Promise of new translational safety biomarkers in drug development and some challenges to regulatory qualification

One promise of new translational safety biomarkers (TSBs) is their ability to demonstrate that toxicities in animal studies are monitorable at an early stage, such that human relevance of potential adverse effects of drugs can be safely and definitively evaluated in clinical trials. Another is that they would provide earlier, more definitive and deeper insight to patient prognosis compared with conventional biomarkers. Recent experience with regulatory authorities indicates that resource demands for new TSB qualifications under the current framework are daunting and the rate of their expansion will be slow, particularly in light of mounting financial pressures on the pharmaceutical industry. Sponsors face a dilemma over engaging in safety biomarker qualification consortia. While it is clear new TSBs could be considered catalysts to drug development and that patient health, business and scientific benefits, described here using examples, should outweigh qualification costs, concerns exist that early ambiguities in biomarker interpretations at the introduction of such new TSBs might hinder drug development.

KEYWORDS: cardiac biomarker ▪ drug development ▪ gastrointestinal injury biomarker ▪ kidney biomarker ▪ liver biomarker ▪ preclinical imaging ▪ regulatory qualification ▪ translational safety biomarker ▪ vascular injury biomarker

Frank D Sistare[†1] & Joseph J DeGeorge[1]

Safety Assessment and Laboratory Animal Resources, Merck and Co., Inc., West Point, PA 19486-0004, USA
[†]Author for correspondence: frank_sistare@merck.com

Advances in technology and basic research have yielded numerous potential translational safety biomarkers (TSBs) with the promise to enhance pharmaceutical development efficiency and success while ensuring patient safety. Once these additional safety biomarkers are sufficiently qualified and demonstrated to perform translationally across the common animal test species (rat, mouse, dog and nonhuman primate) and to humans, the hope is that they would enhance the efficiency and utility of both the animal toxicology studies that are used to assess the safe conduct of clinical drug trials, as well as the safety of clinical trials themselves. Their value would lie in demonstrating that the toxicities seen in animal toxicology studies are monitorable while any ill effects are still reversible and significant toxicity is preventable. In the clinic, their value would lie in being able to assess whether the drug toxicities seen in animal studies are relevant to humans at the doses and exposures needed for therapeutic benefit. Compared with conventional endpoints, they would provide signals earlier, possibly at lower doses associated with toxicities, and may provide mechanistic insight to toxic effects enabling deeper insight to patient prognosis.

It is one matter to sufficiently define the performance characteristics of new TSBs with the rigor and confidence for applications very early in preclinical stages of drug development in order to enable internal corporate decisions involving compound selection. A much greater fit-for-purpose qualification effort is needed to develop a widespread consensus position where reliance on new safety biomarkers in regulated phases of drug development to ensure the safe conduct of a clinical trial is embraced by both drug sponsors and regulatory authorities worldwide [1]. This is referred to here as regulatory qualification, and requires far more resources and a wider commitment of stakeholders than for biomarkers that are reserved for internal company decision making. Such resources, commitments and consensus for regulatory qualification of new biomarkers, may best be addressed by broad collaborations including regulatory authorities.

To achieve success, the benefits from expanding the toolbox of such regulatory qualified TSBs must be convincingly articulated, so as to garner and maintain commitments to the resources needed to execute the long-term research for biomarker qualification strategies. Furthermore, the real and perceived risks associated with introducing new TSBs with limitations that initially are likely to be incompletely understood, must be acknowledged and carefully considered in early biomarker adoption strategies. This overview and perspective summarizes general scientific and business benefits as well as the anticipated risks associated with

future medicine part of fsg

efforts to expand the TSB toolbox. Hypothetical but realistic drug development scenarios are used for illustration, and provide perspective for capturing opportunities to deploy TSBs at different stages of drug development. These are built upon the progressive framework of evidentiary performance standards, or qualification, that a biomarker has attained [2].

Scope of TSBs

The topic of safety biomarkers in drug development is very broad. The scope of this article is limited to accessible biomarkers of drug response, which may signal significant concern for safety following dosing of animals and humans. Biomarkers of drug exposure are not in scope for this article (e.g., measurements of DNA or protein adducts, measurements of drug and drug metabolites), nor are biomarkers of pre-existent drug susceptibility (e.g., somatic DNA sequence polymorphisms in drug-metabolizing enzymes or leukocyte antigens). The range of analytes that can serve as TSBs of drug response is still broad, and the scope of this article is further limited to those analytes that are easily accessible using noninvasive approaches, and are translational across animal test species and into the clinic. Such a definition includes multiple modalities, some of which apply in certain circumstances but not others. For example, biomarkers of gene expression in a liver or kidney sample, while clearly biomarkers of drug response, are presently not easily accessed and not in scope, while gene expression changes in leukocytes, which are easily accessed from blood collections would be. Noninvasive imaging of organs and tissues using nuclear magnetic resonance, x-rays and ultrasound would be in scope. Analytes in body fluids such as urine, blood (serum, plasma, and circulating cells and cell fragments), tears or saliva, and even cerebral spinal fluid (depending on the views within a country) are relatively accessible and within scope. Analytes in these accessible fluids that have served as fertile substrates for biomarker discovery have included proteins, circulating DNA, circulating cell population distribution measurements, easily accessed epithelial cells (e.g., from skin, buccal mucosa and urine), measures of the contents of cell fragments (e.g., exosomes and micronuclei), endogenous metabolic intermediates of biochemical pathways (e.g., blood glucose, serum creatinine, cholesterol and bilirubin), and, relatively recently, the discovery of the appearance of miRNA in body fluids following drug induced tissue injuries [3,4].

Arguments for qualifying additional TSBs

General considerations

Drug toxicities seen either in animal toxicology studies or in clinical trials contribute significantly to the attrition of drugs from the development pipeline. Analyses of trends across the pharmaceutical industry have indicated that toxicity findings seen in animal studies alone contribute to decisions to terminate development of approximately to 30% of all compounds that have been selected to enter development [5,6]. The inferences from such surveys are that either: compounds need to be more fully characterized at earlier stages to detect and predict later appearing toxicities in both animals and humans, and filter out those destined to fail before entry into development; or compounds that are in fact safe to humans are being terminated because the nonclinical species are not predictive of humans and we lack the tools to safely and convincingly demonstrate this. From an analysis of animal toxicology studies and clinical study findings applied to 150 compounds in drug development, it has been demonstrated that, overall, when toxicities are seen in human clinical trials, a single animal species can be expected to have predicted that human toxicity approximately 40–60% of the time [7]. When a second species is added then 71% of toxicities seen in human trials were found to have been present in at least one of the animal test species. Furthermore, when a human toxicity was correctly predicted by at least one of the animal studies, for half of the compounds the toxicity was seen in only one of the test species. Importantly, but not assessed in this paper, is the observation that each animal toxicology study usually identifies several target organs of toxicity. Therefore, in a limited manner these data address the sensitivity of the animal studies for predicting human toxicities on a set of 150 development compounds that advanced into clinical testing. However, the true positive and false positive prediction rates of human findings for all compounds tested in animals that were intended for the clinic, will never be known with certainty. The conclusion to draw from these data is that animal toxicology studies provide guidance regarding what safety concerns could be anticipated in human clinical trials, but they are imperfect and likely excessive predictors of human response. When the toxicities observed in such animal toxicology studies are not easily monitored and reversible, then the negative impact of animal toxicity findings of suspect relevance for humans becomes one of the

most pressing drivers for the expansion of new TSBs. This driver applies to deployment both prior to and after compound selection for drug development. TSBs are needed to assure patient safety while assessing, in clinical trials, the relevance of animal toxicology study findings that are currently non-monitorable and suspected of irrelevancy to human.

Another pressing driver for expanding the TSB toolbox is the current inability to differentiate and distinguish among the different manifestations of target organ drug-induced toxicities with conventional safety biomarkers at both the animal testing level, as well as the individual patient level in the course of a clinical trial. Is the mode of action responsible for a safety concern decipherable using a panel of accessible biomarkers? Can the subanatomical origin and location of the injury be defined to help provide perspective on risk? Can the histopathologic processes, such as degeneration, apoptosis, regeneration, fibrosis, or acquired or innate immune cell infiltrations activated during injury be differentiated by accessible biomarker measurements? Can such mode-of-action or tissue response biomarkers that are linked to molecular pathways or cellular and tissue pathophysiologic responses, inform patient safety earlier and define legitimate drug treatment-related changes with greater clarity? Additional accessible TSB capable of providing such information would help to sub-categorize target organ injuries, rank concerns, and differentiate levels of risk better among otherwise currently indistinguishable changes seen using only conventional biomarkers. For ALT, for example, lack of sensitivity for detecting liver damage is less of an issue than its inability to diagnose the severity of damage or prognosis for clinical liver failure [8].

