

Development of Human Target Validation Classification that Predicts Future Clinical Efficacy

 Cecilia Karlsson, Peter J. Greasley, David Gustafsson, and Karin Wähländer

Cardiovascular, Renal and Metabolism Translational Medicine Unit, Early Clinical Development, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden (C.K., P.J.G.); Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden (C.K.); Emeriti Pharma, AZ Bioventure Hub, Gothenburg, Sweden (D.G.); and KW Translational Medicine AB, Västra Frölunda, Sweden (K.W.)

Received May 24, 2018; accepted November 13, 2018

ABSTRACT

Fewer new medicines have become available to patients during the last decades. Clinical efficacy failures in late-phase development have been identified as a common cause of this decline. Improved ways to ensure early selection of the right drug targets when it comes to efficacy is therefore a highly desirable goal. The aim of this work was to develop a strategy to facilitate selection of novel targets already in the discovery phase that later on in clinical development would demonstrate efficacy. A cross-functional team at AstraZeneca with extensive experience in drug discovery and development participated in several workshops to identify the critical elements that contribute to building human target validation [(HTV); the relevance of the target from a

human perspective]. The elements were consolidated into a 10-point HTV classification system that was ranked from lowest to highest in terms of perceived impact on future clinical efficacy. Using 50 years of legacy research and development data, the ability of the 10-point HTV classification to predict future clinical efficacy was evaluated. Drug targets were classified as having low, medium, or high HTV at the time of candidate drug selection. Comparing this HTV classification with later clinical development efficacy data showed that HTV classification was highly predictive of future clinical efficacy success. This new strategy for HTV assessment provides a novel approach to early prediction of clinical efficacy and a better understanding of portfolio risk.

Introduction

Recent decades have witnessed a period of significant decline in delivery of new medicines to patients (Booth and Zimmel, 2004; Kola and Landis, 2004; Pammolli et al., 2011; Khanna, 2012; Hay et al., 2014). The issue is industry wide and seems to affect all therapeutic areas (Booth and Zimmel, 2004; Paul et al., 2010; Hutchinson and Kirk, 2011; DiMasi, 2014). A comprehensive, longitudinal review of AstraZeneca's candidate drug (CD) projects during the 2005–2010 period found safety to be the predominant reason for project closures in preclinical and first time in human testing (Cook et al., 2014). However, a recent analysis of mid- and late-phase (phase II and III) clinical attrition across the industry shows that the majority of failures between phase II and submission were due to lack of drug efficacy (Arrowsmith and Miller, 2013; Hay et al., 2014). Specific analysis of AstraZeneca projects further confirmed that efficacy failures accounted for the majority of project closures in late-phase development (Cook et al., 2014; Morgan et al., 2018). Since the majority of drug discovery and development costs are from phase II to launch (Paul et al., 2010), strategies to avoid late-phase failures are urgently needed. Also, from an ethical standpoint, it would be

important to avoid treatment of patients with experimental new medicines having no clear clinical benefit. Since the majority of late-phase failures are caused by lack of efficacy, the focus of this work was to develop a strategy to early on predict which novel drug targets will in the future have the highest likelihood of delivering clinical efficacy, and thus of being worth investing in.

Materials and Methods

Development of Human Target Validation Classification System. A team of eight senior cardiovascular and gastrointestinal (CVGI) translational scientists, bioscientists, and clinicians at AstraZeneca with extensive experience in drug discovery and clinical development participated in two workshops to define the elements that contribute to building human target validation (HTV) knowledge. The list of HTV elements was triaged into a list of definitions, which were then consolidated into a 10-point classification system that was ranked from lowest to highest in terms of perceived level of confidence generated (Table 1).

Evaluation of the Ability of the HTV Classification System to Predict Future Clinical Efficacy. A list of all CDs, which during the last 50 years progressed through clinical development into studies of sufficient length to enable efficacy readouts in CVGI at AstraZeneca Research and Development (R&D), was generated. To determine the level of HTV at the time of CD selection for legacy projects, a subgroup of four of the workshop scientists identified former project/clinical leads for each project. To make an objective assessment of each project they were asked to outline the HTV

The primary laboratory used in this study was at AstraZeneca (Gothenburg, Sweden).

This work was funded by AstraZeneca.
<https://doi.org/10.1124/jpet.118.250894>.

ABBREVIATIONS: CD, candidate drug; CVGI, cardiovascular and gastrointestinal; HTV, human target validation; R&D, research and development.

