FINDING THE DRUGS OF TOMORROW:
THE DRUG DEVELOPMENT PIPELINE

Thomas Nieuwenstein[1]

[1]Master's student Biomedical Sciences, Radboud university medical center, Nijmegen, The Netherlands

Abstract

Developing novel drugs and therapies is a labour-intensive, costly, and tightly regulated process spanning multiple years. As students of the medical faculty will likely interact with this process during or after their studies, this article is aimed at giving a general overview of the drug development pipeline and explains the individual phases in more detail. In the drug development pipeline, researchers examine their disease of interest to find a druggable target. Following this, massive compound libraries are screened for hits that are able to influence the target. These drug candidates are often tested *in vitro* and *in vivo* to establish pharmacokinetics, off-target effects, and proof-of-concept efficacy. Subsequently, the drug candidates are studied in more detail, optimised, and prepared for the clinical phases. In the first couple of clinical phases, the emphasis is on safety. In phase I, extensive safety testing is performed in a small group of healthy participants. In phase II, the efficacy of the drug is tested in a small group of patients with the disease of interest while further examining safety. In phase III, the efficacy is assessed in a larger group of patients with the disease of interest, and more rare side effects are studied. Following a successful, positive phase III trial, the novel drug can be submitted for regulatory approval. If the drug is approved, phase IV focuses on all data gathered from the drug prescriptions to all patients. These data may indicate rare side effects not found in earlier trials or indicate a different efficacy compared to what was found in the controlled clinical trial settings. The regulatory approval may be rescinded if the drug proves to have too little efficacy or too many severe side effects outside of the study setting.

Introduction

As the quality of global health increases, so does the need for novel, efficacious drugs and therapies. Older drugs can be improved, current diseases may develop treatment resistance, or new diseases can surface at any time. When the need for a new drug arises, a multi-step, multi-year, and multi-million-dollar cascade must be traversed before a novel drug can enter the market. From time to time, older ideas can be refined or expanded upon, shortening the process. However, from start to finish, the development of a new drug can take over 15 years, and the mean estimate of costs is around US$2.6 billion per approved drug, although newer studies report a decreasing mean estimate of costs of around US$1 billion per approved drug [1, 2].

The drug development pipeline consists of a preclinical development phase, in which a target is defined and a possible drug candidate is selected, and a clinical development phase, in which a promising drug candidate is tested in humans [1]. Both phases are subdivided into individual, more detailed phases [1]. Due to the high sunk costs into potential drug candidates, failure in a later phase of development proves to be extremely costly. As such, drug development companies will put in great effort to only select the most favourable candidates for the development of a new drug. Potential candidates are optimised as much as possible before moving into human trials.

The drug development pipeline

Preclinical development

During the preclinical development of a novel therapy, the goal is to find a suitable target which is causally involved in the disease of interest and to subsequently find a way to influence this target to combat the disease (Table 1). As the name "preclinical development"

Table 1: *The simplified goals of preclinical phases.*

Preclinical phase	**Goal of the phase**
Target identification	To obtain a list of targets that may be causally related to the disease of interest.
Target validation	To verify that the target is causally related to the disease of interest.
Lead discovery	To test many compounds or methods to attempt to influence the target.
Lead optimisation	To gather all information on the lead, improve the safety and efficacy of the lead, and prepare it for clinical phases.

suggests, this is all performed without applying the therapy in humans. *In vitro* assays on cell lines, disease models, or patient material combined with *in vivo* animal disease models and *in silico* tools are used to find a suitable target and subsequently find a therapy that can influence the target. At the end of the preclinical development, information on safety, toxicology, and efficacy has to be collected from previously mentioned models, and if proven to be favourable enough, the development will continue into the clinical phase. The therapy is finalised and ready to be applied in a human setting. The preclinical development is the shorter of the two, taking roughly five years and 30% of all costs associated with developing a new drug [3].

Target identification

The first step in the drug development pipeline consists of defining a health problem or disease. The goal of the target identification is to identify key molecules or processes which are associated with the disease. However, not all diseases may have molecules or processes which are known to directly influence the disease burden, or these cannot be influenced through external intervention (yet). Target identification requires intricate knowledge of the disease aetiology and may require additional research before a suitable target can be found.