Business case for additional qualified TSBs

From a published estimate as of the year 2000, the mean Phase I clinical trial development costs for a drug were reported to be US$31 million, the mean Phase II clinical trial costs an additional $42 million, and, for a Phase 3 candidate, an additional $129 million [9]. The mean final clinical success rate to achieve a favorable marketing decision, beginning from those compounds starting Phase I trials was 21.5%, and the mean time from the start of clinical testing to marketing approval for a representative drug was over 7.5 years. When adjusted for the estimates of overall clinical approval success rates, the mean out-of-pocket total clinical development phase costs per approved drug were estimated to reach $467 million. Furthermore, the estimate of combined efforts from discovery and preclinical testing needed to enable the initiation of Phase I trials constituted an additional 42% of a successful investigational drug's mean total development costs of $802 million. A cost estimate completed only 5 years later indicates that mean total development costs for small molecules have spiraled between 2000 and 2005 from $802 million to over $1.3 billion [10]. And even more recently [11] the estimated clinical approval rate for small molecule drugs is reported to drop from 21.5 to 13%, so in 2011 current costs are likely even higher. Using the estimate of 42% from 2000 applied to the $1.3 billion mean total development cost from 2005, this translates to an estimated mean investment of $500 million made in earlier discovery and preclinical research before a viable pharmaceutical candidate could even enter clinical testing phases. With such an investment in resources, it would seem foolish not to begin an investigation of expected clinical benefits once requisite animal GLP toxicology studies are complete. However, when a non-monitorable animal toxicity unfolds at low exposure margins in the course of Phase I trials enabling animal toxicology studies, even in one species, the door to the clinic may be closed to that candidate in the interest of preserving patient safety and another chemical would need to be sought if the motivation remains to proceed with development of an agent to interact with that drug target. With research costs so high, development failure rates so significant and development timelines so protracted, it can be readily appreciated that intelligent applications of new well-understood TSBs might therefore allow safe clinical testing and enable continued development of drug candidates that are safe to humans despite the concerning but conflicting animal toxicology study findings, reduce ambiguities to identify truly nonviable drugs from the drug development pipeline before advancing inappropriately in clinical development where costs will escalate, and accelerate the pace and quality of business decision-making to shorten development timelines for a given drug candidate, and, thereby, decrease opportunity cost and lengthen market exclusivity lifetime. These situations could each enhance clinical development success rates or resolve safety concerns earlier and more definitively to have a significant financial impact on reducing overall development costs or timelines. If a new qualified TSB with improved performance over conventional endpoints could

eliminate the need for just one Phase II study by providing a convincing safety signal at relevant exposure margins from Phase I study samples, for example, where conventional biomarkers fail, the savings could exceed $42 million. If the appearance of a toxicity that is non-monitorable with conventional biomarkers and seen for the first time in a longer-duration animal study is demonstrated to be monitorable using a new TSB, then its application to samples from a Phase II study could be used to allay concerns about human relevance, enable development, and eliminate the need for an additional $31 million in Phase I study costs for a back-up molecule. Such successful efforts might also avert months or even years of delay in marketing approval for the lead molecule that could quickly add up to hundreds of millions of dollars. While this exercise is not intended to provide a rigorous economic evaluation of the value of TSBs in the course of medical practice, the cost reviews are meant to provide some general perspective on the business benefits that could be realized for expanding TSB qualification initiatives, so that the resource expenses discussed later for a full qualification of such tools can be compared.

Patient health & scientific case for additional qualified TSBs

The following examples, which are summarized in Table 1, are intended to provide perspective on how improved TSBs can provide enhanced clarity in a practical and tangible manner to help resolve ambiguities seen in the course of drug development, and help to assure patient safety while providing scientific justification to support drug development decisions. Case examples are provided from pre-candidate selection phases, postcandidate selection regulated phases of preclinical drug development and regulated phases of clinical drug development.

Earlier identification of targets & compounds with potential safety liabilities & enabling associated de-risking strategies prior to candidate selection

Target validation, lead identification, lead optimization and compound selection are stages in the development of a pharmaceutical that may be considered for this discussion to be exclusive company internal decision-making stages, and not subject to regulatory oversight. As described above, it is costly for sponsors not to carefully consider potential safety liabilities in these pre-candidate selection stages and to instead have later stage failure. Given this cost, it is critical to have well-qualified test systems providing meaningful data, and the ability to evaluate these safety data against a solid historical experience. Therefore, the confidence and strength of evidence, or level of qualification, even from such early applications must be high. False-positive safety signals could derail the progression of crucial medicines, while false negative safety signals could delay exposure of the truth and lead to very costly failures at later stanges. Getting the balance right at early stages is challenging. Some strategies may set the bar low for demonstration of safety at the cost of killing safe compounds (repeated 'low cost to failure' cycles), others may set the bar high for accepting a clear safety signal that would halt development, and may be willing to tolerate a potential false negative, but then quickly focus attention onto subsequent definitive de-risking studies to allay lingering concerns from these early potential safety signals ('accepting down stream kills'). These data are not being generated to support the safe conduct of a clinical trial and are therefore outside the scope of regulatory inspection and discovery [12]. Rather, they are informing on investment risk applied to a particular discovery strategy. Therefore, there need be no debate with regulatory authorities as to whether such signals may be false or true positives, and where thresholds should be set for such internal decision-making test systems and safety biomarker measurements. In this case, the important debates are held within companies. Definition and traditional assessment of safety will follow if development is continued. It can be considered that such early safety screening and de-risking strategies at or prior to compound selection are outside the domain of regulatory authorities, or even when applied during development for late-occurring risks (e.g., explorations using emerging biomarkers with the potential for predicting tumorigenicity well in advance of the conduct of carcinogenicity studies). It can also be considered that safety screening and de-risking strategies in these early stages are areas of internal company investment and represent potential competitive commercial advantages, and, therefore, may not be appropriate for broad qualification through open consortia. These are discussions and decisions that senior pharmaceutical leaders and managers must face. For any TSBs that may be used as tools for internal decisions, there need only be sufficient confidence and performance characteristics for a company to accept their utility for investment

Table 1. Summary table of hypothetical but reasonable examples of drug development scenarios that support the patient health, scientific and business case for qualifying new translational safety biomarkers.

Phase of Development	Example	Summary description	Estimated benefit from deploying new safety biomarker
Pre-candidate selection phase applications	Novel diuretic target safety concern	New translational kidney safety biomarkers add to weight of evidence that the pattern of kidney toxicity is same as for previously marketed and safe diuretics, and, therefore, due to excessive pharmacology of very sensitive dog test species.	Low risk of additional kidney toxicity from novel diuretic mechanism – target effectively de-risked to support investment.
	Lead ID to minimize risk of cardiac hypertrophy	New translational cardiac hypertrophy plasma safety biomarkers and imaging applications optimize study design to enhance chance of success for reaching a definitive answer.	One of three drug intervention strategies prioritized to guide lead ID strategy and minimize resource spend.
	Vaccine formulation lead optimization	Gene expression changes in peripheral blood leukocytes are used to characterize and subsequently minimize systemic and local vaccine reactogenicity potential.	Potential best-in-class selection strategy to minimize clinical adverse effects.
	Renal injury de-risked at compound selection stage	Nonhuman primate exhibits renal toxicity with lead that is thought to be human relevant. Best of three candidates selected for development based on minimal study design using renal biomarker longitudinal measurements.	Safest of three candidates selected for development to minimize drug development delay.
Postcompound selection phase applications: preclinical GLP animal toxicology studies and/or clinical trials	Rat-only kidney pathology first seen in chronic study	New translational kidney biomarkers demonstrate monitorability of kidney toxicity seen only in a chronic rat study. Shorter rat studies and chronic monkey studies are negative. Clinical studies demonstrate no changes in kidney biomarkers.	Ambiguities about human safety concerns are eliminated. US$31 million in clinical development is preserved. Delays in development are avoided.
	Rat-only GI histopathology first seen in chronic study	Gastrin elevations are demonstrated to be useful for monitoring early onset of GI lesions seen in chronic rat studies. Clinical studies proceed with the inclusion of gastrin measurement collections.	Ambiguities about human safety concerns are eliminaed. $31 million in clinical development preserved. Delays in development avoided.
	Skeletal muscle histopathology in two species but at widely divergent safety margins	Serum skeletal troponin I is used to demonstrate monitorability of muscle injury seen in animal toxicology studies with a lead compound and is deployed in clinical studies. Clinical investigations demonstrate sufficient target engagement at safe exposures but no evidence of disease mitigation.	Human safety concerns minimized to enable clinical investigations and evaluate target validity. Business decision reached sooner to abandon the program and resources shifted sooner.
	Mechanism-based tissue mineralization seen in rats	Serum calcium and phosphorus balance is established as a reasonable mode of action biomarker in rat studies and translated to clinical investigations.	Ambiguities about human safety concerns are minimized. Development timeline for a drug to treat a life-threatening disease mainatained.
	Prostate atrophy seen only in dogs	Noninvasive MRI is demonstrated in a dog study to safely monitor the onset of prostate atrophy and subsequent reversibility after dosing is halted. Imaging is included in clinical trials and no effect on prostate is seen.	Ambiguities about human safety concerns are eliminated. $31 million in clinical development preserved. Delays in development avoided.
	Hemorrhagic cardiomyopathy seen only in a second rat study	Explanations for very different outcomes in two similar rat studies are reached, but serum cTnI measured in patient samples from already completed clinical trials provide assurances of patient safety.	Patient safety is confirmed and continuation of clinical investigations is enabled.

cTnI: GI: DEFINE; GLP: DEFINE; ID: DEFINE.

risk. The experimental rigor expectations, high cost associations, and lengthy debates around thresholds, design and number of prospective studies considered adequate are held internally, as the broad acceptance across industry and regulatory agencies associated with regulatory qualification are not necessary before strategic implementation by a company.

