Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications

Hsin-Pei Shih[1], Xiaodan Zhang[2] and Alex M. Aronov[1]

Abstract | The productivity of the pharmaceutical industry has been widely discussed in recent years, particularly with regard to concerns that substantial expenditures on research and development have failed to translate into approved drugs. Various analyses of this productivity challenge have focused on aspects such as attrition rates at particular clinical phases or the physicochemical properties of drug candidates, but relatively little attention has been paid to how the industry has performed from the standpoint of the choice of therapeutic mechanisms and their intended indications. This article examines what the pharmaceutical industry has achieved in this respect by analysing comprehensive industry-wide data on the mechanism–indication pairs that have been investigated during the past 20 years. Our findings indicate several points and trends that we hope will be useful in understanding and improving the productivity of the industry, including areas in which the industry has had substantial success or failure and the relative extent of novelty in completed and ongoing projects.

The single biggest contributor to the cost of drug research and development (R&D) is the high project attrition rate. Attrition is due to various reasons, including insufficient efficacy, toxicity and strategic reasons, which have been analysed in various previous studies from a range of perspectives[1–14]. The pharmaceutical industry has responded by identifying areas for improvement, with success in some cases. For example, attrition due to poor pharmacokinetics and bioavailability was reduced from ~40% to ~10% between 1991 and 2000 (REF. 1), which can be attributed to improvements in *in vitro* absorption, distribution, metabolism and excretion (ADME) assays[15,16], *in silico* ADME prediction models[17] and the increased translatability of both to human pharmacokinetics[18,19]. More recently, companies have moved investigative toxicology activities into the discovery phase, and there are early signs that this has led to lower toxicity-related attrition in clinical development[20–22].

However, there does not seem to have been a substantial reduction in attrition due to insufficient efficacy. Various groups have reported attrition rates due to lack of efficacy in the past two decades over different periods[1,3,4,6,9,10,13,14,23], and although direct comparison is difficult (in part owing to differences in the data sets), a series of studies using the same database have reported that failure rates in phase II or III trials due to efficacy were consistently ~50–60% between 2008 and 2015 (REFS 6,9,14). This finding indicates that more effective target selection continues to be the major challenge in improving the productivity of drug R&D[24].

This observation has prompted us to evaluate industry productivity from a different angle. Instead of dissecting attrition through R&D milestones or as a function of the physiochemical properties of drug candidates, as has been done in some previous studies (for example, REFS 1,2,4,10,12), we focused on two key characteristics of drug development projects — the therapeutic mechanism of action and the intended indication of the drug candidate — for projects across the industry over the past 20 years. The selection of a given mechanism–indication pair to pursue is based on integrating the understanding of multiple factors, including disease biology, pharmacological tractability of the target, safety risk and commercial opportunities, and we hope that choices that have been made in the past could shed further light on reasons for success or failure and potentially inform future projects.

Data collection and definitions

Data were collected on drug programmes (defined as all indications pursued for a given drug) from the Cortellis database (produced by Clarivate Analytics), a pharmaceutical pipeline database, from 1996 to 2016 (BOX 1). Specifically, we parsed and analysed the programme information, which includes the identity of

[1]*Global Informatics Group, Modelling & Informatics, Vertex Pharmaceuticals Inc., 50 Northern Avenue, Boston, Massachusetts 02210, USA.*
[2]*Global Information Services, Vertex Pharmaceuticals Inc., 50 Northern Avenue, Boston, Massachusetts 02210, USA.*
Correspondence to H.P.S. and A.M.A.
hsin-pei_shih@vrtx.com; alex_aronov@vrtx.com

doi:10.1038/nrd.2017.194
Published online 27 Oct 2017; corrected online 15 Dec 2017

Box 1 | **Sources and process for creation of the data set**

We parsed data for drug projects from the Cortellis pharmaceutical pipeline database as of April 2016 and included project data from across the world (not limited to the United States). The data included the drug names, therapeutic mechanisms, indications, highest phases and the developmental phase of each indication. We organized the data by each drug's therapeutic mechanism, its indications and the developmental phases for the enlisted indications.

To ensure the data quality and enable its use for our purposes, we manually removed selected data in several steps and checked the assignments for drug mechanisms and indications with other databases. First, we eliminated vaccines, imaging agents and combination therapies, while keeping small-molecule, antibody, protein and nucleic acid therapies. Second, we kept only one project for the same active pharmaceutical ingredient related to a given mechanism–indication pair. If more than one formulation was identified, we kept the most advanced project. We also eliminated biosimilars. Third, we verified approved drug mechanisms based on published curated data[78]. Fourth, in cases of drugs with known polypharmacology, we tried to assign the most relevant therapeutic mechanism to the indication (for example, targeting KIT for gastrointestinal stromal tumours and ABL for chronic myeloid leukaemia in the case of imatinib). If multiple mechanisms of a drug are likely to be responsible for therapeutic efficacy in the indication, we assigned multiple mechanisms of this drug to the indication. In cases where the mechanism of the drug appears to be unknown or promiscuous, we eliminated those drugs (for example, doxepin). Fifth, we checked the accuracy of drug information for early discovery projects. For some projects, we were able to confirm or correct the data. However, many early discovery projects have limited preclinical information against which to verify the drug mechanisms. In those cases, we relied on curation from Cortellis, with an understanding that insufficient information and curation quality may potentially serve as a limitation for this data set. It is also worth noting that the classification of drug mechanisms or indications defined by the data source may not always agree with other data sources. We reconciled some inconsistencies in indication and therapeutic mechanism terminology.

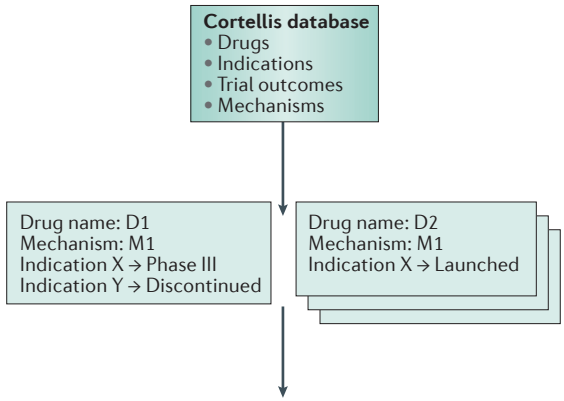
Following this data scrubbing, we transformed the data on the basis of mechanism–indication pairs (see figure, panel **a**). For each mechanism–indication pair, there could be one or more associated drugs in various R&D phases. And for a given drug, there could be multiple associated mechanism–indication pairs if it has been pursued for multiple indications, as is often the case in areas such as cancer and bacterial infection, or if more than one mechanism was assigned for a drug in a given indication.

Overall, at the time when we extracted data for our analysis in April 2016, there were 2,441 therapeutic mechanisms, 1,453 indications and 10,107 unique drugs in various developmental phases recorded in the Cortellis database. In our analysis, a total of 15,101 mechanism–indication pairs were represented by 22,587 projects (defined as a particular drug being pursued for a particular indication).

For our analysis, we further grouped projects into three categories based on their development status: successful (pre-registration, registered and launched projects), discontinued (projects classified as discontinued, withdrawn or suspended) and ongoing (projects classified as in discovery or phase I, II or III trials). Finally, we categorized the mechanism–indication pairs into three non-overlapping groups: validated, unvalidated and emerging, as shown in the figure (panel **b**).

We would also like to point out the limitations in our analysis due to the nature of the data set. First, the indication ontology is not standardized — some indications are broad whereas others may be highly specific. Another issue is related to the temporal progress of drug projects. Discovery-stage projects tend to be evaluated for and assigned to broader indications, while clinical projects focus on more specific indications based on early-stage findings and company strategy. For example, many mechanisms, such as matrix metalloproteinases in oncology, were for projects that mostly did not reach the stage of being linked to specific cancer indications. The annotations in antibacterial indications tend to be very specific, leading to higher counts. Another issue may stem from the assignment of similar indication terms (for example, hypercholesterolaemia and hyperlipidaemia). Other indication ontology issues may be related to a certain amount of inconsistency between the database and regulatory agencies. Overall, we are convinced that the totality of the data outweighs the aforementioned limitations. To reduce the impact of the ontology issues, we manually excluded some problematic pairs when selecting those for the display items, and also split human-targeted and pathogen-targeted mechanisms.

a



Drug	Mechanism	Indication	Developmental phase	Highest phase
D1	M1	X	Phase III	Phase III
D1	M1	Y	Discontinued	Phase III
D2	M1	X	Launched	Launched
:	:	:	:	:

Mechanism–indication		Drug	Developmental phase	Highest phase
M1	X	D1	Ongoing	Ongoing
		D2	Pre-registration and above	Pre-registration and above
M1	Y	D1	Discontinued	Ongoing
:	:	:	:	:

15,101

Mechanism–indication pairs		
Mechanism	Indication	Drugs
M1	X	D1, D2
M1	Y	D1
M2	Z	D3, D4, D5
:	:	:
2,441	1,453	10,107

b

Mechanism–indication pairs	Numbers of projects in terms of R&D status		
	Pre-registration and above	Discontinued	Ongoing
Validated	≥1	Any	Any
Unvalidated	0	≥1	Any
Emerging	0	0	Any

Pharmaceutical pipeline database
A pharmaceutical pipeline database contains extensive information on drug development projects from discovery through to launch, including molecular structures, origins, therapeutic rationales, biological targets, drug properties, indications, licensing details, development history, trial outcomes and scientific references. The information comes from a variety of sources, including press releases, newsletters, conferences, scientific literature and other databases such as clinical data and patents.

the therapeutic mechanism, the highest clinical phase reached and the development status for each specific indication pursued. In total, we collected data for over 10,000 unique drugs with their therapeutic mechanisms and the R&D phases of the intended indications, which corresponded to 22,587 projects (defined as a particular drug being pursued for a particular indication).

To simplify our analysis, the development stage of each project was sorted into three categories: successful, discontinued and ongoing. Successful projects have reached pre-registration status and beyond. Discontinued projects are no longer active and may have been suspended or terminated before reaching the pre-registration phase for various reasons, including (but not limited to) lack of efficacy, clinical safety, developability and strategic considerations[12]. Both successful and discontinued projects are considered completed (see Supplementary information S1 (table) for a list of the projects analysed in this article). Finally, ongoing projects include currently running projects from discovery to phase III clinical trials.