TABLE 1
Human target validation classification

HTV	Definition
High HTV	
Level 1	Drug is on the market
1.1	Mortality/morbidity endpoint met in target population
1.2	Surrogate endpoint (approvable) met in target population
1.3	Confidence-generating endpoints met in target population
Level 2	Clinical tool is hitting the target
2.1	Surrogate endpoint (approvable) met in target population
2.2	Confidence-generating endpoints met in target population
Level 3	Drug is hitting a target in the pathway of interest
3.1	Mortality/morbidity endpoint met in target population
3.2	Surrogate endpoint (approvable) met in target population
3.3	Confidence-generating endpoints met in target population
Level 4	Drug/clinical tool is hitting the target and/or pathway outside the target population
4.1	Surrogate endpoint (approvable) met
4.2	Confidence-generating endpoints met
Medium HTV	
Level 5	Phenotypes/genotypes in the diseased state have been identified
5.1	Extreme (for example, naturally occurring mutants)
5.2	Heterogeneous (for example, based on population databases, patient cohorts)
Level 6	Clinical challenge model
6.1	Nondrug; stress test of pathway (for example, fat meal, lipopolysaccharide [LPS] challenge)
Level 7	Target affected in human ex vivo tissue manipulation
7.1	For example, platelet ADP stimulation, insulin secretion from human islets
Low HTV	
Level 8	Target/target pathway is altered in healthy vs. diseased population
8.1	For example, positron emission tomography tracer data supportive, bioinformatics, in silico mapping
Level 9	Target of interest is present in correct tissue/cell type
9.1	For example, bioinformatics, experimental evidence
Level 10	No relevant human data

evidence available at CD selection and to provide the requisite evidence that was available to provide support. One of the subgroup members then assessed the HTV level and confirmed their assessment with the respective project representative. Each assessment was then peer-reviewed by the other subgroup members to avoid any bias, and if judged necessary further review with project representatives was undertaken.

HTV status at the time of CD selection was then established (low, medium, or high) and compared with the outcome of the efficacy readouts in the clinical studies. This meant that projects tested in humans with closures (safety, pharmacokinetics, etc.) before studies of sufficient length allowing clinical efficacy readouts were conducted were not included in the analysis. Not to bias the high HTV status group, second-generation candidates on a specific target for the same indication were excluded from the analysis. For many historical drug projects, it was the norm to generate backup/second-generation drug candidates to: 1) address any arising safety/pharmacokinetic issues in the front runner and 2) to identify compounds with potentially differentiated efficacy profiles. These second-generation candidates often entered clinical development once clinical efficacy data had become available on the front runner, thereby increasing the HTV. To include these would have caused bias in the number of projects with high HTV and the decision was made that they should be omitted.

Application of HTV Classification to an Existing Drug Portfolio. The HTV classification was applied to the existing cardiovascular and metabolic disease discovery portfolio during the years 2012–2014. Project leaders and translational scientists evaluated the platform of evidence with respect to different HTV levels to define the current status for the current portfolio. Subsequently, each project generated plans on how to progress HTV understanding. Each project team got the task to develop short-term (6 months), medium-term (12 months), and long-term (18+ months) plans with the aim to increase the HTV status at the time of CD selection.

Statistical Analysis. To determine if the likelihood of delivering a successful efficacy readout in the high HTV category was significant, we applied the Fisher's exact test comparing this group to the pooled results for medium and low HTV (due to the limited number of projects in these categories).

Results

HTV Classification System

Herein, we describe the 10 levels of evidence for HTV, progressively building confidence and understanding of the target in humans (Table 1).

Low HTV (Levels 8–10)

Level 10. At the lowest degree of target confidence, there are no human data available supporting the relevance of the target from a human perspective.

Level 9. The next level of confidence generation would be to demonstrate that the target is expressed in the appropriate tissue or cell type in humans. Such evidence may be generated by experimental (e.g., quantitative polymerase chain reaction, in situ hybridization, or protein quantification) or informatics investigations [e.g., databases such as The Human Protein Atlas (<http://www.proteinatlas.org>)]; however, expression of the target in the relevant tissue or cell type does not necessarily predict involvement of the target in the pathophysiological process of the disease. For example, there have been several companies with clinical development programs for melanin-concentrating hormone receptor 1 antagonists as an antiobesity treatment. Although melanin-concentrating hormone receptors are expressed in the human hypothalamus, which is part of the brain critical to energy balance regulation (Tan et al., 2002), and a large preclinical platform of evidence exists supporting melanin-concentrating hormone receptor 1 antagonists as an antiobesity treatment in rodents (Kowalski and McBriar, 2004), no convincing human data on weight loss have been generated to date (MacNeil, 2013).

Level 8. The next level of evidence we defined was demonstration that the target or target pathway is altered when comparing between healthy and diseased states. Methods such as positron emission tomography ligand imaging, metabolomics, transcriptomics, proteomics, and biomarker assessments are all valuable in describing such differences (Phelps, 2000; Frank and Hargreaves, 2003; Kaddurah-Daouk et al., 2008). For example, change in the melanocortin system in obesity has been implied by assessing the concentration of the endogenous MC4R antagonist, agouti-related protein, in the circulation in obese and nonobese men (Katsuki et al., 2001). The MC4R is widely expressed in the central nervous system, including sites that are well known to contribute to the coordinated control of body weight homeostasis. Activation of the

MC4R by the endogenous ligand, α -melanocyte-stimulating hormone, reduces food intake and body weight, whereas binding of the endogenous antagonist agouti-related protein increases food intake and body weight (Adan et al., 2006). In the study by Katsuki et al. (2001), circulating levels of agouti-related protein were increased in obese men and correlated with various parameters of obesity, which shows an altered regulation of the melanocortin system in the diseased state (obesity). However, since level 8 only requires changes between healthy and disease states, nothing can be said about the cause(s) and the consequence(s) for the disease.