The earliest assessments of toxicity potential in the pre-candidate selection space tend to be dominated by *in vitro* model systems owing to the scarcity of adequate test substance [13,14] and well-advertised initiatives have been launched to broadly assess, and potentially expand the value of numerous toxicity screening endpoints for predicting later safety liabilities in environmental chemicals [15] and pharmaceuticals [16,101] with the contribution of test compounds by participating sponsors and disclosure of certain internal test data [17]. However, testing in genetically diverse animal models in later phases of the pre-candidate selection stage allows companies their first opportunity to relate and integrate chemistry, dose, exposure, metabolism, molecular biology, pharmacology and toxicology. These tend to be mouse and rat models, because of the scarcity of compound availability at early stages, and it is here that new and conventional accessible TSBs can first be coupled with the analyses of terminal tissue specimens using conventional histopathological and newer genomic and metabolomic approaches. A battery of multiplexed accessible TSBs for the nonrodent (generally, dog or nonhuman primate) to identify likely safety-related development hurdles can also be deployed in initial tolerability tests prior to candidate selection in a very limited number of animals used primarily to minimize compound need and yet provide some initial insight to exposure and safety relationships.

Therefore, such accessible TSBs are needed, even in this early development space, to translate across animal test species and could be applied by researchers seeking to evaluate targets and select lead molecules based on performance in animal efficacy models that might include mice, rats, dogs, monkeys, or other species, potentially informing on therapeutic margins. Enabling drug-discovery researchers to view potential toxicity early is prudent to inform decisions prior to investing in the scale-up of activities supporting regulatory filings. From a strategic perspective, such early pre-GLP animal studies also offer opportunities to discover, evaluate and calibrate the performance of promising new TSBs wherein sponsors gain valuable experience in compound-selection animal-tolerability experiments, especially given that several related structures are likely to be tested and contrasted.

Hypothetical example 1: the use of translational kidney safety biomarkers to de-risk toxicity concerns associated with target validation

A discovery team is seeking to develop a diuretic with a mechanism of action directed against a novel kidney target. Prior publications indicate that, amongst the standard drug development test species, the dog is the most sensitive to adverse effects associated with excessive pharmacology of diuretics, leading to histopathologic changes in the kidney that are mechanism based, resulting from sustained excessive pharmacologic activation, and are not considered human relevant [18]. Furthermore, diuretics are known to be associated with prerenal azotemia [19] but the intra-renal effects of such a prolonged pharmacologically mediated azotemic state are not well established. Tools to assess whether any kidney toxicities of the new compound class are indeed related to an extension of sustained excessive pharmacology, whether they would exactly resemble that of marketed compounds with a safe clinical track record of decades of use, and whether they would, therefore, be human irrelevant would provide confidence to the company regarding the safety of the target. Once a lead chemical is identified with good pharmacologic activity, the sponsor collects urine from dogs, included in tests of activity of the new diuretic against a marketed diuretic, and also measures a panel of urine kidney injury biomarkers from the urine. They find, using just three instrumented animals, that the dog is exquisitely sensitive to the pharmacology, and, furthermore, the time and dose–response relationship between certain critical pharmacologic activity measurements, BUN and serum creatinine changes, and the pattern and sequence of which new urine kidney biomarkers change and which do not change, is identical between the new test agent and the marketed agent. For example, urinary glomerular biomarkers such as total protein, cystatin C or albumin [20] may be seen to not change; proximal tubular biomarkers such as kim-1 and microalbumin [21,22] may be seen to not change; and only more distal nephron or pan-nephron biomarkers, such as GSTpi, osteopontin and clusterin may be seen to change [23]. However, these changes may only be seen after rises in BUN and serum creatinine, indicating a strong pre-renal effect that is correlated to the extent and duration of excessive diuresis, the intended pharmacology. The weight of evidence gathered at this stage prior to lead optimization, candidate selection, scale-up of compound synthesis and the conduct of extensive terminal animal toxicology studies provides the sponsor confidence that the kidney toxicity that will likely be observed in association with the novel diuretic mechanism will be irrelevant to humans and be a surmountable toxicity hurdle.

Hypothetical example 2: the use of translational cardiac hypertrophy safety biomarkers to confirm lead identification while minimizing resource spend

A sponsor is seeking to modulate the activity of a targeted pathway expected to mitigate a serious disease and wishes to prioritize its options for developing a full receptor antagonist, an allosteric receptor modulator or an inhibitor of local concentrations of the endogenous agonist. Genetic evidence in the literature hints that cardiac hypertrophy may be associated with modulation of activity in the targeted pathway after chronic dosing. A lead is identified among each of the three approaches that can be used to demonstrate sufficient evidence of efficacy in a rat model to support advancement of further investment into lead optimization. The sponsor deploys several approaches to help prioritize which of the three approaches is least likely associated with development of cardiac hypertrophy – longitudinal monitoring of rat equivalent plasma biomarkers to B-type natriuretic peptide/NTpro-B-type natriuretic peptide of drug-induced cardiac hypertrophy [24], weekly monitoring using noninvasive imaging [25], and terminal sacrifice with conventional measurements of heart weight and microscopic assessments of critical cardiac dimensions. The imaging and accessible biomarker approaches allow the sponsor to time the study duration to a point where separation can be clearly seen among the three mechanisms without repeated interim sacrifices or an arbitrarily long study duration, and a definitive answer is gained using heart weights and microscopic histomorphometric measurements.

Hypothetical example 3: the use of translational peripheral blood gene expression safety biomarkers to improve vaccine lead optimization

An example has been presented by investigators at Merck who have screened over a dozen marketed vaccines with a spectrum of low, medium and high incidences in nonhuman primates, and severities of local and systemic reactogenicity related adverse events was recorded in the clinic [26]. They apply a systems biology bioinformatic based approach to correctly categorize the vaccines based on peripheral blood gene expression profiles seen in the monkeys after dosing. The approach is used to optimize development and selection of future vaccine/adjuvant formulations in primates to maximize product efficacy while minimizing patient reactogenicity.

Hypothetical example 4: the use of translational kidney safety biomarkers to accelerate compound selection while minimizing resource spend

A sponsor observes significant treatment-related renal histopathology with a test agent in a 2-week GLP nonhuman primate study at low-exposure margins over intended human targets, with no significant changes in serum creatinine or BUN at any time in the study. Urine collected at study termination confirms that certain new kidney safety biomarkers were elevated in the nonhuman primate. The GLP 2-week rat study did not present with renal histopathology despite high exposure multiples. The sponsor decides, after considering numerous other factors, to abandon further development of this test agent. To minimize resources, a small-animal and compound-sparing, single dose level, 1-month exploratory compound selection study in the nonhuman primate is designed to provide desired exposure margins with three promising backups, with baseline and continual longitudinal assessment of urine biomarkers. Newly qualified urinary kidney biomarkers [2,27] are monitored and begin to change at day 7 with one of the compounds, at day 14 for a second compound, and of continual dosing two study arms are terminated after just 2 weeks and 3 weeks, respectively, and the kidney pathology is confirmed histologically. The sponsor elects to extend the study with the third back-up compound for 3 months but after 6 weeks with regular urine sampling, and no evidence of urine biomarker changes, efforts are initiated to scale-up synthesis and launch further development of this compound. Longitudinal urine biomarker monitoring is continued and, after 3 months, no evidence of renal injury is seen histologically despite good exposure. This compound's lack of renal pathology, appropriate selection decision and trajectory into development are confirmed.

Postcompound selection-regulated stages of GLP animal toxicology testing

In the postcandidate selection and highly regulated development space, the utility of those biomarkers with potential clinical translational utility may first begin to be explored within companies' animal toxicology studies. Here, codified regulations in the USA mandate that industry sponsors provide adequate information about the pharmacological and toxicological properties of a drug, as described in 21 CFR 312 "A sponsor who intends to conduct a clinical investigation subject to this part shall submit an

Investigational New Drug Application (IND) including … adequate information about pharmacological and toxicological studies of the drug involving laboratory animals or *in vitro*, on the basis of which the sponsor has concluded that it is reasonably safe to conduct the proposed clinical investigations. As drug development proceeds, the sponsor is required to submit informational amendments, as appropriate, with additional information pertinent to safety … for each toxicology study that is intended primarily to support the safety of the proposed clinical investigation, a full tabulation of data suitable for detailed review" [102]. Such toxicology studies, used to support the safe conduct of clinical trials are performed under GLP guidelines. To help ensure that clinical trials with test compounds can be conducted "reasonably safely", regulatory guidance further specifies [103] that tests are generally conducted in two species, a rodent and nonrodent with the durations of studies appropriate for supporting the clinical trials of at least equivalent duration [103]. Toxicology studies of a minimum duration of 2 weeks are generally needed while clinical trials longer than 6 months are supported by 6-month rodent and 9-month nonrodent studies.