The projects we evaluated have targeted over 2,400 therapeutic mechanisms for more than 1,400 indications in the past 20 years, accounting for a total of 15,101 unique mechanism–indication pairs (BOX 1). To evaluate the performance of mechanism–indication pairs, we sorted them into three categories based on the progress of associated drug projects: validated, unvalidated and emerging (BOX 1). A validated mechanism–indication pair can be characterized as being represented by at least one drug project at or beyond pre-registration status. Conversely, an unvalidated mechanism–indication pair has at least one discontinued project, without any projects yet reaching pre-registration status. Both validated and unvalidated pairs are considered as completed pairs, as opposed to emerging mechanism–indication pairs, for which all of the projects are ongoing and none have reached beyond phase III or have been discontinued. In this study, the 15,101 unique mechanism–indication pairs include 2,414 validated, 8,066 unvalidated and 4,621 emerging pairs (FIG. 1). Unvalidated pairs can be further categorized into two subgroups: active and inactive unvalidated pairs. Active unvalidated pairs ($n = 752$) have at least one ongoing project, in addition to any number of discontinued projects; they constitute <10% of the unvalidated pair space. Inactive unvalidated pairs ($n = 7,314$), on the other hand, have no currently ongoing projects; these pairs constitute the vast majority of unvalidated pairs.

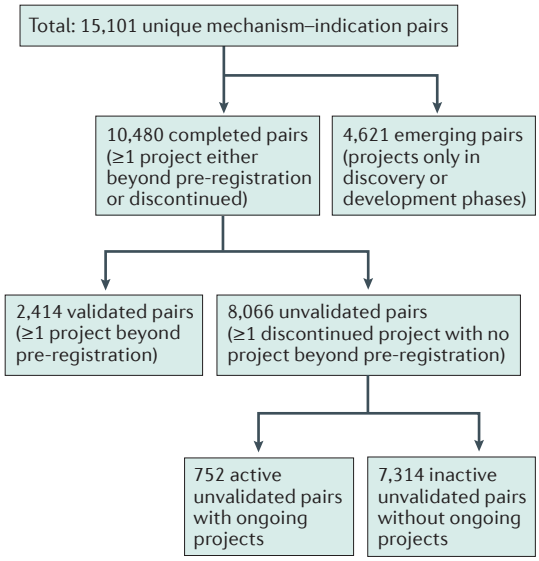


Figure 1 | **Categorization of mechanism–indication pairs.** The 15,101 mechanism–indication pairs in the data set for drug projects from the past two decades (see BOX 1 for details) were classified into three non-overlapping categories: validated, unvalidated and emerging pairs. A validated mechanism–indication pair is one that is represented by at least one drug project at or beyond pre-registration status. Conversely, an unvalidated mechanism–indication pair has at least one discontinued project, without any projects yet reaching pre-registration status. Unvalidated pairs were further classified as active if represented by ongoing projects or inactive for cases where all of the underlying projects have been discontinued. Both validated and unvalidated pairs are considered as completed pairs, whereas emerging mechanism–indication pairs are those for which all of the projects are ongoing and none have reached beyond phase III or have been discontinued.

Validated mechanism–indication pairs

We first explored the industry activity for validated mechanism–indication pairs (FIG. 2a,b). The validated pairs constitute just 16% of all mechanism–indication pairs. Among these validated mechanism–indication pairs, a large majority (74%, $n = 1,782$) correspond to singleton successful projects — that is, a single drug with a particular mechanism for a particular indication has reached pre-registration or beyond (FIG. 2a). Conversely, most of the successful projects correspond to a relatively small fraction of mechanism–indication pairs; of the successful projects ($n = 3,927$), more than half (55%, $n = 2,145$) account for roughly a quarter of the validated mechanism–indication pairs (26%, $n = 632$).

Pairs for which there are ≥5 successful projects (5% of the pairs, $n = 128$) account for 22% ($n = 880$) of the total successful projects overall. Some of these repeatedly validated mechanism–indication pairs represent areas of intense competition, with a number of companies advancing their drugs to the market over a fairly brief period of time (that is, the projects to develop the drugs were probably progressing roughly simultaneously, with similar uncertainty about the likelihood of success). For example, four HIV-1 protease inhibitors — saquinavir, ritonavir, indinavir and nelfinavir — were developed and approved within months of each other[25]. Others are the result of 'me-too' or second-generation or third-generation drug projects (progressing into clinical development after the mechanism–indication pair is considered sufficiently validated; for example, by the approval of a drug) that are pursued due to the lower R&D risk, often with the aim of improving efficacy, safety or pharmaceutical properties or for other commercial reasons. For example, the H^+/K^+-ATPase inhibitor esomeprazole demonstrated better efficacy over its predecessor omeprazole in disorders associated

with gastric acid[26], and the second generation of histamine H_1 receptor antagonists improved the safety profile for allergy-related indications[27]. Regardless, this fraction indicates that a significant proportion of the pharmaceutical industry's output in terms of new drugs (and probably also revenues) in the period studied has come from a small fraction of repeatedly validated pairs, indicating the challenge of identifying truly attractive pairs for which scientific possibility overlaps with clinical need and commercial potential.

As illustrated in FIG. 2a, there have nevertheless been a considerable number of failures seen for projects that pursued validated pairs, with 37% ($n = 2,335$) having been discontinued. In fact, 11% ($n = 263$) of the validated pairs

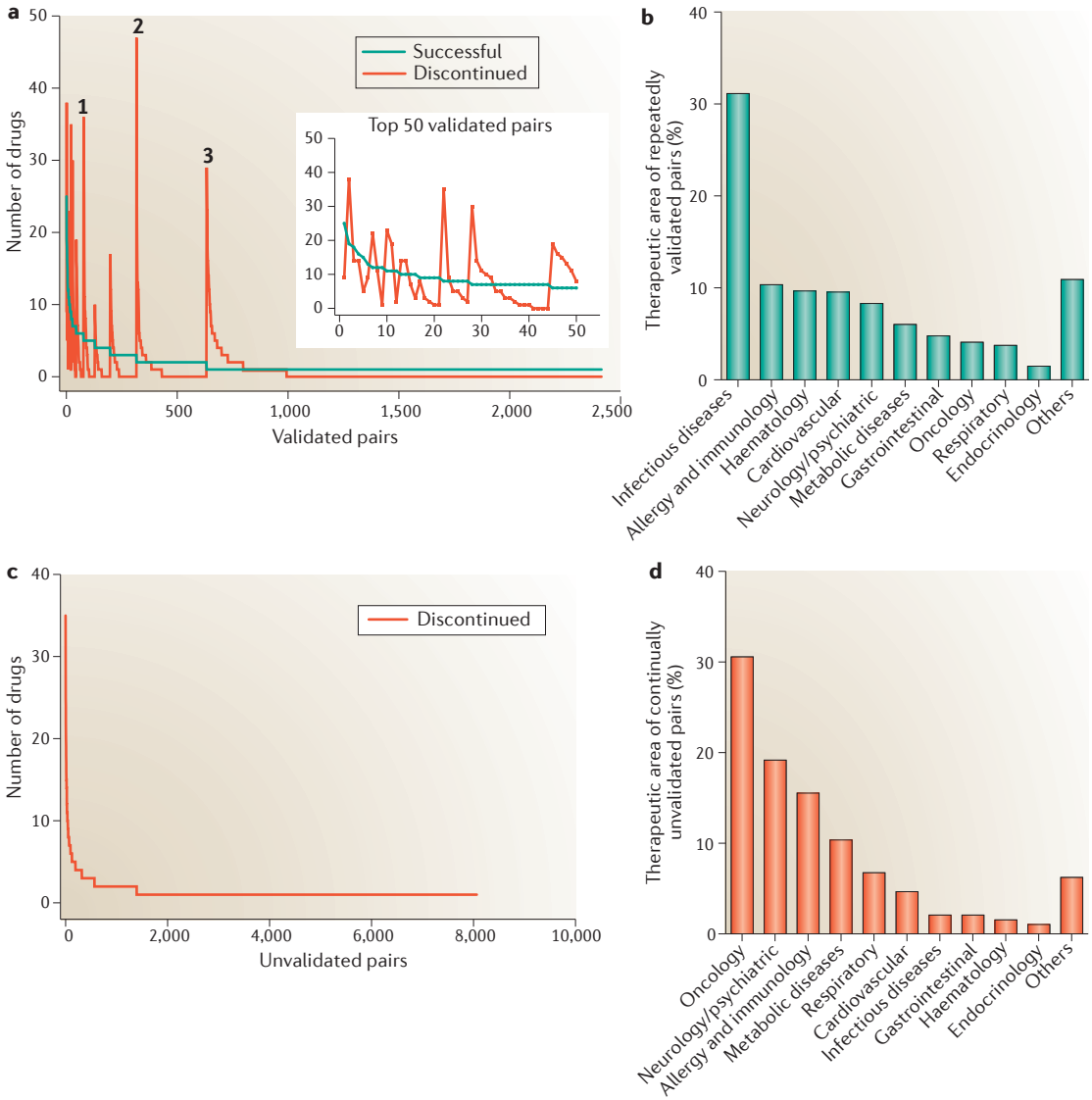


Figure 2 | **Overview of completed projects. a**,**b** | Successful and discontinued projects for the validated mechanism–indication pairs. The majority (74%, $n = 1,782$) of the 2,414 validated pairs correspond to successful singleton projects, and successful projects disproportionately correspond to a small fraction of mechanism–indication pairs. Over a third of projects ($n = 2,335$) for validated pairs were discontinued, and 11% ($n = 263$) of the validated pairs had more discontinued than successful projects. Part **a** shows the distribution of the number of successful and discontinued drug projects for the 2,414 validated pairs. Both successful (green) and discontinued (red) projects are shown, with projects ordered along the *x* axis based on the number of projects that were successful. This highlights various mechanism–indication pairs where many projects have failed relative to the number that have succeeded (red spikes). Angiotensin II receptor inhibition for hypertension (spike 1), hepatitis C virus (HCV) NS5B polymerase inhibition for HCV (spike 2), and 5-lipoxygenase inhibition for asthma (spike 3) are highlighted. The numbers of successful and discontinued projects for the top 50 validated pairs are shown in the insert. Part **b** shows the therapeutic area of indications for repeatedly validated therapeutic mechanism–indication pairs, defined as pairs with ≥5 projects reaching pre-registration status or higher. **c**,**d** | Overview of discontinued projects for the unvalidated mechanism–indication pairs. The majority (83%; $n = 6,671$) of the 8,066 unvalidated pairs correspond to singleton projects, and discontinued projects are disproportionately allocated to a small fraction of unvalidated mechanism–indication pairs. Part **c** shows the distribution of the number of discontinued drug projects for the 8,066 unvalidated pairs. Part **d** shows the therapeutic area of indications for unvalidated mechanism–indication pairs with ≥5 discontinued projects.

had more discontinued than successful projects; the examples of the angiotensin II receptor for hypertension, hepatitis C virus (HCV) NS5B polymerase for HCV infection and 5-lipoxygenase for asthma are indicated in FIG. 2a. This highlights the importance of other factors in addition to the target biology, including project-specific R&D issues (such as toxicity of a chemical class) or commercial reasons[12]. It is also apparent that the number of discontinued projects is not correlated with the number of successful projects within the validated pairs (FIG. 2a), implying that the nature of the R&D issues for each combination of indication and therapeutic mechanism can vary widely.

