = 5.1, probability  $> \chi^2 = 0.024$ . tPSA, topological polar surface area.

candidates (mean calculated logP = 3.2) towards drugs (mean calculated logP = 2.8), it is questionable whether this would significantly reduce development attrition, based on our findings.

### Conclusions

The data set collected and analysed in this study is the largest compiled so far for the purposes of linking attrition to simple molecular properties. We continue to believe that control of physicochemical properties during compound optimization is highly beneficial in identifying compounds of candidate drug quality. This study extends the analysis to their subsequent performance in development and assesses, for the first time, a link to various reasons for failure in development. The most notable finding from this data set is the difference in lipophilicity between compounds failing owing to clinical safety in Phase I and those successfully progressing to Phase II.

It is clear, however, that fully understanding and addressing the multiple causes of attrition in clinical development will require new thinking and fresh insight. This will be a significant challenge owing to the large number of potential causes of attrition and the complexities of fully understanding these causes at a mechanistic level. We can only hope to address such issues by concerted efforts across the pharmaceutical industry and beyond. That four of the largest pharmaceutical companies were able to work together to address this problem is perhaps the most significant outcome of this study, demonstrating that with an appropriate framework it is possible to tackle the key challenge of attrition by working together while still protecting the intellectual property and commercial interests of the individual organizations. We hope that this study will act as a springboard for future collaborations of a similar nature.

- Hay, M. *et al.* Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
- Bunnage, M. Getting pharmaceutical R&D back on target. *Nat. Chem. Biol.* **7**, 335–339 (2011).
- Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–716 (2004).

- Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *Adv. Drug Delivery Rev.* **23**, 3–25 (1997).
- Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **6**, 881–890 (2007).

- This paper shows, perhaps for the first time, a link between physicochemical properties, notably lipophilicity and molecular mass, and *in vitro* promiscuity.
- Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **51**, 817–834 (2008).

7. Hughes, J. D. *et al.* Physicochemical drug properties associated with *in vivo* toxicological outcomes. *Bioorg. Med. Chem. Lett.* **18**, 4872–4875 (2008). **From the observation of a link between logP and PSA and toxicology in preclinical *in vivo* toxicology studies, this study led to what is now known as the 3/75 rule.**
8. Peters, J.-U., Schneider, P., Mattei, P. & Kansy, M. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem* **4**, 680–686 (2009).
9. Gleeson, M. P., Hersey, A., Montanari, D. & Overington, J. Probing the links between *in vitro* potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* **10**, 197–208 (2011).
10. Luker, T. *et al.* Strategies to improve the *in vivo* toxicology outcomes for basic candidate drug molecules. *Bioorg. Med. Chem. Lett.* **21**, 5673–5679 (2011).
11. Kenny, P. W. & Montanari, C. A. Inflation of correlation in the pursuit of drug-likeness. *J. Comput. Aided Mol. Des.* **27**, 1–13 (2013).
12. Van Der Graaf, P. H. *et al.* Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov. Today* **17**, 419–424 (2012). **Based on an analysis of development failures, this study suggests a set of criteria that a compound should meet to successfully progress through clinical development.**
13. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–431 (2014).
14. Watson, D. E. *et al.* Relating molecular properties and *in vitro* assay results to *in vivo* drug disposition and toxicity outcomes. *J. Med. Chem.* **55**, 6455–6466 (2012).
15. Wager, T. T. *et al.* Improving the odds of success in drug discovery: choosing the best compounds for *in vivo* toxicology studies. *J. Med. Chem.* **56**, 9771–9779 (2013).
16. Muthas, D., Boyer, S. & Hasselgren, C. A critical assessment of modeling safety-related drug attrition. *Med. Chem. Commun.* **4**, 1058–1065 (2013). **This analysis of the application of physicochemical descriptors to modelling safety attrition suggests that many approaches, including some of those cited here, may not be generally applicable to other data sets.**
17. Leeson, P. D. & St-Gallay, S. A. The influence of the 'organizational factor' on compound quality in drug discovery. *Nat. Rev. Drug Discov.* **10**, 749–765 (2011).
18. Lovering, F., Bikker, J. & Humblet, C. Escape from Flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009). **This paper suggests that molecules containing a greater degree of saturation, assessed using the fraction of sp<sup>3</sup> atoms, may be more likely to progress through various clinical stages.**
19. Morphy, R. The influence of target family and functional activity on the physicochemical properties of pre-clinical compounds. *J. Med. Chem.* **49**, 2969–2978 (2006).
20. Vieth, M. & Sutherland, J. J. Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* **49**, 3451–3453 (2006).
21. Tetko, I. *et al.* Accurate *in silico* logP predictions: one can't embrace the unembraceable. *QSAR Comb. Sci.* **28**, 845–849 (2009).
22. Waring, M. J. Lipophilicity in drug discovery. *Expert Opin. Drug Discov.* **5**, 235–248 (2010).
23. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922 (2012).
24. Waring, M. J. Defining optimum lipophilicity and molecular weight ranges for drug candidates — molecular weight dependent lower logD limits based on permeability. *Bioorg. Med. Chem. Lett.* **19**, 2844–2851 (2009).
25. Waring, M. J. & Johnstone, C. A quantitative assessment of HERG liability as a function of lipophilicity. *Bioorg. Med. Chem. Lett.* **17**, 1759–1764 (2007).
26. Tomizawa, K., Sugano, K., Yamada & H. Horii, I. Physicochemical and cell-based approach for early screening of phospholipidosis-inducing potential. *J. Toxicol. Sci.* **31**, 315–324 (2006).

#### Acknowledgements

We thank N. Blomberg (AstraZeneca), A. de Dios (Eli Lilly), M. Beaumont and J. Valentine (GlaxoSmithKline) and S. Louise-May (Pfizer) for their roles in the compilation of the dataset.

#### Competing interests statement

The authors declare [competing interests](#): see Web version for details

#### FURTHER INFORMATION

ChemAxon: <http://www.chemaxon.com>  
 eTOX: <http://www.etoxproject.eu>  
 JMP Statistical Discovery: <http://www.jmp.com>

#### SUPPLEMENTARY INFORMATION

See online article: S1 (box)  
**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**