There is an expectation from regulatory authorities that animal toxicology studies will be conducted at sufficiently high doses and exposures to identify and elicit target organ toxicities in each species. Such GLP toxicology studies are rarely expected to demonstrate that a compound has no toxicities in any species at maximal doses. Given that any subsequent clinical trial is expected to be conducted in a reasonably safe manner, and that animal toxicities are expected, sponsors and regulatory authorities must exercise prudent judgment on decisions regarding the conduct of clinical research studies based on relative exposures needed for efficacy versus toxicity, the serious and reversible nature of the toxicity, differences that may be known to account for species sensitivities, and whether or not the onset of the toxicity is monitorable at a stage when toxicity is readily reversible.

Given that regulations specify that sponsors submit "adequate information … pertinent to safety", a point for debate that arises is whether results from measurements in the course of a GLP toxicology study of a promising TSB that has not yet been qualified for regulatory purposes would need to be submitted to an IND. Are these measurements of an emerging TSB in the course of a GLP animal toxicology study considered "adequate information … pertinent to safety" or are they opportunities to learn whether or not these measurements may be informative and, thus, eventually "pertinent to safety"? The effort to sufficiently assess all criteria to qualify and understand the limitations of new safety biomarkers is significant. Therefore, challenges in setting appropriate medically significant thresholds at such an early stage of biomarker applications are daunting. Regulatory authorities seeking to minimize any safety risks are expected to view ambiguities in the interpretations of the significance of any perceptible changes in TSBs conservatively, despite the reality that TSBs that are not adequately qualified are thus of limited practical reliability for safety. Therefore, logical tendency, therefore, is for sponsors to measure only mature, well accepted, highly qualified, hence, 'traditional' safety biomarkers in such GLP studies. The following examples are presented in the context of new TSBs that are sufficiently qualified and established to be considered appropriate for the specific applications to GLP animal toxicology studies and human clinical trials.

New regulatory qualified TSBs, that are easily accessible at interim timepoints during the conduct of GLP studies from 2 weeks to 9 months duration, are useful for satisfactorily demonstrating that the onset of reversible toxicities is monitorable. Such data would support conclusions that a clinical trial can be conducted using the same biomarker strategy in the clinic to ensure and document patient safety. There may be situations where a sponsor may learn that a lead compound for a new untested target is not commercially viable (e.g., caused by deficiencies in half life, metabolic interactions or cost of goods, or nonclinical safety concerns), but the availability of an appropriate backup compound is significantly delayed. Bringing the lead compound into an early short-duration clinical trial, which could be conducted safely only by using a new safety biomarker, could enable the sponsor to make an earlier go/no-go decision for the entire program. An experimental medicine analysis of the relationship between target engagement and potential for disease mitigation, as well as a validation of whether the toxicity observed in test species translates to humans, could be all that is needed to confirm the target. Such timely information could inform drug development and portfolio strategy sooner than waiting for a 'clean' back-up. There are situations where large safety margins exist based on targeted human exposures and exposures where histopathology is noted in animal studies, but

current traditional biomarkers for assessing human toxicity are insensitive. While a clinical trial of short duration may be deemed safe, it could be beneficial to conduct analyses with more sensitive newly qualified TSBs from that short-duration study to inform the likelihood of safety signals being raised by the less sensitive conventional biomarkers in longer-duration clinical studies at the same doses. Such capability could better assure subject safety and prevent the waste of resources and development time on a lead compound that could be better spent on a more appropriate back-up or on a different target altogether. However, the specificity of those new qualified TSBs must be known with high certainty for such a pivotal decision to proceed into longer clinical studies to be made.

Hypothetical example 5: the use of new translational kidney safety biomarkers to investigate concerns of kidney injury

A therapeutic agent under development is demonstrating great promise for a serious unmet medical need in patients in Phase 2 clinical studies after one month of treatment. The compound has been tested in rats and nonhuman primates for one month with no significant toxicologic effects observed at high safety margins. However, in the chronic 6-month rat studies, kidney histopathology is seen with a low exposure safety margin, and no effects of the test agent on BUN and serum creatinine were seen. In the nonhuman primates, after 9 months of dosing with good exposures attained, there is no evidence of any treatment-related histopathologic findings in the kidney. There is already evidence of potential patient benefits, the sponsor has already invested in GLP chronic animal toxicology testing, Phase 1 and 2 clinical testing and, in addition, the nonhuman primate is demonstrating no significant toxicity. The sponsor first measures newly qualified kidney safety biomarkers in urine samples collected longitudinally from a repeat 6-month chronic study in rats and demonstrates that certain biomarkers first appear in the urine after 9 weeks of dosing, that they can be used to monitor the onset of histological change, and that the lesion is reversible within 6 weeks of when dosing is halted and the biomarker changes were first noted. The sponsor then measures concentrations of the responsive set among the new kidney safety biomarkers in the samples of urine collected from the 1 month clinical studies and demonstrates that there are no treatment-related changes. In additions, the sponsor proposes to deploy the new kidney TSBs to monitor renal safety in patients beyond 1 month of continual dosing in subsequent, longer-duration clinical trials. The selection of biomarkers is agreed upon with institutional review boards and the regional regulatory authorities, and the trial is conducted with minimal delay to development time.

Hypothetical example 6: the use of TSBS to investigate concerns of gastric injury

A compound is tested in rats and monkeys, and gastric mucosal histopathology is seen microscopically in the 6-month study only in rats after demonstrating no signs of injury in a 1-month study. Serum gastrin changes are known to occur following both physiologic and pathologic changes to the stomach [28] and have been monitored previously to provide assurances of patient safety concerns for drug injury to the gastrointestinal tract [29]. The sponsor demonstrates a significant treatment-related rise in serum gastrin levels from samples retained from interim bleeds during the chronic rat study. The sponsor decides to proceed into longer Phase 2 studies incorporating longitudinal measurements of gastrin to maintain investment in the lead compound, which is demonstrating promise of target engagement and of efficacy in 1-month clinical studies. No elevation in serum gastrin is noted and the weight-of-evidence conclusion from the subsequent clinical studies is that humans are not responding in the same manner as rats to the test agent. Significant delays in the development timeline are avoided.

Hypothetical example 7: the use of TSBs to investigate concerns of skeletal muscle injury

A compound is tested in rats and dogs, and in rats skeletal muscle pathology is seen microscopically in the 1-month study with a low exposure margin, while in dogs the exposure margin to any histopathology in skeletal muscle is greater than 50-fold over targeted human exposures. There are no clear treatment-related effects on serum creatine phosphokinase levels at doses where histopathology is initially noted. The sponsor demonstrates a significant treatment-related rise in serum skeletal troponin I using a recently developed assay [30] on the same samples retained from interim and final necropsy bleeds during the 1-month rat study and at the high dose of the dog study. Since serum skeletal troponin responses have been reported in clinical instances of muscle injury previously [31,32], the sponsor proceeds into Phase I studies,

incorporating longitudinal measurements of skeletal troponin I along with the conventional creatine phosphokinase measurements to maintain investment in the lead compound and to understand the promise of pharmacologic engagement of this novel target. No elevation in serum skeletal troponin I is noted and the weight-of-evidence conclusion from the subsequent clinical studies is that humans are not responding in the same manner or as sensitively as rats to the test agent. The sponsor learns from continued clinical trials with the test agent that engagement of the novel intended target can be documented. However, no favorable signal of therapeutic benefit results using efficacy biomarkers, so the program is abandoned and chemistry and biology resources are moved to more promising targets.

Hypothetical example 8: the use of TSBs to investigate concerns of tissue mineralization

Investigators at Pfizer have published an account of a development on oncologic test candidate that induced a widespread tissue mineralization after continuous dosing for longer than 2 weeks in rats but not in dogs [33]. The effect appeared to be mechanism-based and the challenge was to establish a safe monitoring strategy to allow introduction of the test agent into clinical trials to begin clinical investigations of the drug intended to treat a life-threatening disease. Could beneficial antitumor effects, using an optimized dosing regimen, be achieved at doses and exposures that did not result in progression of tissue mineralization? A biomarker of tissue mineralization was needed. The authors demonstrated that serum free calcium and phosphorus dysregulation was linked to tissue mineralization and the product of serum calcium multiplied by phosphorus levels served as a sufficiently early mechanism-based toxicity biomarker to allow clinical investigations of a novel promising compound for a life-threatening disease. The sponsor proceeded to clinical trials, and no evidence of calcium and phosphorus dysregulation were seen at desired clinical levels.