Among the 128 repeatedly validated pairs (those with ≥5 successful projects), the top three therapeutic areas (which account for half of the total) are infectious diseases (31%), immune diseases (10%) and haematology (10%) (FIG. 2b). Owing to the level of granularity with which infectious disease indications are defined in our data set (BOX 1), anti-infectives would top a list ranked by the number of successful projects, as antimicrobial mechanisms such as targeting penicillin-binding protein and DNA gyrase have proven highly successful in tackling a broad swathe of infectious pathogens, and thus pathogen-targeted pairs are separated from human-targeted pairs in TABLE 1, which lists the pairs with the highest numbers of successful projects and no failures. In some cases, the same mechanisms can be well translated into different indications based on the disease biology — for example, DNA gyrase inhibition is an effective mechanism for many indications associated with bacterial infection. Notably, the most successful human-targeted validated pairs listed in TABLE 1, such as H^+/K^+-ATPase inhibitors in gastrointestinal indications, represent the alignment of clear target biology, druggability and broad commercial opportunity.

Evidence of intense competition and overcrowding in the validated mechanism–indication pair space echoes earlier findings by Agarwal *et al.*[28], who in addition found that competition steadily increases by clinical development phase. It is worth pointing out the differences between our current study and that of Agarwal and colleagues. The underlying data used in the studies came from two different databases, and Agarwal *et al.* excluded non-human targets and targets with polypharmacology, whereas we included them (see BOX 1). Agarwal *et al.* analysed active projects with respect to targets as of 2013, whereas we evaluated both active and inactive projects for mechanism–indication pairs over the past 20 years. They found that the vast majority (88%) of clinically proven targets (analogous to a validated mechanism in our study) were being pursued by more than one company, with 64% being pursued by ≥5 companies, whereas our analysis showed that a portion of validated mechanisms had a disproportionately high number of successful projects. Despite the differences in data sets and methodology, both studies highlight a heavy distribution of drug discovery activities in the validated mechanism–indication pair space towards a small fraction of mechanisms.

Table 1 | **Validated mechanism–indication pairs with the highest number of successful projects and no failed projects**

Indication	Mechanism	Example drug
Pathogen-targeted pairs		
Streptococcus infection	Penicillin-binding protein inhibitor	Ertapenem
Haemophilus influenzae infection	DNA gyrase inhibitor	Ciprofloxacin
Moraxella catarrhalis infection	DNA gyrase inhibitor	Ciprofloxacin
Streptococcus infection	Topoisomerase IV inhibitor	Levofloxacin
Escherichia coli infection	Penicillin-binding protein inhibitor	Cefepime
Escherichia coli infection	DNA gyrase inhibitor	Moxifloxacin
Enterobacteriaceae infection	DNA gyrase inhibitor	Moxifloxacin
Haemophilus influenzae infection	Penicillin-binding protein inhibitor	Cefditoren
Haemophilus influenzae infection	Topoisomerase IV inhibitor	Levofloxacin
Moraxella catarrhalis infection	Topoisomerase IV inhibitor	Ciprofloxacin
Human-targeted pairs		
Duodenal ulcer	H^+/K^+-ATPase inhibitor	Omeprazole
Female infertility	FSH receptor agonist	Follitropin-α
Allergic conjunctivitis	H_1 receptor antagonist	Cetirizine
Angina	L-type calcium channel inhibitor	Nicardipine
Congenital afibrinogenaemia	Factor I stimulator	Fibrinogen
Ocular inflammation	COX inhibitor	Diclofenac
Bipolar I disorder	D_2 receptor antagonist	Olanzapine
Bipolar I disorder	5-HT_{2A} receptor antagonist	Olanzapine
Renal cell carcinoma	VEGFR inhibitor	Lenvatinib
Breast cancer	GnRH agonist	Goserelin

5-HT_{2A}, 5-hydroxytryptamine 2A; COX, cyclooxygenase; D_2 receptor, dopamine receptor type 2; FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; H_1 receptor, histamine receptor type 1; VEGFR, vascular endothelial growth factor receptor.

Unvalidated mechanism–indication pairs

Next, we explored the unvalidated mechanism–indication pairs; that is, pairs for which there were no drugs beyond pre-registration in the study period. Drug development projects may be discontinued for various reasons, including lack of efficacy, poor safety profile, regulatory issues, insufficient commercial opportunities and competition. Indeed, previous studies have reported that efficacy, safety and commercial and strategic reasons are responsible for varying fractions of the discontinued projects in different therapeutic areas[6,9,14]. Ultimately, regardless of the specific reasons behind project terminations, the resources invested in those projects did not realize a return on investment.

Most of the unvalidated pairs (83%, $n = 6{,}671$) are represented by a single discontinued project (FIG. 2c). Although we do not have enough information to discern whether these single discontinued projects were primarily terminated in the preclinical phase, some of the pairs may have been explored transiently without a strong scientific rationale. It is possible that others may benefit from further investigation. However, there are no ongoing projects at present for the vast majority of these singly unvalidated pairs (93%, $n = 6{,}235$).

Just 2% ($n = 193$) of the unvalidated pairs have ≥5 discontinued projects (FIG. 2c). However, similar to what we observed for validated pairs, we found that a disproportionate 14% ($n = 1{,}580$) of the total of 11,153 discontinued projects corresponded to these 193 continually unvalidated mechanism–indication pairs. This observation suggests that the pharmaceutical industry overall has tended to overcommit to certain mechanism–indication pairs, continuing efforts in the face of mounting evidence against the original rationale for the project. This may be because multiple companies have independently initiated projects pursuing the same therapeutic hypotheses based on the same information, which later turns out to be invalid, or because companies have initiated projects similar to their competitors.

The unvalidated pairs with the highest number of discontinued projects (TABLE 2) are split nearly equally between active and inactive unvalidated pairs. The inactive unvalidated pairs represent a group of therapeutic hypotheses for which the industry appears to have reached a negative conclusion based on the failures so far. Notable examples include acetyl-CoA acetyltransferase 1 (ACAT1) inhibitors in cardiovascular indications[29,30], NMDA (*N*-methyl-D-aspartate) receptor antagonists in cerebrovascular disease[31] and p38 mitogen-activated protein (MAP) kinase inhibitors in rheumatoid arthritis[32] (TABLE 2). Multiple trials for such pairs have been conducted without positive outcomes[33–39], representing a vast amount of resources that did not deliver a return on investment. In these

Table 2 | **Unvalidated mechanism–indication pairs with the most discontinued projects**

Indication	Mechanism	Number of discontinuations for pair	Example drug
Inactive unvalidated pairs with no ongoing projects			
Atherosclerosis	ACAT inhibitor	28	Pactimibe
Hypertension	K^+ channel stimulator	20	Emakalim
Cerebrovascular disease	NMDA receptor antagonist	19	Selfotel
Asthma	PAF receptor antagonist	17	Lexipafant
Obesity	β_3-adrenoceptor agonist	16	Talibegron
Alzheimer disease	M_1 receptor agonist	15	Alvameline
Hypertension	ET_A receptor antagonist	14	Darusentan
Rheumatoid arthritis	p38 MAP kinase inhibitor	14	Pamapimod
Asthma	NK_1 receptor antagonist	13	Nolpitantium
Psoriasis	5-LOX inhibitor	11	Lonapalene
Active unvalidated pairs with ongoing projects			
Alzheimer disease	Amyloid-β synthesis inhibitor	26	Semagacestat
Alzheimer disease	Amyloid-β deposition inhibitor	25	Tramiprosate
Alzheimer disease	Amyloid-β antagonist	24	Ponezumab
Non-insulin-dependent diabetes	PPAR α agonist	22	Aleglitazar
Alzheimer disease	Amyloid-β modulator	18	Lovastatin
Asthma	K^+ channel stimulator	15	Rilmakalim
Depression	$5\text{-}HT_{1A}$ receptor agonist	14	Naluzotan
Breast cancer	EGFR inhibitor	12	Vandetanib
Non-insulin-dependent diabetes	Glucokinase stimulator	11	Piragliatin
Pain	TRPV1 antagonist	11	MK-2295

General indications such as cancer and inflammatory disease were not included. $5\text{-}HT_{1A}$, 5-hydroxytryptamine 1A; 5-LOX, 5-lipoxygenase; ACAT, acetyl-CoA acetyltransferase 1; EGFR, epidermal growth factor receptor; ET_A, endothelin A; NK_1, neurokinin 1; NMDA, *N*-methyl-D-aspartate; PAF, platelet-activating factor; PPAR, peroxisome proliferator-activated receptor; TRPV1, transient receptor potential cation channel subfamily V member 1.

cases, the major reason for failure may have been the lack of translatability from strong therapeutic evidence in animal models into human disease therapeutics[40–42], suggesting the need to carefully evaluate the predictive power of such animal models[43].

By contrast, there are still ongoing projects for the active unvalidated pairs shown in TABLE 2, despite a high number of previously discontinued projects. Of note, several therapeutic approaches based on the amyloid hypothesis[44,45] in Alzheimer disease have failed, such as solanezumab recently[46], suggesting that the industry needs to rethink its strategies in this area[47].