Collectively, levels 8, 9, and 10 provide limited evidence for the target in disease; therefore, these levels constitute a low HTV (levels 8–10).

Medium HTV (Levels 5–7)

Level 7. To achieve level 7, it is necessary to demonstrate target involvement in the physiological process relevant to the disease. This can be achieved by demonstrating that the target is involved in mediating effects resulting from ex vivo manipulation of human tissue by use of disease relevant assays. Examples involve the use of human material such as platelets, pancreatic β cells, or tissue biopsies in conjunction with ex vivo drug administration. An example could be assessment of insulin secretion from human pancreatic β cells following ex vivo drug administration. Another example is the change in human platelet aggregation in ex vivo assays that was shown in response to P2Y₁₂ antagonists, which provided a rapid way of generating human data and confidence in the target (van Giezen and Humphries, 2005).

Level 6. Further confidence in a target can be generated by testing whether the target pathway is altered in vivo during a clinical challenge. While such an approach does not involve an interventional drug, it can use, for example, a dietary or immune provocation to examine target relevance in humans. For example, a target involved in energy balance regulation in humans is expected to demonstrate changed levels in fed and fasting states. This can be exemplified with the changes seen in leptin concentrations after feeding and fasting, which support a role for leptin in energy balance regulation in humans (Boden et al., 1996; Kolaczynski et al., 1996).

Level 5. Noninterventional assessments of target validity may be determined by the prevalence of phenotypes and/or genotypes in the human diseased state (Sioud, 2007). The relevance of the MC4R pathway in body weight regulation was, for example, implicated in a large study showing that 5.8% of subjects with severe early onset obesity had pathogenic mutations in the MC4R (Farooqi et al., 2003). Furthermore, mutations in the SGLT2 in humans result in increased urinary glucose excretion, supporting the rationale for the development of SGLT2 inhibitors in diabetes treatment (Enigk et al., 2011). The examination of genetic and population databases is thus an important strategy to generate evidence, although most projects do not have genetic data available (Plenge et al., 2013; Cook et al., 2014).

Collectively, these three levels provide greater confidence in the target than levels 8–10, but lack the interventional data to ultimately show that an interaction with the target would really generate the desired effect. Therefore, these levels are described as medium HTV (levels 5–7).

High HTV (Levels 1–4)

Level 4. The highest degrees of HTV are achieved from interventional studies using drug molecules. An interventional study at the most basic level will be able to demonstrate that a drug or a clinical tool is impacting either the target or the target pathway outside of the relevant target population, assuming appropriate biomarkers linked to the mechanism of action are in place. For example, in an obesity indication it could mean assessment of body weight (surrogate endpoint) or food intake (confidence-generating endpoint) in an overweight or even normal weight population using a drug that is launched for another indication but hitting the right target or a clinical tool. For example, assessment of body weight is usually captured in clinical studies and may already be available in data collected in studies for another indication.

Level 3. A higher degree of confidence is generated if a drug is influencing a target in the pathway of interest (but not the actual target) within the relevant patient population, with an effect on appropriate endpoints. In addition to level 4, confidence in efficacy in the appropriate target population is also achieved.

Level 2. If a clinical tool is directly affecting the target itself in the target population, confidence in the target is even higher. Such investigations could be conducted, for example, using a drug under development for another indication, allowing its use for testing in a new indication.

Level 1. Ultimately, the highest level of validation that can be achieved is when a drug is on the market hitting the same target and demonstrating effect in the target population. An example of a level 1 HTV is the development of the reversibly binding P2Y₁₂ platelet inhibitor ticagrelor. Here, the aim was to develop a drug that generated greater efficacy and had a differentiated safety profile when compared with clopidogrel. Clopidogrel binds to the P2Y₁₂ receptor in an irreversible manner in contrast to ticagrelor, which has different consequences for offset of action, e.g., if a patient suffers trauma with the risk of bleeding (Teng, 2015). However, it should be noted that from a drug development perspective, while the validation at this level is high, developing new drugs on the same target might not be commercially attractive unless also well differentiated.

The 10 different levels of evidence allow ranking into low, medium, and high HTV as described previously and as presented in Table 1.