Hypothetical example 9: the use of noninvasive imaging approaches to investigate concerns of drug-induced effects on internal organ histomorphology

A compound is tested in dogs and significant treatment related decrements in prostate weights are seen after 4 weeks of dosing, attributed to atrophy. No effects on the prostate were seen in the rat. A MRI strategy is deployed in a dog study to demonstrate that the effect is monitorable after 2 weeks with a match between histology and image-based sensitivity, and the effect is fully reversible within 4 weeks after dosing is halted. The compound is brought to the clinic with baseline and longitudinal MRI assessment after 2 and 4 weeks, respectively, of dosing, and there is no effect on prostate size.

Hypothetical example 10: the use of TSBs of drug-induced cardiac injury to allay concerns of patient harm during a prior completed clinical trial

In the course of compound development, the sponsor sees no evidence of cardiac toxicities in animal toxicology studies conducted following 1 month of dosing with a test compound in either rodents or monkeys. However, in a 6-month rat study, animals are surprisingly presenting with morbidity in a time and dose-dependent manner after just 3 weeks at doses shown in the previous 1-month study to be well-tolerated. Histopathologic examination from early sacrifice animals demonstrated a hemorrhagic cardiomyopathy, and cardiac troponin (cTn)I and T levels are found to be elevated from samples that had been collected at 2 weeks. Following an extensive investigation into the reasons for the discrepancy in findings between the two rat toxicology studies, a plausible and human-irrelevant rationale is identified, serum samples that had been collected from patients after 1 and 4 weeks on study are secured and analyzed for cTnI levels. The weight of evidence, bolstered by the negative findings in serum cTnI levels in the clinical study, support resumption of the clinical investigation with the test compound.

Arguments against investing in the expansion of qualified safety biomarkers

■ At early stages of safety biomarker deployment, limited collective experience & threshold ambiguities are expected

In initial applications of new safety biomarkers, it will be inevitable that there be a feeling of 'raising the safety hurdle' by seeing changes in TSBs at doses and durations where no changes in traditional safety indicators have been observed, and where the real medical risk may not be clear. The outcome, and certainly the concerns, could be that development of a compound may be delayed or even derailed by deploying new TSBs. Following introduction of a new biomarker,

debates are inevitable over what are specific, clinically meaningful thresholds or cutoff levels and where the concern for safety is real, as well as the perception of the harm that would result if the biomarker were to fail. This 'tolerability of risk' theory, describing a critical point for the psychological acceptance of a new biomarker by all stakeholders, has been described recently [34] and has very real implications for setting proper thresholds that balance sensitivity and specificity appropriately. Setting a very low threshold of change would increase the sensitivity of detection but would lead to an increased false-positive rate for the biomarker. An approach that strikes the proper balance in the course of both the qualification process and in the initial applications of newly qualified biomarkers is needed to minimize potential for harm to individual patients while improving access of medicines to patient populations. At the initial application stage of new qualified TSBs, it will be important to seek mutual agreement between industry and regulatory authorities on:

- When it may be reasonable and appropriate to deploy a new TSB on a drug development study;
- If it is included, what magnitude of change will be considered clinically meaningful, considered a legitimate safety concern and, therefore, be actionable.

Establishing appropriate sensitivity limits for a biomarker is not a new issue. For example, debate continues as to where thresholds of clinically significant changes in ALT should be set after 50 years of experience with this clinical biomarker and these are dependent upon other concurrent measures (e.g., 3 × ALT + 1.5× Bili vs 8× ALT alone). Only recently has some level of consensus been acknowledged as evidenced by an US FDA guidance document [104]. However, that is a single regulatory authority's view in a global drug development context.

Concerns also have been raised and the debate continues around the 5–10 ms set point as the appropriate actionable threshold link between the electrocardiogram QTc interval prolongation and risk for Torsades de Pointe arrhythmia [34]. Questions have been raised, citing existing drugs with no apparent risk of Torsades de Pointe as false positives [35], about whether the threshold for this biomarker may be skewed to a conservative value to maximize sensitivity while raising drug development costs and contributing to a large number of drug terminations [34] with minimal safety benefit.

Achieving final consensus on what constitute clinically meaningful rises in other more newly qualified TSBs is certain to take time and requires diverse experience. Setting the threshold of change high would risk patient safety, while conservatively setting the threshold of change low will mean the inevitable appearance of false positives and putting compounds on 'clinical hold' for poor reasons if biomarkers are deployed prematurely. In situations when the application of specific TSBs may not be specifically warranted, and without sufficient data regarding threshold setting, this can be particularly damaging to development, especially in the context of observations in diseased patients as compared with healthy animals. In diseased patients, responses in TSBs may be further impacted by underlying disease pathology, and such experience may be difficult to attain before a pressing need arises. The setting of a threshold or cutoff for a result to be considered significant in the proper context is complex and will take a while to get right. This may be one reason for slow progress in embracing new TSB development.

Higher costs associated with new TSBs & second-tier deployment

We now have a 50-cent test for ALT and for serum creatinine, and they have served us well. The prospect of routinely moving to $200 for each new test among a panel of new kidney or liver TSBs is unattractive and would further add to escalating costs of drug development. Certainly, when there is added cause for concern triggered by findings in animal toxicology studies or other experience with the compounds directed against the same target or of a similar structural class, then additional efforts deploying new TSBs could be justifiable. This would represent a two-tier application strategy. Conventional and lower-cost biomarkers could be deployed for safety screening as a routine first-tier test. Only when there is cause for concern as discussed above, or an opportunity to gain valuable knowledge through further research with new TSBs deployed would the second-tier TSB be evoked.

Resource & time demands associated with high hurdles of regulatory qualification of new TSBs

The effort required to deliver a new set of qualified TSBs will be very resource and time intensive. To understand the sensitivity and specificity, and use characteristics of a new TSB, a broad testing strategy is needed and is best

served through precompetitive collaborations with all collaborators contributing significantly. The resources associated with such ventures that are not considered direct support of a product in the development pipeline will be high and difficult to secure in times of economic belt tightening. The time burden for the FDA, EMA, Pharmaceuticals and Medical Device Agency (PMDA) and other regulatory agencies to review is also not trivial and the PMDA and EMA convey direct charges to sponsors for such biomarker qualification submission reviews. The recent effort by the Critical Path Institute's Predictive Safety Testing Consortium (PSTC) to qualify seven new kidney safety biomarkers serves as a good example to provide some perspective on the costs, time and effort that may be needed for such a systematic safety biomarker qualification approach [2,27,36]. For that successful TSB qualification effort, data from 34 rat toxicology studies were contributed. In addition, for five of the seven biomarkers, peer-reviewed, published clinical data were available and were summarized as part of the claims advanced by PSTC to regulatory authorities. The observations were consistent with published reports of responses seen in humans with these new biomarkers to nephrotoxic agents and supported a weight-of-evidence conclusion that these five new biomarkers could also be considered qualified for clinical drug development applications under very specific contexts of use where animal toxicology studies demonstrated biomarker responses to the test agent. This 'broad-use' clinical qualification claim was not accepted by regulatory authorities despite, for example, the summary of hundreds of published references supporting the strong performance of urinary microalbumin as a clinical biomarker of renal injury and dysfunction. None of the five biomarkers was thought to have been sufficiently qualified for general use in early clinical studies (Phase I study) for detecting drug-induced acute kidney injury, and, in such cases, it was deemed that the utility of these biomarkers should be judged on a case-by-case basis [37,105]. Regulatory authorities indicated that a number of further clinical studies for extensive evaluation would be needed before widespread use of the biomarkers for detection of drug-induced kidney injury in humans. Regulatory authorities stated their expectations that the utility of the novel biomarkers should be continuously and vigorously evaluated in future clinical studies, including exploratory use of the novel biomarkers together with conventional biomarkers. As a result of this feedback, a prospective study design is now preparing for launch by the collaborative project team encompassing the Critical Path Institute's PSTC and the FNIH Biomarker Consortium to look prospectively over a 2-year period at two different nephrotoxic agents in humans with an estimated cost of $3.25 million, as compared with the 34 studies completed in rats estimated to cost less than $2 million, many of which had already been conducted previously for other purposes. This type of investment to qualify new TSBs for clinical applications is unlikely to be sustainable. As an alternative, perhaps the utility of samples from already completed clinical studies that are sufficiently well designed, and that are stored appropriately to ensure stability in freezers, should be considered for meeting regulatory qualification criteria. If thresholds, criteria for performance success and claims for utility can be defined *a priori*, based on data from published studies or limited pilot studies, and the samples are subsequently analyzed and TSB performance is confirmed, this may be a reasonable path for a future fit-for-purpose qualification. This approach is being included along with the PSTC/FNIH prospective clinical study design.