Omics techniques, such as genomics and proteomics, can offer valuable insight into a disease phenotype if used on patient material. Applying several omics techniques on patient materials and combining the results will lead to a large list of factors that are altered, such as upregulation or downregulation of genes or unexpected metabolites. These "changes" can be relevant to the disease aetiology. However, not all of these factors may be related to the disease of interest. Narrowing the list down to a few promising, disease-causing factors is key. By constructing the disease aetiology using the literature and experimental results, key molecules or processes can be pinpointed. These can be cross-referenced with the results from the disease phenotype analysis to yield interesting targets. Once a shortlist of possible targets is constructed, these can move to the target validation phase.

Target validation

During the target validation, the goal is to establish proof that the target truly is a disease-causing factor. Causal relations can be proven if 1) the target is present in the diseased state and 2) influencing the target influences the disease condition. For example, the function of the enzyme angiotensin-converting enzyme (ACE) is to convert angiotensin I into the vasoconstrictive angiotensin II [4]. The activity of this enzyme has been linked to high blood pressure and heart failure [4]. A method to prevent or at least inhibit this conversion would lead to a decrease of angiotensin II, and as a result, a decrease in blood pressure. On the other hand, increasing ACE activity will increase the prevalence of angiotensin II, increasing blood pressure and worsening the disease state. In this case, ACE activity is linked to the diseased state; increasing its activity directly increases the disease burden, and preventing its activity relieves disease burden, making ACE activity a causative factor and thus a suitable target to combat high blood pressure and heart failure [4]. The target validation makes use of patient material and is partially performed *in vitro*, but currently often requires *in vivo* animal studies to confirm whether the selected target is truly disease-causing.

Lead discovery

Once the target is defined, therapy can be designed to influence the target. Note that, depending on the type of target (e.g., genetic, receptor, enzyme, etc.), different types of therapies can be considered (Table 2).

Genetic defects

In the case of genetic defects, numerous approaches can be taken to influence the impact of the genetic defect, such as trying to repair the gene using genomic editing techniques or by influencing the genetic product if no suitable approach can be taken to repair the gene. Before designing a gene-repairing therapy, it is important to thoroughly investigate the genetic defect itself. How often does it occur? What kind of mutation is it (e.g., missense, nonsense, or frameshift)? What is the canonical function of the gene? Can this defect be amended later in life, or is there a specific window of opportunity? Most of these questions can be answered by research performed in the target identification and target validation phases, but knowing the inner workings of the genetic defect at hand is key. Knowledge of the genetic defect can render one approach more suitable in comparison to another (e.g., supplementation of the correct genetic product may be a more suitable approach than attempting to restore the "correct" gene in case of a gene whose function is relatively unknown). Depending on which approach is taken, the process of lead discovery is different.

Table 2: *A few examples of approaches that can be taken to find a hit, depending on the type of target.*

Type of target	**Examples of therapy methods**
Genetic	Gene repair techniques; Supplementation of correct genetic product (in case of loss-of-function of genetic product); Inhibition of the diseased genetic product (in case of gain-of-function genetic product)
Enzymatic	Exogenous enzyme supplementation (in case of loss-of-function); Enzyme inhibition (in case of gain-of-function);
Receptor	Receptor agonists (if receptor must be activated); Receptor antagonists (if receptor must be inactivated)

If gene repair appears to be the most promising route, several types of delivery vectors can be used. A delivery vector is a method to get the correct genetic material to the correct place. Viruses such as adenoviruses or herpesviruses are versatile vectors as they can deliver a relatively large genetic load but cannot integrate this into the host DNA, causing only temporary expression [5]. However, other viruses, such as retroviruses and lentiviruses, can insert a smaller genetic load into the host DNA using reverse transcriptase and integrase enzymes [5]. Viruses are unique in terms of which cells they can infect, how effective they are at delivering the genetic load, and how they must be prepared. Newer techniques of gene editing, such as the CRISPR-Cas9 complex, can also be used. However, as the potential effectivity and preparation for each technique are different, it is necessary to choose a specific method of gene-editing early in the process.