Historical Data Validate the HTV Strategy

Testing of the HTV classification system was performed by a retrospective analysis of all CDs selected during the previous 50-year period that entered into human studies, across nine indications within the CVGI therapy area (Fig. 1). This list of more than 50 CDs was then reviewed for the available evidence pertaining to HTV at the time of CD selection (decision point when an investment to take a novel drug candidate into human is made) using the HTV classification described in Table 1. Only CDs with studies of sufficient duration to allow evidence of clinical efficacy readouts were included in the evaluation. In a number of cases there were several generations of CDs selected within a project that produced clinical efficacy data. Not to bias the analysis, HTV assessment was only counted for the first CD generating

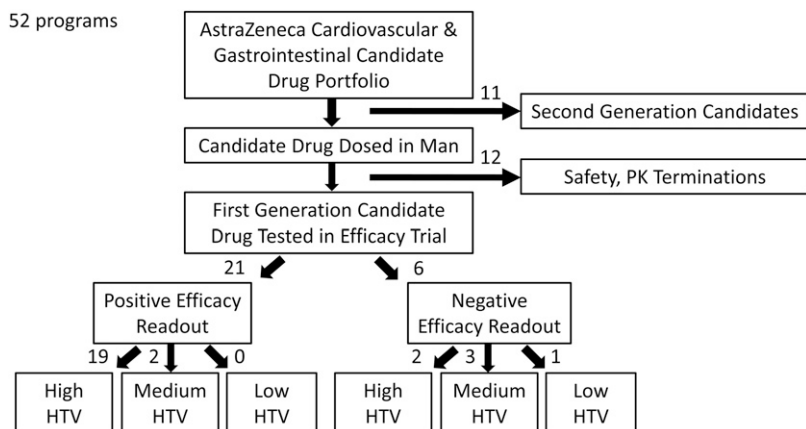


Fig. 1. Schematic process demonstrating project filtering. CDs selected during 50 years of R&D at the AstraZeneca CVGI therapy area were filtered for second-generation molecules and nonefficacy-based attrition, and whether the efficacy readout was generated prior to the retrospective evaluation of the 10-point HTV classification with regard to its ability to predict future clinical efficacy of a CD.

clinical efficacy, avoiding duplicate counting in the high HTV group (Fig. 1).

This analysis showed that if the HTV was high (that is, interventional evidence using a drug or tool existed, levels 1–4), then the probability of achieving clinical efficacy was high, reaching 90% in our analysis (Fig. 2A). The success rate for high HTV compared with medium and low HTV was significant ($P < 0.05$; Fisher's exact test), highlighting the value of selecting CDs with this level of validation. Given the criteria for high HTV, it is conceivable that best in class programs (when a drug molecule targeting the mechanism is already on the market or in development) disproportionately favor success in this category; however, when we subdivided this group into best ($n = 11$) and first ($n = 10$) in class we found that the success rate was approximately 90% for both categories. If HTV was only classified as medium (levels 5–7), the analysis showed a significant decline to 40% likelihood of demonstrating clinical efficacy; by definition all programs in this group were first in class. If HTV was classified as low (levels 8–10), then achieving clinical efficacy in humans was determined to be zero. This final observation comes with the caveat that we found only three examples of CDs that had been taken into clinical development with low HTV, likely reflecting the value attached to generating relevant HTV evidence. Only one of these programs progressed to efficacy testing.

Application of HTV Classification to an Existing Drug Portfolio

The data from our 50-year evaluation of projects clearly demonstrated that HTV status at the time of CD nomination is linked to the probability of clinical efficacy success. Thus, it follows that increasing the level of HTV before CD selection will improve the likelihood of clinical efficacy success and thereby decrease the risk of development failure. Therefore, HTV assessment was applied to the cardiovascular and metabolic disease discovery portfolio and this information was then used to devise strategies to increase the level of HTV for all projects.

In 2012, each project team was encouraged to define and subsequently set up activities over short-term (6 months), medium-term (12 months), and longer-term (18+ months) perspectives that would enhance HTV status based on the 10-level HTV classification (Table 1). Accordingly, each project underwent review at 6-month time points over an 18-month (or longer) period to assess progress against the HTV status assessment. Project teams were encouraged to ensure that HTV activities were truly decision making, either to enhance confidence in the target, or to provide a clear rationale for discontinuation.

At the initial HTV status assessment in 2012, 50% of projects were found to have low HTV status (levels 8–10). By implementing the HTV strategies over the subsequent time