- Aside from the difficulty of diversion of resources from direct pipeline support to biomarker tool qualification, other questions that must be addressed by pharmaceutical company management are:
- Is the type of biomarker being evaluated likely to meet a significant need for their own company in the near future and, therefore, provide sufficient return on investment?
- Can corporate priorities and resource constraints afford a contribution to biomarker development that will need to be made globally available?
- Is a multiyear commitment to biomarker development sufficiently meritorious to include in the multiyear project budget being requested and thus prioritize resources away from future drug project work?
- With all of the other competing external collaboration opportunity requests, is any particular proposal sufficiently important to make the cut?
- Which kind of efforts will be joined, those that start from scratch with a prolonged time for the qualification effort to achieve goals or those where companies contribute existing

data, samples, experience up front from earlier internal efforts to accelerate progress and reduce resource requirements?

- Is the collaboration one that allows mutual benefit as competitors will contribute equivalent intellectual pre-existing knowledge and experience, or will one company drive strategy and assist competitors more than the benefit received?
- Is the scope of the effort one that involves the need for regulatory involvement, public academic collaboration- and data sharing with competitors, or is the objective one that would provide a competitive advantage and require disclosure of efforts from prior internal company investments?
- Are other consortia already engaged in exactly the same effort or are they able to share strategy, data or samples, and coordinate activities so as to be complementary and not to be redundant?
- Is the structure, management, leadership, sponsorship and program support of the consortium likely to generate success?

Clearly there is a need for leadership among industry, regulatory, professional academic organizations, not-for-profit consortia support organizations, and possibly even government or nongovernment funding organizations to align and optimize on prioritizing and defining the execution strategies for the most critical TSB qualification opportunities.

A diversity of approaches & considerations for evaluating & qualifying new TSBs

What types or categories of data provide sufficient evidence needed for facilitating qualification of a TSB? This question of evidentiary criteria for biomarker qualification has been dealt with in several recent publications including a very recent draft regulatory guidance [106,107]. Statistics-based categorical rules have been described [38,108], general performance criteria of biomarker success have been offered based on a weight-of-evidence causality association assessment of the strength of the linkage between biomarker alterations and biological outcomes [2,39], and cost–effectiveness considerations to biomarker qualification and acceptance have been raised [34].

The steps needed will be similar to get to the stage where these TSBs are well recognized, sufficiently evaluated, and fit-for-purpose qualified to be ready for: internal company decision-making involving compound selection; regulatory decision-making in animal toxicology studies, used to support the safe conduct of clinical trials; application to monitoring patient safety in a clinical drug development trial; or implementation into widespread medical practice. By necessity, the extent of research needed for those progressively more critical and broader needs will be experimentally more rigorous, require more resources and require a broader base of stakeholder involvement, and this has been termed progressive qualification [2].

It is important to recognize that there are safety biomarkers that may not be used routinely and may not be considered conventional, but have attained fit-for-purpose utility and, may therefore, be considered qualified for drug development applications. For example, in the scenarios cited in this article the application of cTnI to provide assurances that a drug candidate under development was not causing direct injury to the myocardium would be appropriate and, therefore, could be considered fit-for-purpose qualified for that particular drug development context of use even though there may be no formally approved regulatory qualification submission. Although never subjected to a formal regulatory qualification process, the cTnI and T were first reported to have utility for detecting acute myocardial infarction in 1987 [40], and are now widely viewed as effective interspecies TSBs for myocardial injury [41]. Furthermore, the cardiac troponins had been endorsed by authoritative cardiology organizations as a gold standard for redefining diagnoses of ischemic cardiac damage and acute myocardial infarction in 2000 [42]. The application of gastrin [42], serves as another example of an appropriate application of a safety biomarker to investigate the human relevance of drug-induced pathology to the stomach and may therefore be considered qualified as a biomarker for that particular drug development context, as the biomarker is known to respond similarly in humans and was shown to respond in a satisfactory treatment-related and time-dependent manner in animal toxicology studies, demonstrating its utility for monitoring safety for that compound. Gastrin monitoring has been used for example to help de-risk the entire pharmacologic class of highly effective and beneficial proton-pump inhibitors and support conclusions that findings of gastric tumors seen in rats, following prolonged stimulation by gastrin as a result of sustained proton pump inhibition, does not occur at therapeutic doses in

humans [29,43]. While serum calcium and phosphorus are routinely measured, their utility for the context of monitoring tissue mineralization had not been formally qualified. The qualification of their application to this specific context for monitoring MEK inhibitor-induced tissue mineralization was, in effect, established in the course of the drug development strategy using the investigative studies conducted in the rat. Testing of an extensive series of different agents causing tissue mineralization in animals and humans was not needed to establish the utility of this biomarker for its particular context of use. The point of these few examples is to stress that not every biomarker needs the same additional new and formal level of effort to be considered fit-for-purpose qualified, even when the contexts of use would be essentially similar – for monitoring and ensuring patient safety in a clinical trial with a specific test agent. Some biomarkers may have already been heavily evaluated with valid clinical assays established, many years of application experience and scores of supporting peer-reviewed publications. Another example is the use of noninvasive imaging, such as echocardiography, to assess progression of cardiac hypertrophy. Such imaging applications are firmly in medical practice and, at most, a demonstration project may be beneficial to establish utility for a particular novel context such as in an animal toxicology study with the specific test agent, but running ten to 20 new studies to qualify the general utility of imaging for monitoring drug-induced cardiac hypertrophy would be superfluous.

For those situations, where TSBs may be very new and yet expected to have a broad impact, and to be fairly frequently desired for second-tier deployment in drug development, the justification may be made of the need for a concerted investment in a broad-based safety biomarker qualification strategy that will necessitate a regulatory decision-making level of rigor. For such situations, while certainly drug developers and regulatory authorities are logical stakeholders, it is worthwhile to also include biomarker qualification project teams, clinical academic opinion leaders, who are positioned well to represent the viewpoints of professional societies and other organizations and networks vested in declaring on appropriate standards of medical practice with established conventional safety biomarkers. While drug development applications may be the focus of the immediate goals of new TSB qualification, it is prudent to seek input and to forge alignment early with medical practitioners. Consortia serve a crucial role, enabling frameworks for such collaboration with critical stakeholder groups. The EU has provided stimulus funding to encourage the development of consortia efforts focused on improving drug development and several are launched with goals that include qualification of TSBs [109], new safety model test systems [44] or new safety-focused technology platforms [45]. In the USA, resources directed at the qualification of new TSBs to support drug development have been pledged almost entirely by pharmaceutical companies to consortia-organized under organizations such as the Critical Path Institute Predictive Safety Testing Consortium [105], the International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) [110], and the Foundation for the National Institutes of Health [111]. For each of the efforts focused on TSB qualification where pharmaceutical collaborators are engaged, there may be obvious synergies and opportunities to engage in biomarker qualification in the course of ongoing drug development programs where the additional resources needed may be minimized and completely separate studies may not be needed. For example, collection of control serum, plasma or urine samples from normal healthy volunteers, to establish age and sex ranges of normal variations in new TSB could be included in specimen collection protocols. Collection of serum, plasma or urine samples from patients included as controls in studies of common diseases, such as diabetes, hypertension, or Alzheimer's, would help establish the effect of the underlying disease state on new biomarker baselines. The inclusion of the collections of such specimens on all study protocols, where already-marketed comparator compounds are included in study designs, could allow expanded assessments of biomarker specificity and less frequently even biomarker sensitivity when the current standard-of-care compound may be known to cause certain target organ toxicities in the course of treatment. Consortia are the logical framework for such samples and data to converge to reach the critical mass that would support qualification decisions and minimize additional resource spend. Optimizing sample storage, particularly for unknown analytes, is a critical issue to plan for here.

A need for discovering & qualifying additional TSBs

Additional drug development scenarios exist where the way forward remains obscure owing, in a large extent, to the lack of qualified TSBs.

Because TSBs do not exist for these toxicities, decision-making remains ambiguous, and drug development delays and challenges persist. A list of examples of drug-induced toxicities encountered in animal studies is provided in Box 1 where the availability of effective TSB monitoring strategies would enable drug development by providing essential assurances of patient safety and reduce decision-making ambiguities.

Conclusion & future perspective

In 5–10 years from now, sponsors and regulatory authorities will align on a consistent and coherent approach to TSB qualification that recognizes one-size-does-not-fit-all but that each TSB qualification proposal should be tailor made based on a thorough analysis of existing data strengths and real deficiencies in order to minimize qualification time and resource burdens. However, it must be done in such a manner as to afford global regulatory acceptance – no small feat for case-by-case strategies.

Measurements of new biomarkers with demonstrated storage stability of samples from existing completed GLP animal toxicology studies and completed clinical studies will be integrated with existing published data to minimize the need for additional prospective animal toxicology studies and costly new prospective clinical biomarker bridging qualification clinical studies, while achieving broad fit-for-purpose qualification decision agreements on TSBs. An appropriate balance will be reached on the benefits and risks of measuring emerging and promising (but not yet qualified) TSBs in GLP animal toxicology studies and in early clinical drug trials, where the objective of such measurements is to further understand the emerging TSB.