The lead discovery for gene-editing of genetic defects consists mostly of testing different genetic codes of the vector. In the case of CRISPR-Cas9-mediated gene editing, this consists of testing multiple sequences of guide RNA.

If gene repair is not an option, downstream elements can be targeted. For example, the transcribed mRNA can be intercepted and degraded, the formed product can be inhibited, or a functional version of the genetic product can be supplemented.

Non-genetic defects

In the case of non-genetic defects, therapies are focused on affecting the target directly. This is often in the form of small molecules or biologicals. Long lists of known compounds, called compound libraries, are screened to see whether any of the existing compounds interact with the target in a process called high throughput screening (HTS). Thousands of compounds are screened for interactions with the target. Any compounds that are found to interact with the target in the initial HTS are tested again to confirm their activity using a positive selection and a negative selection. Once a compound is confirmed as active against the target, its selectivity for the target is tested, and a dose-response curve for the target is constructed. If the compound passes this test, it is considered a "hit".

Finally, after a promising hit is selected, it is further tested in *in vitro* systems. Data on cellular toxicity and genotoxicity is gathered, off-targets are tested, and a primitive absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile is generated. If these appear favourable, the hit is considered a "lead" and can be tested in *in vivo* systems. Testing the lead *in vivo* gives valuable information on the characteristics of the compound in a systemic setting. New off-targets can be found, unexpected metabolites can be analysed, toxicities can be discovered, and an *in vivo* ADMET profile can be generated. Additionally, this will also serve as the first proof-of-concept that the drug or therapy can achieve the desired goal.

Lead optimisation

The lead optimisation focuses on improving the safety and toxicology of the lead. If the lead is to ever become a successful therapy for humans, the benefit versus risk ratio must be favourable. In the lead optimisation phase, *in vitro* assays, cellular disease models, and *in vivo* disease models are used to ascertain the efficacy and the safety of the lead. The lead optimisation is aimed at finding out everything there is to know about the lead, amplifying the positives, mitigating the negatives, and creating the optimal ADMET profile. For example, the compound may be coated or slightly altered to improve the oral bioavailability, or the route of administration may be changed to reduce off-target exposure. At the end of the lead optimisation, the therapy is ready to be tested in humans. During the lead optimisation, the lead is tested in similar doses and routes of administration as planned in humans. Generally, after the lead optimisation, no more changes are made to the molecular contents of the lead.

The dose administered to the animal test subjects in the *in vivo* disease models at which no adverse effects were observed is used as a basis for human dosing in the clinical phases [6]. Interspecies and intraspecies safety factors are applied to reduce the dose to a level that is thought to be safe for humans. Generally, the intraspecies factor is 10, while the interspecies factor depends on which animal model was used in the *in vivo* safety testing [6].

Clinical development

The clinical phases of drug development are aimed at finding out whether a treatment is safe to use and if it has the intended effect in humans (Table 3). Next to this, the dosing regimen is also established and confirmed. The clinical development of a therapy is generally the most time consuming and costly part of the drug development pipeline, taking up to 10 years and 70% of all costs [3]. The following sections describe the clinical development phases of a "classic" drug candidate, a drug candidate to be used for relatively common diseases. The clinical phases can be different in the clinical development of "orphan" drugs or therapies, drugs that would not be profitable to develop without government funding due to the rarity

Table 3: *The simplified goals of the clinical phases.*

Clinical phase	**Goal of the phase**
Phase I	To establish whether the therapy is safe to use in humans in a small group.
Phase II	To establish whether there is an effect of the therapy on the disease of interest in a small group. Includes further safety assessment.
Phase III	To assess the effectiveness of the therapy and possibly compare it to other therapies in a large study group. Includes further safety assessment.
Phase IV/post-marketing surveillance	To continually check for the safety of using the therapy in all patients using the therapy.

of the treated condition. Specific immunotherapy for a rare cancer or gene repair therapy for a rare, severe hereditary disease falls under the orphan drug development. For example, phases I and II can sometimes be performed simultaneously, and market approval may be easier to attain in these cases.