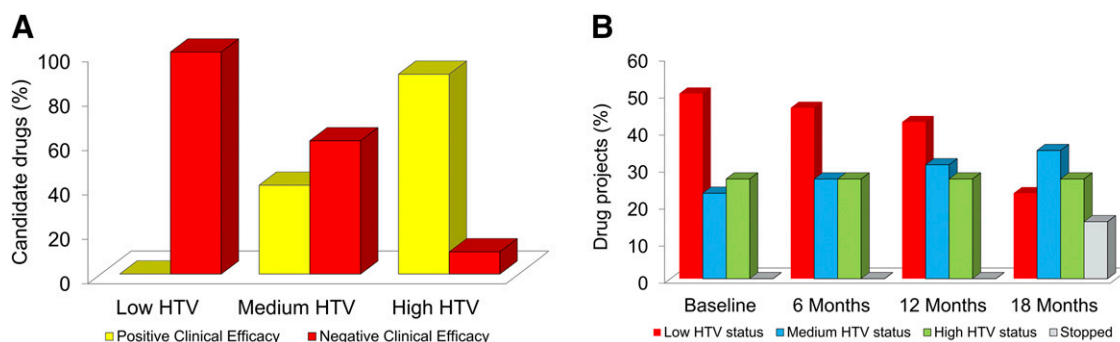


Fig. 2. HTV assessment: relationship of HTV status to clinical efficacy and portfolio impact of applying HTV strategies. (A) Historical HTV status at the time of CD selection. A retrospective evaluation of HTV for all CDs that entered into human studies of sufficient length to generate efficacy readouts in the past 50 years within AstraZeneca's CVGI therapy area. Data show HTV status at the time of CD selection and relationship to clinical efficacy. The number of projects in each group can be found in Fig. 1. (B) Portfolio impact of developing and applying HTV strategies. Application of HTV strategies to projects resulted in a decrease in the percentage of projects with low HTV status, an increase in the percentage of projects with medium HTV status, and no change in the overall proportion of projects with high HTV status over a period of 18 months and beyond.

period, the number of projects with low HTV status decreased by around half, to 23%. At the same time, the percentage of projects with medium HTV status (levels 5–7) increased from 23% to 35%, while 15% of projects were stopped due to a lack of confidence. The proportion of projects with high HTV status (levels 1–4) remained unchanged at 27% throughout the 18-month time frame (Fig. 2B). The projects making up this group changed as a result of terminations for competitive positioning. However, successful progression of other projects was made from medium to high HTV status, refilling this portfolio segment.

A clear benefit to projects of the HTV assessment comes from the facilitation of discussions on how to build confidence in the target. This is of particular relevance for projects in which it proved to be difficult to develop simple strategies to enhance the HTV status. Such projects may require a more thorough scrutiny than would have been the case without continuous HTV assessment.

An example of an HTV strategy is shown for the GPR103 antagonist project that only had a preclinical rationale for obesity and weight management. At the time of project initiation, the target was found to have a level 9 HTV (classified as low HTV status). This level was based on mRNA expression data demonstrating that GPR103 is expressed in human brain, including the hypothalamus, which is the region of the brain controlling food intake (Neveu et al., 2014). Additionally, pyroglutamylated RFamide peptides (the endogenous ligands of GPR103) expression and binding was found in the human hypothalamic neurons (Ramanjaneya et al., 2013). A strategy to enhance the level of HTV through a set of sequential plans was defined and is described in Table 2. A short-term key HTV activity would be to establish evidence that the GPR103 pathway and/or its ligands are dysregulated in obesity, which would be undertaken by *in silico* analysis of internal and external databases. This activity would increase the scoring to level 8, albeit still low HTV status. In the medium term, a clinical challenge study could be conducted in which samples are collected from the same individuals

following long-term fasting and in a fed state. If GPR103 was to play a role in energy balance regulation the determination of altered ligand levels (pyroglutamylated RFamide peptides 26 and 43) in humans in fasting and fed states would be supportive of a role for this mechanism in energy balance regulation, providing a level 6 (medium HTV) validation. Longer-term planning would include an early exploratory human study with a tool compound using clinically relevant endpoints. Such a study would require front loading from different functions including synthesis, formulation, and safety, as outlined by Kummar et al. (2008). The outcome of such a strategy would have maximum impact on the project (especially if started at the earliest opportunity that a tool compound was available) with the possibility to stop many years in advance if no efficacy was proven. Strategic decisions resulted in this project not progressing.

A second example of an HTV strategy is that of TB4, which had been implicated in tissue repair and regeneration. TB4 was considered as a target for intervention in myocardial infarction and acute coronary syndrome patients. At the time of project initiation there was no human evidence linking the target to the specific disease biology, only published data generated in preclinical models, providing a level 10 HTV. Initial activities were to use *in silico* bioinformatic analysis to map the downstream pathways of TB4 using a selection of tools and human data sets lifting the HTV to level 9. Once an understanding of the TB4-associated pathways was established they would be interrogated in bio-samples collected from acute myocardial infarction and acute coronary syndrome patients, and by comparing the data generated to that from healthy samples demonstrate differences in TB4-associated pathways in the diseased state, generating a level 8 HTV. It would then be possible to try and establish a TB4 pathway biomarker profile that would be descriptive of a phenotype-predicting outcome. This level 5 HTV understanding would not only establish increased confidence in TB4 for these indications but would also offer the potential to develop a personalized healthcare strategy once an investigational

TABLE 2

Example of HTV strategy (for GPR103 antagonist in obesity and weight management)

Timescale	Project Start	6 mo	12 mo	18+ mo
HTV status	Low (level 9) Target present in correct tissue	Low (level 8) Target/pathway is altered in healthy vs. diseased population	Medium (level 6) Pathway-related effect in clinical challenge model	High (level 2) Clinical tool is hitting the target
Evidence	mRNA expression data demonstrated GPR103 expression in human brain, including the hypothalamus, and QRFP (endogenous ligand) expression and binding in human hypothalamic neurons			
Key activities		Is there evidence that the GPR103 pathway and/or its ligands are dysregulated in obese patients (in <i>in silico</i> analysis of internal and external databases)?	Is there evidence that the GPR103 receptor ligands QRFP26 and QRFP43 are present in either the circulation or cerebrospinal fluid at different levels in humans between fasting and fed states?	Does an exploratory clinical tool GPR103 antagonist reduce food intake in humans?