In the future, sponsors are expected to be collecting not only DNA samples, but also

Box 1. Examples of unmet needs in drug development for additional accessible translational safety biomarkers.

Liver injury
- Improved biomarkers are needed with:
  - Enhanced tissue specificity over ALT for liver injury;
  - More sensitive and earlier detection to inform liver dysfunction, injury response, molecular toxicology mode-of-action (and not just report membrane leakage);
  - Capability to inform patient prognosis and differentiate high-risk from low-risk occurrences of seemingly similar ALT elevations;
  - Translational biomarkers with sensitivity and specificity for biliary hyperplasia.

Vascular injury
- Biomarkers or imaging approaches are needed to inform acute vessel damage, including degeneration, inflammation and hemorrhage, and also to allay hypothetical concerns over atherosclerotic plaque progression, thrombus formation and destabilization.

Neurotoxicity
- Accessible TSBs (including imaging) of neuronal injury including CNS pathologies as well as peripheral neuropathies are needed.

Pancreatitis
- Early TSBs with improved sensitivity over amylase and lipase are needed.

Tissue fibrosis
- Accessible biomarkers or imaging approaches could address tissue specific safety concerns relating to chronic organ injury and fibrosis, as well as perhaps even monitor disease progression and response to intervention (e.g., liver fibrosis and hepatitis C).

Bone & cartilage damage
- Arthropathies may follow destruction of chondrocytes and joint cartilage; qualified biomarkers are needed to address treatment-related concerns regarding acceleration of osteoporosis and enhanced risk for bone fracture.

Phospholipidosis
- Phospholipidosis histomorphological changes may be seen in rodent toxicology studies, and often not in other species. Human relevance and toxicological significance remain questionable. A satisfactory monitorable biomarker strategy could enable drug development and dispel lingering concerns over long-term patient safety.

Testicular injury
- Such findings may limit early clinical trial designs to females, excluding males, until greater understanding of the effect is achieved, or until a back-up compound devoid of such findings is identified.

Gastrointestinal
- Accessible mucosal damage biomarkers, for example, that could specifically alert this safety concern may reduce the need for endoscopic examinations in clinical trials and assist with patient enrollment.

Tumorigenesis
- A diverse set of qualified biomarker panels that could be integrated in the course of drug development animal toxicology studies in order to elucidate mode-of-action for nongenotoxic compounds and enable human relevance assessments using specimens collected from both the animal studies as well as matching duration clinical trials.

ALT: DEFINE.

strategically and opportunistically collecting critical specimens longitudinally from clinical studies, which will allow efficient assessments of normal biological variation in new safety biomarker baselines, and quick engagement of biomarker sensitivity and specificity evaluations on certain studies in order to gain more experience with new promising TSBs.

Proteins have historically been measured from accessible patient fluid samples as enzymatic activity, by chromatographic separation and mass spectrometric detection, or by immunoassay. Accessible fluids will be measured for new TSBs that will include miRNA, more novel proteins and critical new intermediary biochemical metabolites.

Such biomarkers will be multiplexed using measurement platforms in both animal toxicology studies and in early clinical trials that will allow not only indications of whether or not damage has occurred, but also an earlier and more sensitive awareness of mode of action, of activated pathophysiologic processes, and a better differential patient prognosis with insight into molecular toxicologic response pathways.

Point-of-care biomarker measurement devices are likely to gain more frequent use in clinical trials conducted on an outpatient basis, allowing instantaneous electronic transmission of results, a more complete and rigorous time-course assessment of safety biomarker responses and real-time integration with all other pertinent patient data and prior knowledge of compound properties.

Imaging will be integrated with the deployment of new accessible TSBs more progressively in animal toxicology studies to provide a more complete justification for the safe continuation of clinical trials using fewer animals, and convincingly and safely address questions of human relevance of suspect animal toxicology findings.

Exploratory clinical trials will be performed with greater frequency, with modifications allowing safe progression to doses and exposures that will allow and enable investigators to answer questions using new TSBs concerning the relevance to humans of questionable animal toxicology study safety findings where TSBs are shown to provide safety monitoring assurances.

New qualified TSBs will be investigated for their utility for monitoring disease progression, response to new treatments and disease regression, and also find utility in certain contexts as drug efficacy biomarkers [46].

Executive summary

- Animal toxicology study findings are imperfect predictors of human safety. There is hope that new regulatory qualified translational safety biomarkers (TSBs) will enable new drug development to reduce attrition rates and lower costs of drug development by:
  - Ensuring safe monitorability in the clinic of more animal toxicology findings suspected to be irrelevant to human
  - Providing greater prognostic dimension to conventional safety biomarkers and better inform earlier go, no-go drug development decision points.
- Patient health, business and scientific arguments are made in favor of qualifying new TSBs to support new drug development. These include:
  - Eliminating the need for repeating, or avoiding unwarranted progression of just one Phase 1 or 2 trial will save over US$31–42 million in development expenses when specific safety concerns can confidently be addressed only by adding data from new qualified TSBs;
  - Ten hypothetical but reasonable case examples are described typical of drug development scenarios where new higher performing TSBs could ensure patient safety when conventional biomarkers could not, resulting in significant savings in resources and development time.
- The business and scientific arguments against investing in qualifying new TSBs for drug development are critically discussed. These include:
  - The higher assay costs associated with new as compared with conventional biomarkers;
  - The time commitment and monetary investment needed to qualify TSBs for regulatory acceptance are substantial;
  - Concerns that potential requests from regulatory authorities for a premature deployment and or disallowing the appropriate application of recently introduced qualified TSBs could jeopardize, rather than enable, certain development programs.
- Sponsors must weigh up numerous conflicting factors when trying to decide whether to contribute data or resources to numerous consortia initiatives dedicated to regulatory qualification of new TSB qualification. A framework is needed to add leverage to and catalyze the growing interest and momentum toward TSB qualification in order to maximize efficiencies and minimize resource demands.
- Evidence indicates that the mechanism remains unresolved for assigning value and incorporating data from completed clinical and preclinical studies, and integrating the weight of evidence drawn from similarities in biomarker behavior across species to formulate a regulatory opinion regarding TSB qualification status.
- If biomarker qualification processes continue to be very cumbersome and costly, and are not individualized, future progress will be stymied. There is a need for resolution in a globally acceptable manner.

Leadership from drug development industry, regulatory organizations, not-for-profit consortia support organizations, professional academic societies, and government funding institutions will optimize execution strategies to support the many collaboration opportunities that may exist to qualify promising new and emerging TSB candidates.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as:

- ▪ of interest
- ▪▪ of considerable interest

1 Wagner JA. Strategic approach to fit-for-purpose biomarkers in drug development. *Annu. Rev. Pharmacol. Toxicol.* 48, 22.1–22.21 (2008).

▪▪ **Comprehensive review describing the principles of biomarker qualification for drug development.**

2 Sistare FD, Troth SP, Holder DJ *et al.* Towards consensus practices to qualify safety biomarkers for use in early drug development. *Nat. Biotechnol.* 28, 446–454 (2010).

▪ **The principles established by a consortium that succeeded in qualifying new kidney safety biomarkers.**

3 Wang K, Zhang S, Marzolf B *et al.* Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl Acad. Sci. USA* 106, 4402–4407 (2009).

4 Laterza OF, Lim L, Garrett-Engle PW *et al.* Plasma microRNAs as sensitive and specific biomarkers of tissue injury. *Clin. Chem.* 55, 1977–1983 (2009).

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

6 Schuster D, Laggner C, Langer T. Why drugs fail – a study on side effects in new chemical entities. *Curr. Pharmaceutical Design* 11, 3545–3559 (2005).

7 Olsen H, Betton G, Robinson D *et al.* Concordance of the toxicity of pharmaceuticals in humans and animals. *Reg. Toxicol. Pharmacol.* 32, 56–67 (2000).

▪ **Seminal manuscript describing the results from a collaboration of pharmaceutical companies conducting a retrospective analysis of 150 compounds in drug development with both animal toxicology and clinical trial data.**

8 Watkins PB, Kaplowitz N, Slattery JT *et al.* Aminotransferase elevations in healthy adults receiving 4 grams of acetaminophen daily: a randomized controlled trial. *JAMA* 296, 87–93 (2006).

9 DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22, 151–185 (2003).

10 DiMasi JA, Grabowski HG. The cost of biopharmaceutical R&D: is biotech different? *Managerial Decision Economics* 28, 469–479 (2007).

11 DiMasi JA, 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).

12 Sistare FD, DeGeorge JJ. Applications of genomics to nonclinical drug development: regulatory science considerations. *Methods Mol. Biol.* 460, 239–261 (2008).

13 O'Brien PJ, Irwin W, Diaz D *et al.* High concordance of drug-induced human hepatotoxicity with *in vitro* cytotoxicity measured in a novel cell-based model using high content screening. *Arch. Toxicol.* 80, 580–604 (2006).

14 Dambach DM, Andrews DA, Moulin F. New technologies and screening strategies for hepatotoxicity: use of *in vitro* models. *Toxicol. Pathol.* 33, 17–26 (2005).