Overall, despite the large body of scientific literature supporting the top unvalidated pairs, their performance in the context of pharmaceutical R&D has been disappointing, mainly due to lack of efficacy and side effects (for example, see REFS 48–51). The huge amount of resources invested in these pairs by the industry overall, with no return on investment in terms of new drugs, also suggests that challenging 'groupthink' and more critically questioning whether a given project really has a meaningful chance of success when multiple previous projects have failed could help reduce costly late-stage attrition.

Looking at the subset of unvalidated pairs with ≥5 discontinued projects, the largest proportions are in oncology (30%) and neurology/psychiatric-related therapeutic areas (19%) (FIG. 2d). This is unsurprising, given that both areas historically have had widely acknowledged challenges related to a lack of understanding of the disease biology and poorly predictive preclinical models[52–57]. However, at least for cancer, there have been substantial advances in more recent years that may lead to higher success rates if similar analyses were to be conducted in the future.

Success and failure for popular pairs

To gain an overview of the performance of mechanism–indication pairs for the most tested indications and mechanisms, we identified the indications ($n = 196$) for which ≥10 mechanisms have been tested and the mechanisms ($n = 297$) tested for ≥10 indications from the completed pairs. We constructed a combined heatmap to visualize these mechanism–indication pairs based on the success rate (that is, the ratio of the number of successful projects to the sum of successful and discontinued projects) for validated pairs and the number of discontinued projects for unvalidated pairs (FIG. 3 and Supplementary information S2 (figure)).

In total, there are 719 validated and 2,573 unvalidated mechanism–indication pairs in this group; that is, quite a small proportion have led to a successful drug so far. FIGURE 3 illustrates that there are several mechanism–indication pair clusters by indication, reflecting that the pharmaceutical industry has tested certain mechanisms over sets of related indications in the same disease area (FIG. 3a,b, inflammatory diseases; FIG. 3c, central nervous system diseases and FIG. 3d,e, oncology). The likely reason is the view that a drug that shows success for a particular indication in such areas could be expanded to related indications in the same disease area or that a drug that targets a mechanism that could be relevant to multiple indications in a broad area is often tested at an early stage in several indications, before the most promising one is selected to progress first towards the market (a common strategy in cancer indications). In some cases, this strategy has been hugely successful, such as for tumour necrosis factor (TNF) signalling blockers in inflammatory diseases[58] (FIG. 3a) and a number of receptor tyrosine kinases (RTKs) in oncology[59–61] (FIG. 3d). However, there are also notable examples of cases where the strategy has failed, such as 5-lipoxygenase and the leukotriene B4 (BLT) receptor in inflammatory diseases[62,63] (FIG. 3b), the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) receptor in neurology[64] (FIG. 3c) and insulin-like growth factor 1 receptor (IGF1R) signalling in oncology[65] (FIG. 3e). Overall, our interpretation of the historical data is that pursuing the same mechanism across multiple related indications has not generally been successful, except in those cases where there is a clear understanding of the underlying disease biology and the key role of the targeted mechanism. In addition to the examples of TNF blockers and some RTKs in cancer, this is also often the case for infectious diseases (for example, antibiotics in bacterial infections and interferon in viral infections).

Success variation for mechanism–indication pairs

Although validated target biology is critical to the likelihood of success for a drug development project, the importance of other factors such as druggability should not be underestimated. To explore this further, we identified indications ($n = 31$) corresponding to ≥10 different validated mechanisms. Data for a selected set of indications shown in FIG. 4 indicate a wide range of success rates observed for different mechanisms of action within a given indication.

In the case of HCV infection, both NS3 protease inhibition (for example, with simeprevir) and NS5B polymerase inhibition (for example, with sofosbuvir) are validated mechanisms. Developers of NS3 protease inhibitors, the first class of oral direct-acting antiviral drugs to make it to market for HCV, had to contend with poor active site druggability, leading to larger, less drug-like compounds[66], but nevertheless, a third of the completed projects (6 out of 19 in the study period) have resulted in marketed drugs (FIG. 4). NS5B polymerase inhibitors have been a game changer for HCV treatment, increasing cure rates, shortening treatment time and enhancing treatment safety, but this has been an even more difficult target due to toxicity and resistance issues[67], with a mere 2 out of a total of 49 completed projects having been successful in the study period (FIG. 4; TABLE 3).

Hypertension has been successfully targeted by multiple mechanisms over the years[68], such as L-type calcium channel inhibitors and angiotensin-converting enzyme inhibitors (FIG. 4), and multiple projects targeting these two mechanisms have been successful. By contrast, renin inhibition is an interesting and stark example of a validated mechanism corresponding to the approved drug aliskiren, where all but one (23 of 24) project failed (FIG. 4; TABLE 3), principally as a result of poor pharmacokinetic properties of the candidate drugs[68,69].

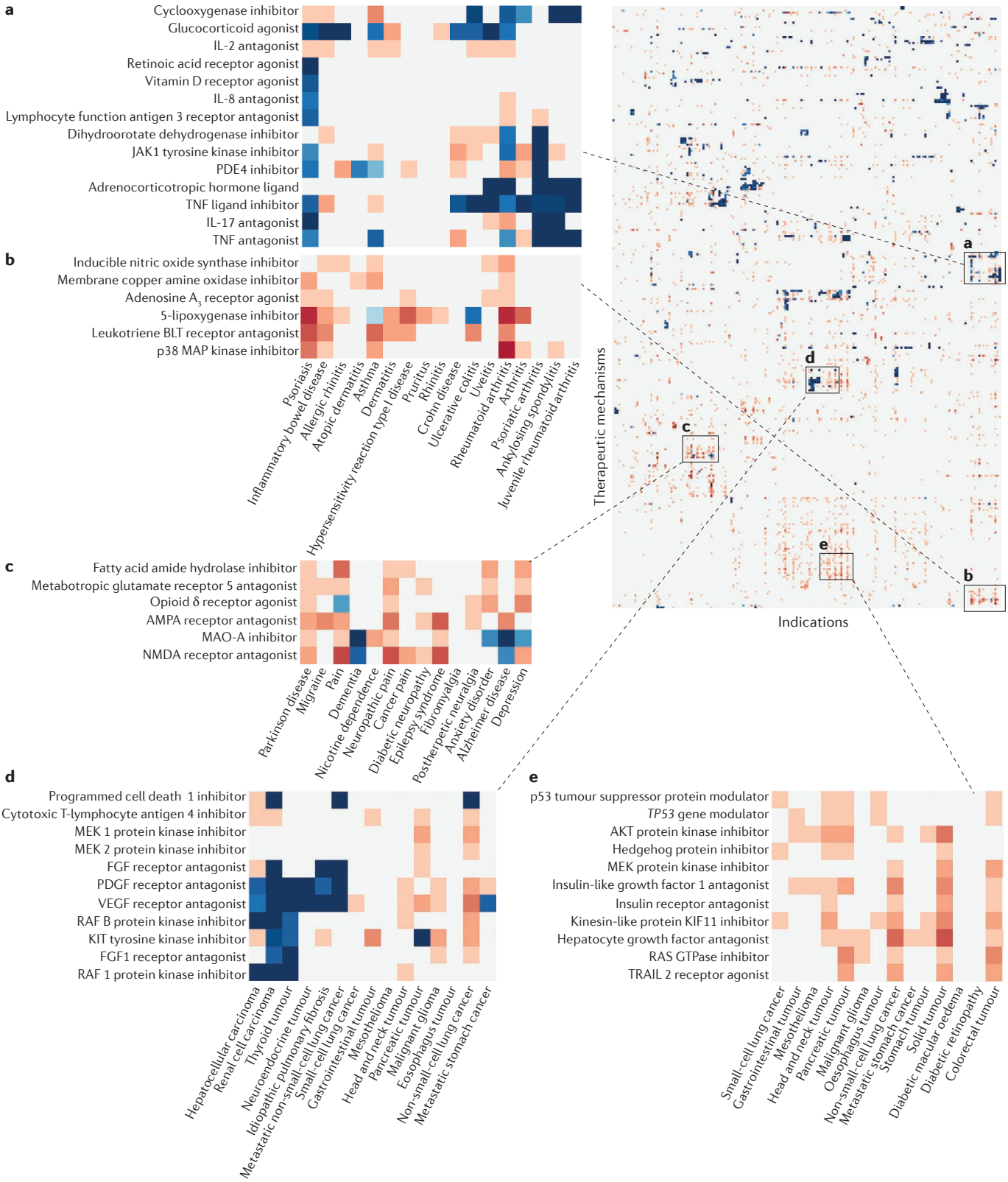


Figure 3 | **Combined drug discovery effectiveness heatmap for the most tested mechanism–indication pairs.** Validated pairs are shown in blue; darker shading reflects a higher success rate as measured by the ratio of the number of successful projects over the sum of the successful and discontinued projects. Unvalidated pairs are shown in red; darker shading represents a higher number of discontinued projects. Only those mechanisms tested in ≥10 indications and the indications tested by ≥10 mechanisms are included. There are 719 validated pairs (blue) and 2,573 unvalidated pairs (red). The transformed data were organized by hierarchical clustering of mechanisms and diseases based on cosine similarity. The insets detail pair clusters in inflammatory diseases (parts **a** and **b**), central nervous system diseases (part **c**) and oncology (parts **d** and **e**). FGF, fibroblast growth factor; JAK, Janus kinase; IL, interleukin; MAO-A, monoamine oxidase type A; MAP, mitogen-activated protein; NMDA, *N*-methyl-D-aspartate; PDE, phosphodiesterase; PDGF, platelet-derived growth factor; TNF, tumour necrosis factor; TRAIL, tumour necrosis factor-related apoptosis-inducing ligand; VEGF, vascular endothelial growth factor.