QRFP, pyroglutamylated RFamide peptide.

drug was developed. Before this work was completed the availability of specific TB4 tool molecules demonstrated that the published preclinical data were not reproducible and the program was discontinued.

Discussion

The past decades have witnessed decreasing pharmaceutical R&D productivity. The current cost of bringing a new drug to market is estimated to be as high as \$1.8 billion and a vast majority of investigational products that enter clinical trials fail (Paul et al., 2010). Especially in late-phase clinical trials, attrition rates have increased (Pammolli et al., 2011) and most of the late-phase failures are due to lack of efficacy (Arrowsmith and Miller, 2013; Cook et al., 2014; Hay et al., 2014). In the Critical Path Report (“Challenge and Opportunity on the Critical Path to New Medical Technologies”) from March 2004, the US Food and Drug Administration concluded that a new product development toolkit was urgently needed to reduce the time and resources expended on candidate products that are unlikely to succeed. More specifically, the Food and Drug Administration requested collective action to predict the safety, effectiveness, and manufacturability of medical products. In this work, we have focused on development of a strategy that already at the time of CD selection predicts the likelihood of a drug target to demonstrate future clinical efficacy. The HTV classification system developed was validated within the CVGI therapy area across nine indications over a period of 50 years. The cardiovascular therapy area often requires long studies to prove clinical efficacy, and together with oncology is the disease area with the lowest rate of Phase III success (Hay et al., 2014). The CVGI therapy area is thus a good example of where a tool predictive of clinical efficacy could be of great value in prioritizing between projects.

The focus on validating targets from a human perspective has received increased attention over the last years, most likely driven by the decrease in R&D productivity and the late-phase efficacy failures. In fact, selection of the most safe and potent molecule will still fail if the wrong target is selected (Cook et al., 2014). In the literature, different aspects of human target validation have been discussed, including use of human tissue for analysis of target expression and function (Coleman and Clark, 2003), genetics (Plenge et al., 2013; Nelson et al., 2015), small clinical trials for early test of proof of concept in a “fast-fail” model (Paul et al., 2010; Karlsson, 2015a,b; Owens et al., 2015), and the need for predictive and robust biomarkers supporting a “fast-fail” model (Cook et al., 2014). Also, integrated combinations of the aforementioned aspects have been proposed to improve R&D productivity (Plenge, 2016). In this study, we aimed to take science one step further by introducing a grading of human evidence supporting a target that could be used to predict the likelihood of later clinical efficacy success. To our knowledge, this is the first time such a strategy has been developed that takes all aspects of human target validation into account in a structured way for future clinical efficacy prediction.

It is clear from our analysis across 50 years that the higher the HTV status at the time of CD selection (levels 1–4), the higher the chance of clinical efficacy success; vice versa, the lower the HTV status at the time of CD selection, the lower the chance of clinical efficacy success. It is already possible to proactively build HTV into project planning from project initiation to support early and pragmatic decision

making. The aim of challenging the HTV status is to build confidence (or not) of future clinical efficacy early in a project through results from human experiments. If HTV status can be improved there is an increased chance of future clinical efficacy. On the other hand, if HTV status cannot be further improved from a low status despite efforts, it may indicate that the drug target would not be relevant from a human perspective.

HTV classification is preferably conducted in the cross-functional project team at project initiation. The project team members should be aware of all available external and internal data allowing evaluation of the current level of HTV. If a target ends up with a confidence level of high, no further HTV activities would be needed. However, if a target ends up with a confidence level of low or medium, plans should be put in place to design experiments that would have the highest impact on the HTV status, taking into account the available toolbox and project timelines. A prerequisite for building HTV understanding is availability of the toolbox needed. Key components of the toolbox are relevant biomarkers, human tissue, clinical and bioinformatics databases from healthy and diseased populations, as well as tool compounds suitable for use in humans. An initial lack of interventional data on a target (and thus no high HTV status) should not be seen as a stop for generating interventional data. Instead, projects could from the start plan to take a tool compound with proven efficacy in animal models into small decision making human studies to get a stop/go decision. Such an approach is associated with upfront costs and planning, but has the potential to save money in the longer term while avoiding treatment of patients with experimental new medicines with limited clinical benefit. The ultimate HTV would be human testing already done before CD selection, with the use of a tool compound and biomarkers predictive of efficacy. Such data will enable a rapid stop/go decision based on assessment of an early efficacy signal. It is important that the right biomarkers are selected to avoid a false positive result, and therefore progression of a CD that is destined to fail. One example in which this happened was the development of the novel GABA_B receptor agonist lesogaberan for improvement of reflux symptoms in patients with persistent gastroesophageal reflux disease symptoms despite receiving proton pump inhibitor therapy. Data on transient lower esophageal sphincter relaxations and gastroesophageal reflux inhibition from pre-clinical models translated well to clinical studies in healthy volunteers and patients with gastroesophageal reflux disease. However, in later phase studies these biomarkers did not translate into clinical meaningful effects on symptom relief, and thus continued development occurred based on a partially inappropriate selection of surrogate endpoints where the relationship to symptoms was not yet established in this heterogeneous patient population (Boeckxstaens et al., 2011). Determining what degree of validity of the biomarker is required to balance the risk of incorrect decisions versus the investment involved in biomarker validation is critical, and will vary by project (Peck, 2007).