15 Dix, DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95, 5–12 (2007).

16 Parham F, Austin C, Southall N, Huang R, Tice R, Portier C. Dose–response modeling of high-throughput screening data. *J. Biomol. Screen.* 14, 1216–1227 (2009).

17 Goozner M. Failed drug candidates getting re-screened by EPA to find toxicity markers. *The Pink Sheet*, July 26, 34–35 (2010).

18 Garthoff B, Hoffmann K, Luckhaus G, Thurau K: Adequate substitution with electrolytes in toxicological testing of "loop" diuretics in the dog. *Toxicol. Applied Pharmacol.* 65, 191–202 (1982).

19 Choudhury D, Ahmed Z. Drug-associated renal dysfunction and injury. *Nat. Clin. Practice* 2, 80–91 (2006).

20 Dieterle F, Perentes E, Cordier A *et al.* Urinary clusterin, cystatin C, β2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat. Biotechnol.* 28, 463–469 (2010).

21 Vaidya VS, Ozer JS, Dieterle F *et al.* Kidney injury molecule-1 outperforms traditional biomarkers of kidney injury in preclinical biomarker qualification studies. *Nat. Biotechnol.* 28, 478–485 (2010).

22 Yu Y, Jin H, Holder D *et al.* Urinary biomarkers trefoil factor 3 and albumin enable early detection of kidney tubular injury. *Nat. Biotechnol.* 28, 470–477 (2010).

23 Bonventre JV, Vaidya VS, Schmouder R, Feig P, Dieterle F. Next-generation biomarkers for detecting kidney toxicity. *Nat. Biotechnol.* 28, 436–440 (2010).

24 Arnold JMO, Howlett JG, Dorian P *et al.* Canadian Cardiovascular Society Consensus Conference recommendations on heart failure update 2007: prevention, management during intercurrent illness or acute decompensation, and use of biomarkers. *Can. J. Cardiol.* 23, 21–45 (2007).

25 Urboniene D, Haber I, Fang Y-H, Thenappan T, Archer SL. Validation of high-resolution echocardiography and magnetic resonance imaging vs. high-fidelity catheterization in experimental pulmonary hypertension. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 299, L401–L412 (2010).

26 Wang I-M. Development of a preclinical model to predict reactogenicity of vaccines/adjuvants in humans. Presented at: *The 2010 Keystone Symposium on Immunological Mechanisms of Vaccination*. Seattle, WA, USA, 27 October–1 November 2010.

27 Dieterle F, Sistare F, Goodsaid F *et al.* Renal biomarker qualification submission: a dialog between the FDA-EMEA and Predictive Safety Testing Consortium. *Nat. Biotechnol.* 28, 455–462 (2010).

▪▪ **A view of the differences in viewpoints between regulatory authorities and the pharmaceutical industry, which both stand to benefit from the qualification of new**

safety biomarkers.

28 Dockray G, Dimaline R, Varro A. Gastrin: old hormone, new functions. *Pflu. Arch. Eur. J. Physiol*. 449, 344–355 (2005).

29 Burek JD, Patrick DH, Gerson RJ. Weight of biological evidence for assessing carcinogenicity. In: *Carcinogenicity*. Grice HC, Cimina JL (Eds). Springer-Verlag, New York, NY, USA, 83–95 (1998).

30 Sun DQ, Hamlin D, Butterfield A, Watson D, Smith H. Electrochemiluminescent immunoassay for rat skeletal troponin I (Tnni2) in serum. *J. Pharmacol. Toxicol. Methods* 61, 52–58 (2010).

31 Onuoha GN, Alpar EK, Dean B *et al.* Skeletal troponin-I release in orthopedic and soft tissue injuries. *J. Orthop. Sci.* 6, 11–15 (2001).

32 Simpson JA, Labugger R, Hesketh GG *et al.* Differential detection of skeletal troponin I isoforms in serum of a patient with rhabdomyolysis: markers of muscle injury? *Clin. Chem.* 48, 1112–1114 (2002).

33 Brown AP. Development of serum calcium and phosphorus as clinical biomarkers for drug-induced systemic mineralization: a case study with a MEK inhibitor. In: *Biomarkers in Drug Deevlopment: A Handbook of Practice, Application, and Strategy.* Bleavins MR, Carini C, Jurima-Romet M, Rahbari R (Eds). John Wiley and Sons, Inc., Hoboken, NJ, USA (2010).

34 Williams SA, Slavin DE, Wagner JA, Webster CJ. A cost–effectiveness approach to the qualification and acceptance of biomarkers. *Nat. Rev. Drug Discov.* 5, 897–902 (2006).

35 Espeland M, O'Leary D, Terry J, Morgan T, Evans G, Mudra H. Carotid intima-medial thickness as a surrogate for cardiovascular disease events in trials of HMG-CoA reductase inhibitors. *Curr. Control. Trials Cardiovasc. Med.* 6, 1–6 (2005).

36 Mattes WB, Gribble Walker E, Abadie E *et al.* Research at the interface of industry, academia and regulatory science. *Nat. Biotechnol.* 28, 432–433 (2010).

37 Chetty RK, Ozer JS, Lanevschi A *et al.* A systematic approach to preclinical and clinical safety biomarker qualification incorporating Bradford Hill's principles of causality association. *Clin. Pharmacol. Therap.* 88, 260–262 (2010).

38 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* 8, 431–440 (1989).

39 Altar CA, Amakye D, Buonos D *et al.* A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clin. Pharmacol. Ther.* 83, 368–371 (2008).

40 Cummins B, Auckland ML, Cummins P. Cardiac-specific troponin-I radioimmunassay in the diagnosis of acute myocardial infarction. *Am. Heart J.* 113, 1333–1344 (1987).

41 O'Brien PJ. Cardiac troponin Is the most effective translational safety biomarker for myocardial injury in cardiotoxicity. *Toxicology* 245, 206–218 (2008).

42 Alpert JS, Antman E, Apple F *et al.* Myocardial infarction redefined – a consensus document of the Joint European Society of Cardiology/ American College of Cardiology Committee for the Redefinition of Myocardial Infarction. *J. Am. Coll. Cardiol.* 36, 959–969 (2000).

43 Betton GR, Dormer CS, Wells T, Pert P, Price CA, Buckley P. Gastric ECL-cell hyperplasia and carcinoids in rodents following chronic administration of H2-antagonists SK&F-93479 and oxmetidine and omeprazole. *Toxicol. Pathol.* 16 288–298 (1988).

44 Vinken M, Doktorova T, Ellinger-Ziegelbauer H *et al.* The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based *in vitro* carcinogenicity screening assays. *Mutat. Res. Rev. Mutat. Res.* 659, 202–210 (2008).

45 Suter L, Schroeder S, Meyer K *et al.* EU framework 6 project: predictive toxicology (PredTox) – overview and outcome. *Toxicol. Appl. Pharmacol.* 252(2), 73–84 (2010).

46 Sistare FD, Devarajan P, Blank M *et al.* Assessing and predicting kidney safety. In: *Accelerating the Development of Biomarkers for Drug Safety: Workshop Summary*. National Academies Press, Washington, DC, USA, 29 (2009).

Websites

101 Austin CP, Kavlock R, Tice RR. Tox21: putting a lens on the vision of toxicity testing in the 21st century http://alttox.org/ttrc/overarching-challenges/way-forward/austin-kavlock-tice

102 Code of Federal Regulations. Title 21 Food and Drugs, Chapter I Food and Drug Administration Department of Health and Human Resources, Subchapter D Drugs for Human use, Part 312 Investigational New Drug Application, Subpart B Investigational New Drug Application http://a257.g.akamaitech.net/7/257/2422/26mar20071500/edocket.access.gpo.gov/cfr_2007/aprqtr/pdf/21cfr312.23.pdf

103 International Conference on Harmonization ICH Guideline M3 (R2): Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/M3_R2/Presentation/A._Jacobs_ICH_M3_R2__presentation_to_the_GCG_and_the_Tokyo_Symposium_6–09.pdf

104 FDA Guidance for Industry – drug-induced liver injury: premarketing clinical evaluation www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/UCM174090.pdf

105 Critical Path Institute PSTC Qualified Biomarkers www.c-path.org/PSTCRegulatory.cfm

106 ICH Guideline E16 Genomic biomarkers related to drug response: context, structure and format of qualification submissions www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500097060 - 2010–2009–24

107 FDA Draft Guidance on a Qualification Process for Drug Development Tools (2010) www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf

108 International Conference on Harmonization. Guidance for Industry, E9 statistical principles for clinical trials www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073137.pdf

109 IMI SAFE-T www.imi-safe-t.eu/biomarker/drug-induced-injury/imi

110 ILSI HESI Biomarkers of Nephrotoxicity Project Committee www.hesiglobal.org/i4a/pages/Index.cfm?pageID=3430

111 Biomarker Consortium www.fnih.org/Biomarkers%20Consortium/Biomarkers_home.shtml