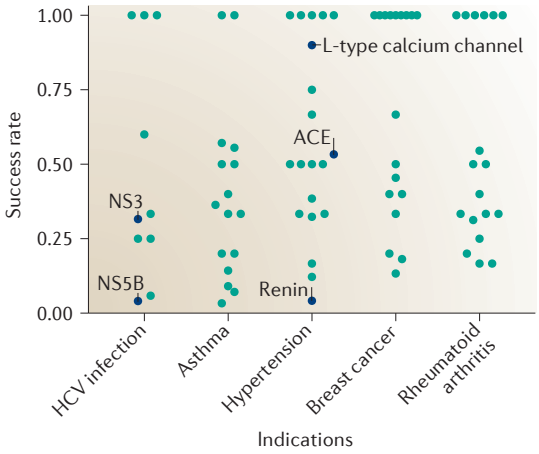


Figure 4 | **Success rate distribution for validated mechanisms in selected indications.** Indications that correspond to ≥10 validated therapeutic mechanisms were selected. Each dot in the figure represents a mechanism–indication pair. A wide range of success rates across different validated mechanisms illustrates how other factors beyond validated target biology can contribute to drug discovery success. Specific highlighted mechanisms include NS3 protease inhibitors and NS5B polymerase inhibitors for the treatment of hepatitis C virus (HCV) infection and L-type calcium channel inhibitors, angiotensin-converting enzyme (ACE) inhibitors and renin inhibitors for the treatment of hypertension.

Mechanisms with extremely low success rates, such as targeting renin or NS5B, frequently have mechanism-specific R&D issues and are largely responsible for driving down the overall R&D effectiveness for the validated pairs (TABLE 3). In other words, validated mechanisms are not all equal, with the success rate being a characteristic of the mechanism–indication pair rather than the indication per se (FIG. 4). There are cases when this latter distinction is blurred, such as in therapeutic areas like Alzheimer disease, for which projects have been unsuccessful for nearly all the mechanisms investigated[47,70]. However, the mechanism–indication pair perspective adds a useful granularity to the average therapeutic area success rate that is typically analysed (for example, as in REF. 10).

Next, we examined the performance of mechanisms by evaluating their corresponding validated and unvalidated indications, identifying the best performers (TABLE 4). It is not surprising that several antimicrobial mechanisms rank highly, in part due to the bias in the way indications are defined as discussed above (BOX 1; TABLE 1). However, we should not ignore the fact that better understanding of the microbiology and host–microorganism interactions makes these mechanisms particularly effective. Among the human-targeted mechanisms, therapies aimed at the endocrine and immune systems have been the most effective mechanisms to date, probably due to our understanding of the diseases relative to human physiological processes. For example, the physiological role of interferon-α in immunity makes its use an effective mechanism for antiviral therapy[71].

Lastly, we analysed the diseases for which the most therapeutic mechanisms have been tested (TABLE 5). It is reasonable to expect that diseases with complex aetiologies will have had a larger range of attempted mechanisms and that diseases that pose a particular burden to patients and health care systems would attract substantial R&D investment. Indeed, the diseases shown in TABLE 5 include cancers (for example, breast cancer), neurological diseases (for example, Alzheimer disease) and inflammatory diseases (for example, rheumatoid arthritis), which have all seen a steady rate of investment by the industry, including testing a variety of mechanisms, several of which have been validated and highly successful. For example, drugs with two different mechanisms (targeting TNF and CD20) for rheumatoid arthritis currently constitute several of the top-selling drugs in the US. Nevertheless, the majority of tested mechanisms in these diseases have been less fruitful, with ~10 times or more mechanisms being unvalidated than validated in most cases.

Table 3 | **Validated mechanism–indication pairs with the lowest success rates**

Indication	Mechanism	Success rate*	Example of successful drug
Asthma	5-LOX inhibitor	1/30	Zileuton
Thrombosis	Factor IIa antagonist	1/28	Bivalirudin
HCV infection	NS5B polymerase inhibitor	2/49	Sofosbuvir
Hypertension	Renin inhibitor	1/24	Aliskiren
HCV infection	NS5A inhibitor	1/17	Daclatasvir
Anxiety	5-HT_{1A} receptor agonist	1/16	Buspirone
HIV infection	CCR5 antagonist	1/15	Maraviroc
Asthma	PDE4 inhibitor	2/28	Doxofylline
Schizophrenia	5-HT_{1A} receptor agonist	1/13	Brexpiprazole
Thrombosis	Factor Xa antagonist	1/13	Bemiparin
Non-insulin-dependent diabetes	PPARγ agonist	2/23	Rosiglitazone
Asthma	Immunoglobulin E antagonist	1/11	Omalizumab

General indications such as cancer, bacterial infection and inflammatory disease were not included; pairs with a success rate <10% are shown. 5-HT_{1A}, 5-hydroxytryptamine 1A; 5-LOX, 5-lipoxygenase; CCR5, CC chemokine receptor 5; HCV, hepatitis C virus; PDE, phosphodiesterase; PPAR, peroxisome proliferator-activated receptor. *Success rate is measured by the ratio of the number of the successful projects (reaching pre-registration and beyond) to the number of successful and discontinued projects.

Table 4 | **Mechanisms with the highest number of validated and lowest number of unvalidated indications**

Mechanisms	Examples of indications
Pathogen-targeted mechanisms	
DNA gyrase inhibitor	Various conditions caused by bacterial infection
Topoisomerase IV inhibitor	Various conditions caused by bacterial infection
PBP inhibitor	Various conditions caused by bacterial infection
Lanosterol 14 α-demethylase inhibitor	Various fungal infections
Bacterial RNA polymerase inhibitor	Conditions caused by *Escherichia coli*, *Mycobacterium* and *Clostridium* infection
Human-targeted mechanisms	
IFNα2 ligand	Melanoma, hairy cell leukaemia, Kaposi sarcoma and certain viral infections
β_1-adrenoceptor antagonist	Angina, atrial fibrillation, hypertension, glaucoma and myocardial infarction
H^+/K^+-ATPase inhibitor	Gastroesophageal reflux disease, duodenal and gastric ulcers, *Helicobacter pylori* infection and Zollinger–Ellison syndrome
GnRH agonist	Breast cancer, prostate cancer, endometriosis, uterine fibroids and precocious puberty
IgG	Immunodeficiency, Kawasaki disease, Guillain–Barre syndrome and chronic lymphocytic leukaemia

GnRH, gonodotrophin-releasing hormone; IFNα2, interferon α2; IgG, immunoglobulin G; PBP, penicillin-binding protein.

Success rates for orphan diseases

In recent years, owing to factors such as the exclusivity provisions in the US Orphan Drug Act legislation and a series of highly successful orphan drug launches, the industry has turned greater attention to rare diseases, which were once neglected and thought to be commercially non-viable. We were interested in determining whether the pharmaceutical industry performed better in this subset of diseases in the period studied. Based on data from the National Organization for Rare Disorders, we were able to identify over 1,200 rare diseases. Of these, just 221 rare diseases were listed in the Cortellis database, and 173 of them had completed projects. For non-rare diseases, we found 1,039 indications with completed projects (FIG. 5a). Comparing the distribution of the success rate for mechanism–indication pairs indicates that the success rate for rare disease projects is slightly better compared with non-rare diseases, with a higher interquartile range and higher mean for the success rate (26% versus 19%) (FIG. 5b).

One likely explanation is that many of the rare diseases for which drug development has been pursued so far have a clear disease biology, such as well-defined genetic causes. Genetic evidence supporting a particular target and/or mechanistic hypothesis has been highlighted for its potential to serve as a guide in target selection[72], and a recent study of approved drugs showed that drug mechanisms with genetic support have higher chances of success[73]. Another factor that could have improved the likelihood of success in rare diseases in the period studied is the lack of any existing therapies for many of the diseases, which means there could be a lower bar for demonstrating a clinically useful treatment effect, compared with developing a new drug for an indication for which several classes of therapy already exist, such as type 2 diabetes or hypertension. On the other hand, rare disease drug development also often has particular challenges, including identifying and validating appropriate clinical end points if no approved therapies exist and recruiting patient cohorts of suitable size and homogeneity.

Characteristics of ongoing projects

We next investigated if ongoing R&D projects (that is, those at the discovery and phase I, II and III stages) are following the trend from the past 20 years, in terms of a bias towards either repeatedly validated mechanism–indication pairs (≥5 projects having reached pre-registration or beyond) or continually unvalidated pairs (≥5 discontinued projects), as seen in FIG. 2. For those indications with repeatedly validated mechanisms, 103 ongoing projects are targeting the same repeatedly validated mechanisms and 128 ongoing projects are targeting other validated mechanisms (that is, with at least one project that has reached pre-registration or beyond), for a total of 231 out of 1,241 total projects for these indications that are targeting validated mechanisms (FIG. 6a); that is, ~8% of ongoing projects in these indications are targeting repeatedly validated mechanisms and ~18% are targeting validated mechanisms, with presumably a lower risk of failure due to lack of efficacy or safety but potentially challenges in achieving sufficient differentiation from the existing competition. Of the 1,010 projects not targeting validated mechanisms, 315 target unvalidated mechanisms (those with at least one previous failed project but no successful projects) and 695 target emerging mechanisms (those for which there have not yet been any successful or failed projects).

In addition, for those indications with validated mechanisms (that is, at least one project that has reached pre-registration or beyond), we found that 642 out of

a total of 5,424 ongoing projects (12%) are targeting a validated mechanism. Thus, 4,782 (88%) of all ongoing projects in such indications are exploring mechanisms for which no project has yet reached pre-registration or beyond, representing a greater risk; of these, 17% of projects are pursuing unvalidated therapeutic mechanisms and 71% are pursuing emerging mechanisms. This indicates a strong commitment by the industry to the identification of drugs with substantially better efficacy and/or safety than existing therapies for these indications, with little emphasis on 'me-too' or follow-on drugs. This may reflect the need in the current environment to develop new drugs that are not just effective and safe but also clinically differentiated from existing therapies, in order to justify reimbursement when they reach the market.

A further 1,080 projects were pursuing novel mechanisms in indications that had no projects that had reached pre-registration or beyond (see also the discussion below on novel mechanism–indication pairs).

In the indications with continually unvalidated mechanisms, we were interested in the impact of pre-existing data on choices for therapeutic mechanisms to pursue. It would be expected that the multiple failures of projects for continually unvalidated pairs would alert the industry and prompt it to focus attention elsewhere. For example, there were 14 discontinued p38 MAP kinase inhibitor projects for rheumatoid arthritis in the period studied (TABLE 2), with no project advancing to or beyond pre-registration due to mediocre efficacy[32], and indeed, there were no such ongoing projects in our data set.