Phase I

Once the green light is given for human testing, phase I can begin. Phase I is focused on finding information regarding the safety of systemic exposure to the drug or therapy in humans. The subjects in phase I are usually healthy subjects without the disease of interest unless the therapy is known to actively harm healthy subjects, such as new chemotherapeutic drugs. In that case, a small patient group is selected. Phase I usually uses between 20-50 test subjects.

The primary outcomes of phase I studies are focused on parameters regarding the ADMET profile of the compound in humans. Next to this, phase I studies often include a dose-escalation. Dose-escalation means that a fraction of the study subjects is exposed to increasing doses of the drug to assess the safety of higher than intended exposures. For example, while overexposure to antibiotic Vancomycin will almost certainly kill off a bacterial infection, it may also cause significant nephrotoxicity. The dose-escalation is divided into phase Ia (single dose escalation) and phase Ib (multi-dose escalation). The dose-escalation allows for dose-related toxicities to be found and an estimate of the optimal dose to be selected. This optimal dose is the dose at which the positive effects of the treatment maximally outweigh the negative effects of the treatment. This optimal dose is confirmed in later phases and will become the therapeutic dose if the drug gains market approval.

Phase I lasts around one and a half years, with estimated success rates of up to 65% [7].

Phase II

If a drug or therapy is proven to be safe in phase I, phase II can commence. For some drug candidates that have a high expected toxicity, phase I and phase II are combined. In phase II, the goal is to demonstrate the efficacy of the drug or therapy. Next to this, additional safety data is gathered. The results of a phase II trial will determine the sample size, expected effect size, time schedule, and other parts of the following phase III trial. Phase II uses a hundred to several hundred participants with the disease of interest. Phase II is also used to establish the final, optimal dosing regimen, based on the safety data gathered from the phase I dose escalation.

Even though the phase may not seem that complicated, only 33% of compounds pass from phase II to phase III, with phase II lasting around three years [7]. The failure to pass phase II can often be attributed to interspecies and intraspecies differences, rendering the novel drug or therapy not effective for human subjects.

Phase III

Phase III is the most critical phase of the clinical phase. Phase III trials, sometimes referred to as "pivotal trials", include hundreds to thousands of patients to confirm the dosing regimen, the safety, and the efficacy of the novel therapy or drug [8]. Phase III trials are incredibly expensive due to the large number of participants and the monitoring involved. Costs may even be driven up if the novel therapy or drug is compared to the current golden standard, as this requires additional study participants. If the novel therapy or drug is proven to be effective at treating the disease and is considered safe (enough) to use in patients, the trial can be considered a success. Following a successful phase III trial, the researchers can submit a request for regulatory approval in the tested population to the regulatory agencies. The European Medicines Agency (EMA, for approval in the European Union) and the Food and Drug Administration (FDA, for approval in the United States of America) carefully review all the evidence gained from the preclinical and clinical phases. If the novel therapy or drug is deemed to be efficacious and safe, approval for marketing will be granted. The novel therapy or drug can now be prescribed for the indicated disease in the indicated population.

Phase III trials take the longest, taking a median of 3.8 years to completion, with a success rate varying between 50% and 60% [7].

The applicability of the drug or therapy can be extended to other diseases or other populations (e.g., some antivirals that were approved for treatment against HIV also display efficacy for treating hepatitis B), but these require new trials to prove the safety and efficacy profile in the novel population.

Phase IV/Post-marketing surveillance

After a new drug or therapy is approved, more patients gain access to the product. This can be seen as the final "trial". As more patients are exposed to the drug or therapy, much more data will become available regarding the safety and efficacy of the product. It is possible for new effects – positive and negative – to only show up after the product has already gained regulatory approval. The manufacturer must keep collecting data regarding the safety of their product and periodically submit this data to the regulatory authorities, who will act accordingly. If severe health effects are found, or a drug with severe side effects is found to not be as efficacious as thought after the regulatory approval, the regulatory authorities can withdraw marketing approval.