Validation of the 10-point HTV classification system to predict future clinical efficacy was based on 50 years of legacy R&D data in the CVGI therapy area. The validation part was thus run as a retrospective analysis as described in *Materials and Methods*. A prospective analysis would ideally be a more robust way for validation, but by taking the time frame into

account a prospective validation is not feasible due to the number of years required to generate the data.

This work was conducted in the framework of the CVGI therapy area, which often requires long and sizable studies to prove clinical efficacy. Together with oncology, cardiovascular is an area with the lowest rate of Phase III success (Hay et al., 2014), and a tool to derisk late-phase efficacy failures is therefore warranted. There is no reason to believe that the utility should be different for other therapy areas, although this still needs to be proven.

Highly innovative targets come with a limitation that HTV is low and the toolbox is limited. In our opinion, such projects should be regarded as a high risk in the portfolio, but acceptable given the potential they offer, as long as there is good risk balance in the portfolio. It is not an absolute certainty that low HTV equates to a failure to deliver clinical efficacy for a target; however, such targets merit additional investment to enhance their validity where possible.

To conclude, the HTV assessment strategy as presented here has been shown to be effective in predicting clinical efficacy. It has been validated using 50 years of AstraZeneca data in the CVGI therapy area and goes far beyond genetic validation, which is only available for a fraction of all drug targets. We have shown that HTV strategies can be easily incorporated into project planning, and HTV assessment front-loads hypothesis testing and reduces the risk of late-phase development failures for lack of efficacy. Using guidance from this easy-to-use HTV classification system would speak for delivery of an increase in new medicines to patients.

Acknowledgments

We thank Jesper Havsol at AstraZeneca for conducting statistical analysis and Sue Harris and Diane Kwiatkoski at IQVIA, (Durham, NC) for their editorial assistance with the manuscript, which was funded by AstraZeneca. We are indebted to all colleagues who helped formulate the HTV classification system, evaluated project HTV status, and provided critical review of the manuscript.

Authorship Contributions

Participated in research design: Karlsson, Greasley, Wähländer.
Conducted experiments: Karlsson, Greasley, Gustafsson.
Contributed new reagents or analytic tools: Karlsson, Greasley, Wähländer.
Performed data analysis: Karlsson, Greasley.
Wrote or contributed to the writing of the manuscript: Karlsson, Greasley, Gustafsson, Wähländer.

References

Adan RA, Tiesjema B, Hillebrand JJ, la Fleur SE, Kas MJ, and de Krom M (2006) The MC4 receptor and control of appetite. *Br J Pharmacol* **149**:815–827.

Arrowsmith J and Miller P (2013) Trial watch: phase II and phase III attrition rates 2011–2012. *Nat Rev Drug Discov* **12**:569.

Boden G, Chen X, Mozzoli M, and Ryan I (1996) Effect of fasting on serum leptin in normal human subjects. *J Clin Endocrinol Metab* **81**:3419–3423.

Boeckxstaens GE, Denison H, Jensen JM, Lehmann A, and Ruth M (2011) Translational gastrointestinal pharmacology in the 21st century: 'the lesogaberan story'. *Curr Opin Pharmacol* **11**:630–633.

Booth B and Zimmel R (2004) Prospects for productivity. *Nat Rev Drug Discov* **3**:451–456.

Coleman RA and Clark KL (2003) Target validation using human tissue: from gene expression to function. *Targets* **2**:58–64.

Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, and Pangalos MN (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov* **13**:419–431.

DiMasi JA (2014) Pharmaceutical R&D performance by firm size: approval success rates and economic returns. *Am J Ther* **21**:26–34.

Enigk U, Breitfeld J, Schleinitz D, Dietrich K, Halbritter J, Fischer-Rosinsky A, Enigk B, Müller I, Spranger J, Pfeiffer A, et al. (2011) Role of genetic variation in

the human sodium-glucose cotransporter 2 gene (*SGLT2*) in glucose homeostasis. *Pharmacogenomics* **12**:1119–1126.

Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, and O'Rahilly S (2003) Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**:1085–1095.

Frank R and Hargreaves R (2003) Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov* **2**:566–580.

Hay M, Thomas DW, Craighead JL, Economides C, and Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nat Biotechnol* **32**:40–51.

Hutchinson L and Kirk R (2011) High drug attrition rates—where are we going wrong? *Nat Rev Clin Oncol* **8**:189–190.

Kaddurah-Daouk R, Kristal BS, and Weinshilboum RM (2008) Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol* **48**:653–683.

Karlsson C (2015a) Exploratory clinical studies ("phase 0 trials"), in *Principles of Translational Science in Medicine: From Bench to Bedside* (Wehling M ed), 2nd ed, pp 229–233, Elsevier Amsterdam, Netherlands.

Karlsson C (2015b) The pharmaceutical R&D productivity crisis: can exploratory clinical studies be of any help? in *Principles of Translational Science in Medicine: From Bench to Bedside* (Wehling M ed), 2nd ed, pp 227–228, Elsevier Amsterdam, Netherlands.

Katsuki A, Sumida Y, Gabazza EC, Murashima S, Tanaka T, Furuta M, Araki-Sasaki R, Hori Y, Nakatani K, Yano Y, et al. (2001) Plasma levels of agouti-related protein are increased in obese men. *J Clin Endocrinol Metab* **86**:1921–1924.

Khanna I (2012) Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug Discov Today* **17**:1088–1102.

Kola I and Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* **3**:711–715.

Kolaczynski JW, Ohannesian JP, Considine RV, Marco CC, and Caro JF (1996) Response of leptin to short-term and prolonged overfeeding in humans. *J Clin Endocrinol Metab* **81**:4162–4165.

Kowalski TJ and McBriar MD (2004) Therapeutic potential of melanin-concentrating hormone-1 receptor antagonists for the treatment of obesity. *Expert Opin Investig Drugs* **13**:1113–1122.

Kummar S, Rubinstein L, Kinders R, Parchment RE, Gutierrez ME, Murgo AJ, Ji J, Mroczkowski B, Pickeral OK, Simpson M, et al. (2008) Phase 0 clinical trials: conceptions and misconceptions. *Cancer J* **14**:133–137.

MacNeil DJ (2013) The role of melanin-concentrating hormone and its receptors in energy homeostasis. *Front Endocrinol (Lausanne)* **4**:49.

Morgan P, Brown DG, Lennard S, Anderton MJ, Barrett JC, Eriksson U, Fidock M, Hamrén B, Johnson A, March RE, et al. (2018) Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nat Rev Drug Discov* **17**:167–181.

Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, et al. (2015) The support of human genetic evidence for approved drug indications. *Nat Genet* **47**:856–860.

Neveu C, Dulin F, Lefranc B, Galas L, Calbrix C, Bureau R, Rault S, Chuquet J, Boutin JA, Guilhaudis L, et al. (2014) Molecular basis of agonist docking in a human GPR103 homology model by site-directed mutagenesis and structure-activity relationship studies. *Br J Pharmacol* **171**:4425–4439.

Owens PK, Raddad E, Miller JW, Stille JR, Olovich KG, Smith NV, Jones RS, and Scherer JC (2015) A decade of innovation in pharmaceutical R&D: the Chorus model. *Nat Rev Drug Discov* **14**:17–28.

Pammolli F, Magazzini L, and Riccaboni M (2011) The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov* **10**:428–438.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, and Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9**:203–214.

Peck RW (2007) Driving earlier clinical attrition: if you want to find the needle, burn down the haystack. Considerations for biomarker development. *Drug Discov Today* **12**:289–294.

Phelps ME (2000) Positron emission tomography provides molecular imaging of biological processes. *Proc Natl Acad Sci USA* **97**:9226–9233.

Plenge RM (2016) Disciplined approach to drug discovery and early development. *Sci Transl Med* **8**:349ps15.

Plenge RM, Scolnick EM, and Altshuler D (2013) Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**:581–594.

Ramanjaneya M, Karteris E, Chen J, Rucinski M, Ziolkowska A, Ahmed N, Kagerer S, Jöhren O, Lehnert H, Malendowicz LK, et al. (2013) QRFP induces aldosterone production via PKC and T-type calcium channel-mediated pathways in human adrenocortical cells: evidence for a novel role of GPR103. *Am J Physiol Endocrinol Metab* **305**:E1049–E1058.

Sioud M (2007) Main approaches to target discovery and validation. *Methods Mol Biol* **360**:1–12.

Tan CP, Sano H, Iwaasa H, Pan J, Sailer AW, Hreniuk DL, Feighner SD, Palyha OC, Pong SS, Figueroa DJ, et al. (2002) Melanin-concentrating hormone receptor subtypes 1 and 2: species-specific gene expression. *Genomics* **79**:785–792.

Teng R (2015) Ticagrelor: pharmacokinetic, pharmacodynamic and pharmacogenetic profile: an update. *Clin Pharmacokinet* **54**:1125–1138.

van Giezen JJ and Humphries RG (2005) Preclinical and clinical studies with selective reversible direct P2Y₁₂ antagonists. *Semin Thromb Hemost* **31**:195–204.

Address correspondence to: Dr. Cecilia Karlsson, Cardiovascular, Renal and Metabolism Translational Medicine Unit, Early Clinical Development, IMED Biotech Unit, AstraZeneca Gothenburg, Pepparedsleden 1, SE-431 83 Mölndal, Sweden. E-mail: cecilia.karlsson@astrazeneca.com