Table 5 | **Diseases with most therapeutic mechanisms tested**

Indication	Example of mechanism (drug)		
	Validated	**Unvalidated**	**Emerging**
Breast cancer	HER2 inhibitor (trastuzumab)	PI3K inhibitor (apitolisib)	PD1 inhibitor (nivolumab)
Rheumatoid arthritis	TNF inhibitor (infliximab)	p38 MAP kinase inhibitor (pamapimod)	CD40 inhibitor (CFZ-533)
Non-small-cell lung cancer	EGFR inhibitor (gefitinib)	HGFR inhibitor (tivantinib)	HDAC inhibitor (entinostat)
Asthma	β_2-adrenoceptor agonist (formoterol)	IL-13 inhibitor (anrukinzumab)	IL-23 inhibitor (BI-655066)
Alzheimer disease	Acetylcholinesterase inhibitor (tacrine)	Amyloid-β antagonist (ponezumab)	FYN inhibitor (saracatinib)
Colorectal cancer	VEGFA inhibitor (bevacizumab)	HDAC inhibitor (vorinostat)	CD49b inhibitor (E-7820)
Pancreatic cancer	EGFR inhibitor (erlotinib)	Thymidylate synthase inhibitor (plevitrexed)	FAK inhibitor (defactinib)
Psoriasis	IL-17 inhibitor (secukinumab)	5-LOX inhibitor (lonapalene)	STAT3 inhibitor (MOL-4249)
Pain	COX2 inhibitor (celecoxib)	TRPV1 inhibitor (MK-2295)	TRPA1 inhibitor (GRC-17536)
Multiple myeloma	Proteasome inhibitor (bortezomib)	CD40 inhibitor (dacetuzumab)	CD19-specific CAR T cell therapy (tisagenlecleucel-T)
Prostate cancer	GnRH agonist (leuprolide)	DPYD inhibitor (eniluracil)	SERCA inhibitor (mipsagargin)
Ovarian cancer	PARP inhibitor (olaparib)	AKT inhibitor (afuresertib)	CTLA4 inhibitor (ipilimumab)
Melanoma	PD1 inhibitor (nivolumab)	Tubulin inhibitor (rhizoxin)	PDL1 inhibitor (atezolizumab)
Multiple sclerosis	S1P receptor agonist (fingolimod)	IL-12/IL-23 inhibitor (ustekinumab)	Lingo 1 inhibitor (opicinumab)
Cerebrovascular disease	Plasminogen activator, tissue type (alteplase)	NMDA receptor (perzinfotel)	MASP2 inhibitor (OMS-721)
Acute myeloid leukaemia	Aminopeptidase inhibitor (ubenimex)	CSF1R inhibitor (pexidartinib)	CXCR4 inhibitor (BL-8040)
Non-insulin-dependent diabetes	DPP4 inhibitor (sitagliptin)	β_3-adrenoceptor agonist (talibegron)	Apolipoprotein C3 inhibitor (volanesorsen)
Obesity	GLP1 agonist (liraglutide)	CB_1 receptor antagonist (drinabant)	MTTP inhibitor (Slx-4090)
Head and neck cancer	EGFR inhibitor (nimotuzumab)	KIF11 inhibitor (ispinesib)	BTK inhibitor (acalabrutinib)
Parkinson disease	D_2 receptor agonist (cabergoline)	AMPA receptor antagonist (perampanel)	NAD(P)H oxidoreductase modulator (vatiquinone)

General indications such as inflammatory diseases, cancer and cardiovascular diseases were not included. 5-LOX, 5-lipoxygenase; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor, BTK, Bruton tyrosine kinase; CAR, chimeric antigen receptor; CB_1 receptor, cannabinoid receptor type 1; COX, cyclooxygenase; CSF1R, colony stimulating factor 1 receptor; CXCR4, CXC chemokine receptor type 4; CTLA4, cytotoxic T lymphocyte-associated protein 4; D_2 receptor, dopamine receptor type 2; DPP4, dipeptidyl peptidase 4; DPYD, dihydropyrimidine dehydrogenase; EGFR, epidermal growth factor receptor; FAK, focal adhesion kinase; GLP1, glucagon-like peptide; GnRH, gonadotropin-releasing hormone; HDAC, histone deacetylase; HER2, human epidermal growth factor receptor 2; HGFR, hepatocyte growth factor receptor; IL, interleukin; KIF, kinesin family member 11; Lingo 1, leucine rich repeat and immunoglobin-like domain-containing protein 1; MAP, mitogen-activated protein; MASP2, mannan-binding lectin serine protease 2; MTTP, microsomal triglyceride transfer protein; NMDA, *N*-methyl-D-aspartate; PARP, poly(ADP ribose) polymerase; PD1, programmed cell death 1; PDL1, PD1 ligand 1; PI3K, phosphoinositide 3-kinase; S1P, sphingosine 1 phosphate; SERCA, sarco/endoplasmic reticulum calcium ATPase; STAT3, signal transducer and activator of transcription 3; TNF, tumour necrosis factor;TRPA1, transient receptor potential cation channel A1; TRPV1, transient receptor potential cation channel subfamily V member 1; VEGFA, vascular endothelial growth factor A.

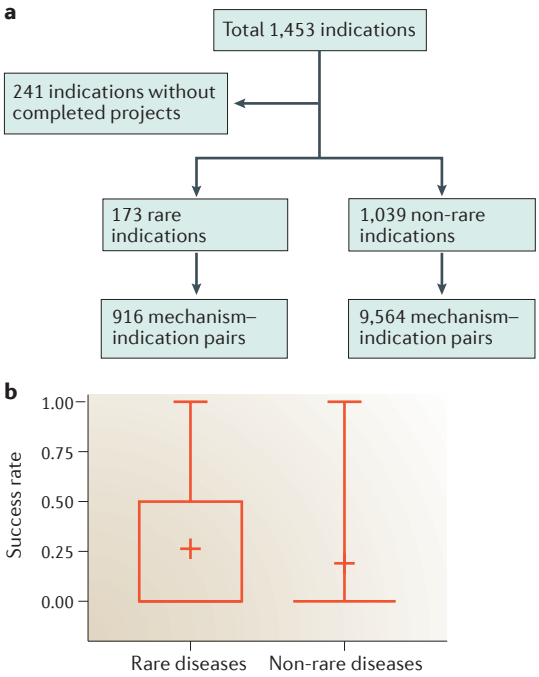


Figure 5 | **Comparison of success rates for rare and non-rare diseases. a** | To analyse the relative success rates for rare and non-rare diseases, a subset of the data for completed projects was processed and compared as shown in the flow chart. **b** | The box plot shows the success rate distributions for 916 mechanism–indication pairs in 173 rare diseases compared with 9,564 pairs in 1,039 non-rare diseases. The mean values of success rate in rare diseases (26%) and non-rare diseases (19%) are labelled with a +. The success rates do not follow the normal distribution, and thus a non-parametric statistical analysis, the Kolmogorov–Smirnov test, was used to assess the difference, which is statistically significant ($P < 0.0001$).

However, we still identified 151 (8%) ongoing projects corresponding to continually unvalidated mechanisms, compared with 1,853 projects in all mechanisms for the same indications (FIG. 6b). Unless these projects are able to adequately address the issues that plagued their predecessors, they probably face a higher probability of discontinuation later in the R&D process.

Overall, our analysis of ongoing projects suggests a positive trend in the industry, while leaving room for improvement. A strong focus on testing novel therapeutic hypotheses in the diseases with well-validated therapeutic mechanisms is a highly encouraging trend, indicating that the industry is trying hard to improve on current therapies. Conversely, there is an indication that the industry could improve its success rates overall by re-evaluating ongoing projects in mechanisms that have already repeatedly failed and by reducing such investments unless there is a compelling rationale that the issues responsible for these earlier failures can be addressed.

Changing target space

For the repeatedly validated or continually unvalidated mechanism–indication pairs discussed above, we should be able to learn lessons from earlier projects. But what about the novel cases where we do not have much prior knowledge, such as when a new mechanism is proposed for an indication or when mechanisms are explored for new indications? We identified such emerging mechanism–indication pairs that do not have any projects that have advanced beyond phase III or that have been discontinued (FIG. 1; BOX 1). These pairs constitute 80% ($n = 4,621$) of total mechanism–indication pairs ($n = 5,782$) currently in development, and 75% of currently ongoing projects map to these pairs, reinforcing the earlier finding by Agarwal *et al.*[28] that the majority of ongoing projects are testing new therapeutic ideas.

Further investigation of these emerging therapeutic hypotheses indicates that there has been a substantial shift in the target classes being pursued now versus those completed in the 20-year period we analysed (FIG. 7a). Whereas in the completed projects, G protein-coupled receptors (GPCRs) and non-kinase enzymes were the best-represented drug target classes, GPCRs account for a significantly lower percentage of the emerging ongoing projects, and there are more projects for emerging pairs targeting kinases and cytokine signalling. This observation suggests that the industry has been moving into different, sometimes less conventional, target classes. To gauge whether this shift might be related to success or failure with drugging certain target classes of targets, we looked at the success rate of target classes for the top 100 mechanisms ranked by the most completed projects. We found that most target classes have success rates of ~15–30%, except non-host targets, which have a much greater success rate of ~70% (FIG. 7b). Thinking speculatively, the popularity of cytokine signalling targets among mechanisms tested in emerging ongoing projects might be considered to be supported by the success of historical projects (35%), but conversely, the low success rate (16%) for historical kinase projects does not seem to have deterred efforts to pursue kinases in the ongoing projects. Furthermore, non-host targets are not being heavily pursued compared with other target classes despite a high historical success rate, probably due in part to the challenging economics for antibacterial R&D[74]. Overall, the target class shift seen in emerging ongoing projects in FIG. 7a appears to be largely unrelated to historical success rates for a given target class. Other possible reasons could include recent advances in human disease biology (for example, in immuno-oncology and cancer genomics) and the growing availability of a broader armamentarium of treatment modalities (including monoclonal antibodies and other recombinant proteins, nucleic acid-based therapeutics and viral vectors), which have expanded the 'druggable target space' compared with traditional small-molecule drugs alone. Only time will tell how well these novel mechanistic hypotheses progress in the clinic, but it is encouraging that the industry is willing to invest substantially in innovative strategies (as also highlighted in REF. 28).