The most well-known cases are likely the recall of approval for diethylstilbestrol (DES) and the recall of thalidomide (Softenon). DES is a synthetic oestrogen approved to prevent miscarriage and other pregnancy-related complications, but the use of DES resulted in clear cell adenocarcinomas and breast cancer not only in the women who took the drug but also in their daughters and possibly their granddaughters [9]. Prescribing DES to pregnant women was finally banned in the USA in 1971 and in 1978 in Europe [9]. Thalidomide was approved for use against anxiety and morning sickness [10]. However, following a series of cases in which severe birth defects were observed, thalidomide was withdrawn from the European market in 1961 [10]. Both incidents had a large impact on the way clinical trials and drug approval are handled today. Pregnant women are now a special population for which a new drug or therapy must gain separate regulatory approval. The post-marketing surveillance system is a direct result of the thalidomide effects.

Discussion

The drug development pipeline is a long and meticulous process. Each individual section of the pipeline is important and plays its own part in bringing a new drug to the market. However, the pipeline is changing. With novel SARS-CoV-2 vaccines achieving emergency approval for clinical testing and bringing a working vaccine from an idea to the public in as little as 11 months, it is clear to some that the pipeline can be altered [11]. The massive gain in time-to-product in the development of COVID-19 vaccines came mostly from running clinical trials in multiple phases back-to-back. Additionally, the preclinical development was also significantly reduced, as most vaccines were developed using prior research performed for SARS-CoV or other viral infections [12]. However,

the vaccine manufacturers must continue to submit data to the regulatory authorities to prove long time safety and efficacy. This will continue until the manufacturers have fully completed the regular clinical phase studies.

Parts of the preclinical development are also up for debate, as the voices of animal rights activists grow louder every year. Currently, animals are used in multiple steps of the preclinical development, such as disease models to confirm the target in the target validation phase or in tests for the systemic exposures in the lead discovery phase. *In vitro* systems such as organs- or humans-on-a-chip models or *in silico* modelling could serve as a suitable replacement for animal testing [13, 14]. However, it is not clear yet whether these systems can model the real human situation sufficiently. As about only 10% of all leads currently make it through the clinical phases, using new systems must at least be as good as the old systems and preferably provide a benefit in predictive value for the human situation. Only time will tell how the drug development pipeline will develop in the future and what this will mean for research and researchers alike.

Acknowledgements

RAMS would like to thank prof. dr. R.W.J. Collin, professor at the Department of Human Genetics at the RadboudUMC Nijmegen, and Arbaaz, BSc., for providing the author with feedback.

References

1. Mohs, R.C. & Greig, N.H. Drug discovery and development: Role of basic biological research. *Alzheimers Dement (N Y)* **3**, 651-657 (2017).
2. Wouters, O.J.*, et al.* Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844-853 (2020).
3. Harrer, S.*, et al.* Artificial Intelligence for Clinical Trial Design. *Trends Pharmacol Sci* **40**, 577-591 (2019).
4. Messerli, F.H.*, et al.* Angiotensin-Converting Enzyme Inhibitors in Hypertension: To Use or Not to Use? *J Am Coll Cardiol* **71**, 1474-1482 (2018).
5. Bouard, D.*, et al.* Viral vectors: from virology to transgene expression. *Br J Pharmacol* **157**, 153-165 (2009).
6. Nair, A.B. & Jacob, S. A simple practice guide for dose conversion between animals and human. *J Basic Clin Pharm* **7**, 27-31 (2016).
7. Wong, C.H.*, et al.* Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273-286 (2019).
8. Umscheid, C.A.*, et al.* Key concepts of clinical trials: a narrative review. *Postgrad Med* **123**, 194-204 (2011).
9. Giusti, R.M.*, et al.* Diethylstilbestrol revisited: a review of the long-term health effects. *Ann Intern Med* **122**, 778-788 (1995).
10. Lenz, W. A short history of thalidomide embryopathy. *Teratology* **38**, 203-215 (1988).
11. Pearson, H. How COVID broke the evidence pipeline. *Nature* **593**, 182-185 (2021).
12. Padron-Regalado, E. Vaccines for SARS-CoV-2: Lessons from Other Coronavirus Strains. *Infect Dis Ther*, 1-20 (2020).
13. Festing, S. & Wilkinson, R. The ethics of animal research. Talking Point on the use of animals in scientific research. *EMBO Rep* **8**, 526-530 (2007).
14. Pinero, J.*, et al.* In silico models in drug development: where we are. *Curr Opin Pharmacol* **42**, 111-121 (2018).