Discussion and summary

In summary, we analysed the past 20 years of drug project history with the aim of understanding more about how the pharmaceutical industry has been performing

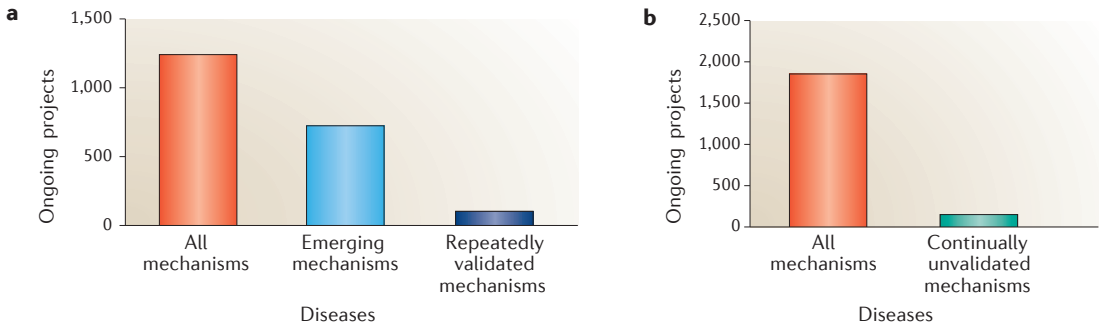


Figure 6 | **Ongoing projects for the repeatedly validated and continually unvalidated mechanism–indication pairs.** **a** | Ongoing projects for repeatedly validated mechanisms (defined as those that have ≥5 successful projects at or beyond pre-registration status) constitute 8% of total ongoing projects in the same indications, with those targeting validated mechanisms (those that have at least one successful project) constituting ~18% of the total for these indications. So, >80% of projects in these indications are targeting unproven — and thus potentially riskier — mechanisms. **b** | Ongoing projects for continually unvalidated mechanisms (defined as those that have ≥5 discontinued projects) constitute 8% of total ongoing projects in those same indications. These projects probably have a lower chance of success later in development.

with regard to therapeutic mechanisms and their intended indications. Although restricted to a single database (Cortellis), the amount and breadth of data provides a good industry-wide reference point. Our analysis suggests that industry output in terms of successful projects in this period has come primarily from a limited set of well-validated therapeutic mechanisms. There is no doubt a place for second-generation or 'me-too' drugs. The former may bring important advances in efficacy, safety or ease of use, while the latter serve to increase competition and ultimately lower the price of drugs. However, our analysis indicates that current pipelines are increasingly focusing on more innovative mechanisms. This shift has presumably been catalysed by factors such as the growing need to demonstrate meaningful clinical advances over existing approaches in order to achieve commercial success, as well as regulatory incentives provided for projects considered to have this potential, such as the US Food and Drug Administration's breakthrough designation programme and the European Medicine Agency's PRIME programme. We confirm that the industry has had slightly greater success in rare disease indications than in those for non-rare diseases (as also noted in REF. 13), probably owing to the clear monogenic origins of many of the rare diseases investigated so far. This notion, coupled with the favourable economics of orphan drug R&D[75], has spurred a boom in drug discovery and development for rare diseases[76,77].

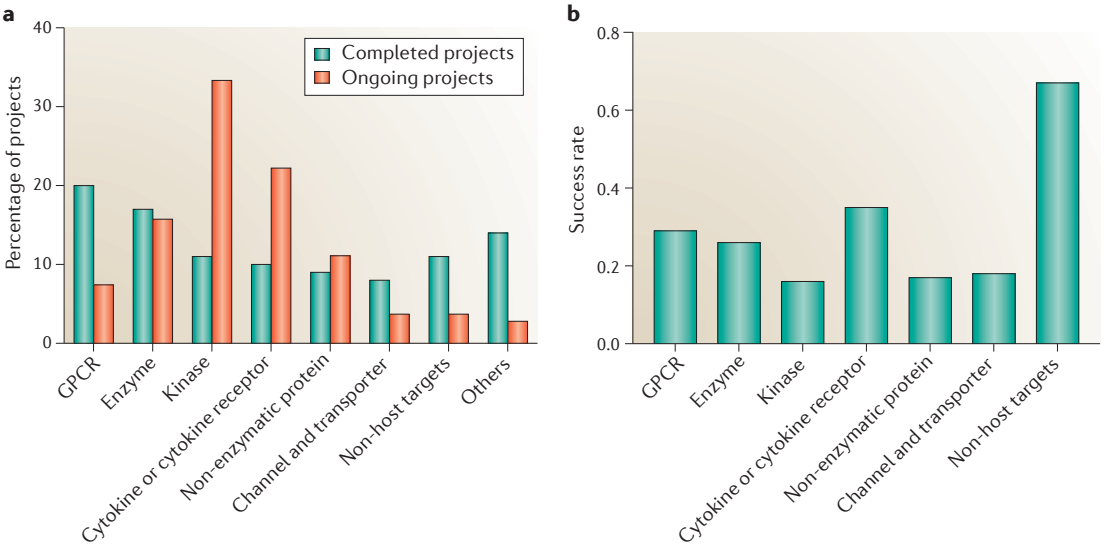


Figure 7 | **Target class distribution for completed versus emerging ongoing projects.** Most therapeutic mechanisms can be classified into target classes: enzyme (non-kinase enzymes), G protein-coupled receptor (GPCR), cytokine or cytokine receptor, kinase, non-enzymatic protein (for example, signalling molecules, transcription factors, cell-surface molecules and structural proteins), channel or transporter, and non-host targets. **a** | Target classes of the top 100 mechanisms with the highest number of completed projects (successful and discontinued) and the target classes of the top 100 mechanisms with the most ongoing projects for the emerging mechanism–indication pairs. The target classes for the emerging mechanism–indication pairs are substantially different from those in the completed mechanism–indication pairs. **b** | Success rate of target classes for the top 100 mechanisms with the highest number of completed projects.

Our analysis also highlights inefficiencies in the industry due to continued investment in frequently discontinued therapeutic mechanisms. It indicates that the industry could benefit from paying more attention to lessons learned from other projects and avoiding initiating projects for previously studied failed therapeutic mechanisms without rigorous and independent validation. In general, we suggest that the analysis of new potential projects, whether internal (early research) or external (business development), through the prism of the mechanism–indication pair formalism could help in learning lessons from historical data. A significant shortcoming of this retrospective analysis is that it is not possible to identify novel fruitful therapeutic mechanism–indication pairs a priori, but we have started working on methods that would integrate existing data and extend knowledge to a forward-looking model for prioritization of novel mechanism–indication pairs.

Finally, our analysis indicates that the majority of ongoing projects are pursuing novel mechanism–indication pairs, even in the indications with existing therapeutics. This vigorous testing of new therapeutic hypotheses by the pharmaceutical industry is highly encouraging. It is ultimately our hope that many of them will turn out to be successful therapeutics, driving the effectiveness of future drug discovery.

1. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–715 (2004).
2. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
3. Empfield, J. R. & Leeson, P. D. Lessons learned from candidate drug attrition. *IDrugs* **13**, 869–873 (2010).
4. DiMasi, J. A., Feldman, L., Seckler, A. & Wilson, A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin. Pharmacol. Ther.* **87**, 272–277 (2010).
5. Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* **10**, 428–438 (2011).
6. Arrowsmith, J. Trial watch: phase II failures: 2008–2010. *Nat. Rev. Drug Discov.* **10**, 328–329 (2011).
7. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).
8. Ringel, M., Tollman, P., Hersch, G. & Schulze, U. Does size matter in R&D productivity? If not, what does? *Nat. Rev. Drug Discov.* **12**, 901–902 (2013).
9. Arrowsmith, J. & Miller, P. Trial watch: phase II and phase III attrition rates 2011–2012. *Nat. Rev. Drug Discov.* **12**, 569 (2013).
10. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
11. Peck, R. W., Lendrem, D. W., Grant, I., Lendrem, B. C. & Isaacs, J. D. Why is it hard to terminate failing projects in pharmaceutical R&D? *Nat. Rev. Drug Discov.* **14**, 663–664 (2015).
12. Waring, M. J. *et al.* An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486 (2015).
13. Smietana, K., Siatkowski, M. & Møller, M. Trends in clinical success rates. *Nat. Rev. Drug Discov.* **15**, 379–380 (2016).
14. Harrison, R. K. Phase II and phase III failures: 2013–2015. *Nat. Rev. Drug Discov.* **15**, 817–818 (2016).
15. Kassel, D. B. Applications of high-throughput ADME in drug discovery. *Curr. Opin. Chem. Biol.* **8**, 339–345 (2004).
16. Wang, J. Comprehensive assessment of ADMET risks in drug discovery. *Curr. Pharm. Des.* **15**, 2195–2219 (2009).
17. Wishart, D. S. Improving early drug discovery through ADME modelling: an overview. *Drugs R D* **8**, 349–362 (2007).
18. De Buck, S. S. *et al.* Prediction of human pharmacokinetics using physiologically based modeling: a retrospective analysis of 26 clinically tested drugs. *Drug Metab. Dispos.* **35**, 1766–1780 (2007).
19. Jones, H. M., Gardner, I. B. & Watson, K. J. Modelling and PBPK simulation in drug discovery. *AAPS J.* **11**, 155–166 (2009).
20. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922 (2012).
21. Segall, M. D. & Barber, C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discov. Today* **19**, 688–693 (2014).
22. Blomme, E. A. G. & Will, Y. Toxicology strategies for drug discovery: present and future. *Chem. Res. Toxicol.* **29**, 473–504 (2016).
23. Prentis, R. A., Lis, Y. & Walker, S. R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies. *Br. J. Clin. Pharmacol.* **25**, 387–396 (1988).
24. Bunnage, M. E. Getting pharmaceutical R&D back on target. *Nat. Chem. Biol.* **7**, 335–339 (2011).
25. Deeks, S. G., Smith, M., Holodniy, M. & Kahn, J. O. HIV-1 protease inhibitors. A review for clinicians. *JAMA* **277**, 145–153 (1997).
26. Olbe, L., Carlsson, E. & Lindberg, P. A proton-pump inhibitor expedition: the case histories of omeprazole and esomeprazole. *Nat. Rev. Drug Discov.* **2**, 132–139 (2003).
27. Simons, F. E. R. Advances in H1-antihistamines. *N. Engl. J. Med.* **351**, 2203–2217 (2004).
28. Agarwal, P., Sanseau, P. & Cardon, L. R. Novelty in the target landscape of the pharmaceutical industry. *Nat. Rev. Drug Discov.* **12**, 575–576 (2013).
29. Burnett, J. R. & Huff, M. W. Cholesterol absorption inhibitors as a therapeutic option for hypercholesterolaemia. *Expert Opin. Investig. Drugs* **15**, 1337–1351 (2006).
30. Stoekenbroek, R. M., Kastelein, J. J. P. & Hovingh, G. K. Recent failures in antiatherosclerotic drug development. *Curr. Opin. Lipidol.* **24**, 459–466 (2013).
31. Ikonomidou, C. & Turski, L. Why did NMDA receptor antagonists fail clinical trials for stroke and traumatic brain injury? *Lancet Neurol.* **1**, 383–386 (2002).
32. Ho, L.-J. & Lai, J.-H. Small-molecule inhibitors for autoimmune arthritis: success, failure and the future. *Eur. J. Pharmacol.* **747**, 200–205 (2015).
33. Nissen, S. E. *et al.* Effect of ACAT inhibition on the progression of coronary atherosclerosis. *N. Engl. J. Med.* **354**, 1253–1263 (2006).
34. Meuwese, M. C. *et al.* ACAT inhibition and progression of carotid atherosclerosis in patients with familial hypercholesterolemia: the CAPTIVATE randomized trial. *JAMA* **301**, 1131–1139 (2009).
35. Damjanov, N., Kauffman, R. S. & Spencer-Green, G. T. Efficacy, pharmacodynamics, and safety of VX-702, a novel p38 MAPK inhibitor, in rheumatoid arthritis: results of two randomized, double-blind, placebo-controlled clinical studies. *Arthritis Rheum.* **60**, 1232–1241 (2009).
36. Cohen, S. B. *et al.* Evaluation of the efficacy and safety of pamapimod, a p38 MAP kinase inhibitor, in a double-blind, methotrexate-controlled study of patients with active rheumatoid arthritis. *Arthritis Rheum.* **60**, 335–344 (2009).
37. Davis, S. M. *et al.* Selfotel in acute ischemic stroke: possible neurotoxic effects of an NMDA antagonist. *Stroke* **31**, 347–354 (2000).
38. Albers, G. W., Goldstein, L. B., Hall, D. & Lesko, L. M. Aptiganel hydrochloride in acute ischemic stroke: a randomized controlled trial. *JAMA* **286**, 2673–2682 (2001).
39. Sacco, R. L. *et al.* Glycine antagonist in neuroprotection for patients with acute stroke: GAIN Americas: a randomized controlled trial. *JAMA* **285**, 1719–1728 (2001).
40. Heinonen, T. M. Inhibition of acyl coenzyme A-cholesterol acyltransferase: a possible treatment of atherosclerosis? *Curr. Atheroscler. Rep.* **4**, 65–70 (2002).
41. Hoyte, L., Barber, P. A., Buchan, A. M. & Hill, M. D. The rise and fall of NMDA antagonists for ischemic stroke. *Curr. Mol. Med.* **4**, 131–136 (2004).
42. McNamee, K., Williams, R. & Seed, M. Animal models of rheumatoid arthritis: how informative are they? *Eur. J. Pharmacol.* **759**, 278–286 (2015).
43. Denayer, T., Stöhr, T. & Van Roy, M. Animal models in translational medicine: validation and prediction. *New Horiz. Transl Med.* **2**, 5–11 (2014).
44. Karran, E., Mercken, M. & De Strooper, B. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat. Rev. Drug Discov.* **10**, 698–712 (2011).
45. Soejitno, A., Tjan, A. & Purwata, T. E. Alzheimer's disease: lessons learned from amyloidocentric clinical trials. *CNS Drugs* **29**, 487–502 (2015).
46. Le Couteur, D. G., Hunter, S. & Brayne, C. Solanezumab and the amyloid hypothesis for Alzheimer's disease. *BMJ* **355**, i6771 (2016).
47. Selkoe, D. J. The therapeutics of Alzheimer's disease: where we stand and where we are heading. *Ann. Neurol.* **74**, 328–336 (2013).
48. M. U.I. R., K. Glutamate-based therapeutic approaches: clinical trials with NMDA antagonists. *Curr. Opin. Pharmacol.* **6**, 53–60 (2006).
49. Langmead, C. J., Watson, J. & Reavill, C. Muscarinic acetylcholine receptors as CNS drug targets. *Pharmacol. Ther.* **117**, 232–243 (2008).
50. Pal, P., Gandhi, H., Giridhar, R. & Yadav, M. R. ACAT inhibitors: the search for novel cholesterol lowering agents. *Mini Rev. Med. Chem.* **13**, 1195–1219 (2013).
51. Karran, E. & Hardy, J. A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Ann. Neurol.* **76**, 185–205 (2014).
52. Pangalos, M. N., Schechter, L. E. & Hurko, O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat. Rev. Drug Discov.* **6**, 521–532 (2007).
53. McGonigle, P. Animal models of CNS disorders. *Biochem. Pharmacol.* **87**, 140–149 (2014).
54. Pankevich, D. E., Altevogt, B. M., Dunlop, J., Gage, F. H. & Hyman, S. E. Improving and accelerating drug development for nervous system disorders. *Neuron* **84**, 546–553 (2014).
55. Hait, W. N. Anticancer drug development: the grand challenges. *Nat. Rev. Drug Discov.* **9**, 253–254 (2010).
56. Gould, S. E., Junttila, M. R. & de Sauvage, F. J. Translational value of mouse models in oncology drug development. *Nat. Med.* **21**, 431–439 (2015).
57. Zitvogel, L., Pitt, J. M., Daillère, R., Smyth, M. J. & Kroemer, G. Mouse models in oncoimmunology. *Nat. Rev. Cancer* **16**, 759–773 (2016).

58. Cessak, G. *et al.* TNF inhibitors — mechanisms of action, approved and off-label indications. *Pharmacol. Rep.* **66**, 836–844 (2014).
59. Traxler, P. Tyrosine kinases as targets in cancer therapy — successes and failures. *Expert Opin. Ther. Targets* **7**, 215–234 (2003).
60. Gschwind, A., Fischer, O. M. & Ullrich, A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat. Rev. Cancer* **4**, 361–370 (2004).
61. Gross, S., Rahal, R., Stransky, N., Lengauer, C. & Hoeflich, K. P. Targeting cancer with kinase inhibitors. *J. Clin. Invest.* **125**, 1780–1789 (2015).
62. Pergola, C. & Werz, O. 5-Lipoxygenase inhibitors: a review of recent developments and patents. *Expert Opin. Ther. Pat.* **20**, 355–375 (2010).
63. Di Gennaro, A. & Haeggström, J. Z. Targeting leukotriene B4 in inflammation. *Expert Opin. Ther. Targets* **18**, 79–93 (2014).
64. Lee, K. *et al.* AMPA receptors as therapeutic targets for neurological disorders. *Adv. Protein Chem. Struct. Biol.* **103**, 203–261 (2016).
65. King, H., Aleksic, T., Haluska, P. & Macaulay, V. M. Can we unlock the potential of IGF-1R inhibition in cancer therapy? *Cancer Treat. Rev.* **40**, 1096–1105 (2014).
66. Goudreau, N. & Llinàs-Brunet, M. The therapeutic potential of NS3 protease inhibitors in HCV infection. *Expert Opin. Investig. Drugs* **14**, 1129–1144 (2005).
67. Gentile, I., Coppola, N., Buonomo, A. R., Zappulo, E. & Borgia, G. Investigational nucleoside and nucleotide polymerase inhibitors and their use in treating hepatitis C virus. *Expert Opin. Investig. Drugs* **23**, 1211–1223 (2014).
68. Jadhav, M., Yeola, C., Zope, G. & Nabar, A. Aliskiren, the first direct renin inhibitor for treatment of hypertension: the path of its development. *J. Postgrad. Med.* **58**, 32–37 (2012).
69. Jensen, C., Herold, P. & Brunner, H. R. Aliskiren: the first renin inhibitor for clinical treatment. *Nat. Rev. Drug Discov.* **7**, 399–410 (2008).
70. Graham, W. V., Bonito-Oliva, A. & Sakmar, T. P. Update on Alzheimer's disease therapy and prevention strategies. *Annu. Rev. Med.* **68**, 413–430 (2017).
71. Antonelli, G., Scagnolari, C., Moschella, F. & Proietti, E. Twenty-five years of type I interferon-based treatment: a critical analysis of its therapeutic use. *Cytokine Growth Factor Rev.* **26**, 121–131 (2015).
72. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
73. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
74. Renwick, M. J., Simpkin, V. & Mossialos, E. *Targeting innovation in antibiotic drug discovery and development. The need for a One Health – One Europe – One World Framework* (WHO, 2016).
75. Meekings, K. N., Williams, C. S. M. & Arrowsmith, J. E. Orphan drug development: an economically viable strategy for biopharma R&D. *Drug Discov. Today* **17**, 660–664 (2012).
76. Braun, M. M., Farag-El-Massah, S., Xu, K. & Cotē, T. R. Emergence of orphan drugs in the United States: a quantitative assessment of the first 25 years. *Nat. Rev. Drug Discov.* **9**, 519–522 (2010).
77. Melnikova, I. Rare diseases and orphan drugs. *Nat. Rev. Drug Discov.* **11**, 267–268 (2012).
78. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).

Acknowledgements
Information reported in this article is derived from Cortellis Competitive Intelligence, a database produced and owned by Clarivate Analytics. For further information about Cortellis and Clarivate Analytics, please use the following link: https://clarivate.com/products/cortellis/cortellis-competitive-intelligence/. Clarivate Analytics will not be liable for any inaccuracy in the information provided or the way in which it is used by any reader of this article. The authors gratefully thank G. McGaughey and J. Come for helpful comments in the course of manuscript preparation. This work is not funded by Clarivate Analytics, and the authors are solely responsible for the analysis and conclusions put forward in it.

Competing interests statement
The authors declare competing interests: see Web version for details.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF