OPINION

Considerations for the successful co-development of targeted cancer therapies and companion diagnostics

Jane Fridlyand, Richard M. Simon, Jessica C. Walrath, Nancy Roach, Richard Buller, David P. Schenkein, Keith T. Flaherty, Jeff D. Allen, Ellen V. Sigal and Howard I. Scher

Abstract | As diagnostic tests become increasingly important for optimizing the use of drugs to treat cancers, the co-development of a targeted therapy and its companion diagnostic test is becoming more prevalent and necessary. In July 2011, the US Food and Drug Administration released a draft guidance that gave the agency's formal definition of companion diagnostics and introduced a drug–diagnostic co-development process for gaining regulatory approval. Here, we identify areas of drug–diagnostic co-development that were either not covered by the guidance or that would benefit from increased granularity, including how to determine when clinical studies should be limited to biomarker-positive patients, defining the diagnostically selected patient population in which to use a companion diagnostic, and defining and clinically validating a biomarker signature for assays that use more than one biomarker. We propose potential approaches that sponsors could use to deal with these challenges and provide strategies to help guide the future co-development of drugs and diagnostics.

Molecular diagnostic tests are being used with increasing frequency, especially in oncology[1], where they are used to identify the genes, proteins or pathways that are disrupted during tumorigenesis. Advances in molecular biology and an improved understanding of cancer pathogenesis have helped to identify numerous new potential cancer drug targets. Drugs that are directed at these targets — often referred to as targeted therapeutics — interact with mutated proteins or other specific molecules to disrupt a molecular signalling pathway on which the cancer cells depend for their growth and/or survival.

There are several examples of such drugs[2]: the tyrosine kinase inhibitor imatinib mesylate (Gleevec; Novartis) inhibits the Philadelphia chromosome fusion product in chronic myelogenous leukaemia[3] and inhibits KIT (also known as CD117) in gastrointestinal stromal tumours[4] and other solid tumours; the monoclonal antibody trastuzumab (Herceptin; Genentech) binds to human epidermal growth factor receptor 2 (HER2; also known as ERBB2) and is approved in some breast cancers as well as gastric or gastroesophageal cancers[5,6]; and the small-molecule tyrosine kinase inhibitor sunitinib (Sutent; Pfizer) blocks vascular endothelial growth factor (VEGF) and platelet-derived growth factor receptor (PDGFR) signalling to inhibit angiogenesis and cell proliferation, and is approved for the treatment of patients with metastatic renal cell carcinoma and other tumours[7,8].

Because some of these drugs may show more activity or only be active in subsets of tumours that have been identified by predictive biomarkers or biomarker signatures, it has become increasingly important to identify those patients who are most likely to benefit from a given therapy. Therefore, the need to co-develop diagnostic devices to measure the biomarker or biomarker signature in order to properly treat patients is also increasingly important. The US Food and Drug Administration (FDA) refers to these as "*in vitro* diagnostic (IVD) companion diagnostic devices", which it formally defines as a "diagnostic device that provides information that is essential for the safe and effective use of a corresponding therapeutic product"[9]. Throughout this article we refer to these diagnostic devices as companion diagnostics.

The development of new drugs is increasingly becoming a co-development process involving a targeted therapeutic and its companion diagnostic, because the test results obtained from the use of the diagnostic are essential to properly guide therapy. Appropriate use of the drug depends on the diagnostic test result; analytical failure of the companion diagnostic could result in the misidentification of patients who are eligible for treatment, thus potentially denying the drug to patients who would have benefited from it or, conversely, subjecting patients to toxicities without a resulting benefit. Therefore, the FDA requires that the safety and effectiveness of both the drug and the companion diagnostic are assessed (for regulation in the European Union, see BOX 1).

A company that is applying for FDA approval of a companion diagnostic must provide evidence that the assay to be used as the companion diagnostic is analytically validated; that is, the technical performance of the device and/or diagnostic test and the range of conditions under which it provides reproducible and reliable results should have been determined. The company must also provide evidence that the companion diagnostic has demonstrated clinical validity (evidence that there is an association between the biomarker and a clinical outcome) and utility (evidence that treatment based on a diagnostic assay results in measurable improved patient outcomes relative to the currently available therapy). These validations help to ensure that the companion IVD will successfully select the appropriate patient populations in therapeutic clinical trials[10].

Box 1 | **Regulation of companion diagnostics in the European Union**

Like the US Food and Drug Administration (FDA), the European Medicines Agency (EMA) encourages the use of companion diagnostics in drug development, and requires that testing with companion diagnostics is performed for the prescription of certain targeted therapies (such as crizotinib (Xalkori; Pfizer) and vemurafenib (Zelboraf; Roche/Plexxikon))[1,13]. However, the EMA has taken a less active leadership role in the regulation of companion diagnostics, at least in part because the agency does not have regulatory authority over *in vitro* diagnostics. In both the United States and the European Union, companion diagnostics are considered to be *in vitro* diagnostics, but *in vitro* diagnostics are regulated separately from other medical devices in the European Union. In contrast to pharmaceutical products, which receive central approval by the EMA, the approval of diagnostics occurs separately in each European country[45]. Diagnostics are regulated under the European Directive 98/79/EC on *in vitro* diagnostic medical devices (known as the *in vitro* diagnostics directive), and devices that receive the CE (Conformité Européene) mark — a mandatory marking for products sold in the European Union — can be distributed to all members of the European Union[46]. The approval of companion diagnostics is based on a process whereby the manufacturer itself determines whether the product meets a set of harmonized standards for *in vitro* diagnostics, and affixes the CE mark. In addition, devices must "work as advertised" and meet legislative quality standards but they are not assessed for efficacy[45].

Therefore, the European Union is planning substantial reform to strengthen the current regulatory system, especially in the areas of oversight, transparency and rigor. Although many feel that the current system is fundamentally sound[45], the weakness of the *in vitro* diagnostics directive have also been revealed following more than 10 years of its implementation[47]. For example, there is generally unanimous consent among stakeholders that, unlike under current law, *in vitro* diagnostics should be reviewed by a notified body[47]. These reforms began in 2008 in order to update the role of notified bodies in device regulation, as well as to adjust to changing scientific and technological capabilities and more properly incorporate new diagnostic devices, such as companion diagnostics, into the framework.

In late 2012, the European Commission submitted a set of proposals on the innovation and regulation of medical devices[48] to the European Parliament, and a series of votes on the amendments were scheduled throughout 2013, with the first now on 18 September (having been postponed from 10 July). The updated European regulatory framework maintains the separation of the regulation of medical devices from that of *in vitro* diagnostics. Co-diagnostics will be classified as high-risk devices (that is, they could cause serious risk of injury or death if they did not function properly) that require regulation by a notified body and stricter compliance standards. There will also be increased requirements for the demonstration of clinical evidence. Importantly, both new and existing devices would need to meet these new requirements.

The European Commission's proposal also includes the implementation of the Global Harmonization Task Force's benefit–risk classification system[49]. European Union regulators consider this classification — which has already been adopted by other countries[47] — to be a good model, and its adoption by the European Union would be a good step towards the international harmonization of medical device regulations. The FDA, EMA and Japanese regulators have also been involved in efforts to communicate on companion diagnostic regulation.

To address the challenges associated with the co-development of drugs and companion diagnostics, in July 2011 the FDA distributed a guidance document for comment purposes[9]. The goals of the draft guidance were to define IVD companion diagnostic devices, to clarify the need for FDA oversight and approval for their safe and effective use, and to provide direction for both industry and FDA staff on possible developmental pathways and the approval requirements for the labelling of therapies that require IVD companion diagnostic devices[9].

In this article, we address issues related to the co-development of drugs and companion diagnostics that were not discussed in the FDA's July 2011 draft guidance (in addition, many of these suggestions were further reinforced in a later FDA draft guidance on enrichment strategies)[11] and propose approaches for drug–diagnostic co-development that introduce flexibility and provide options for present-day drug development. We begin to form a consensus opinion on some of the most important questions faced by sponsors and regulators during the drug–diagnostic co-development process. This has led to the development of a more detailed blueprint for drug–diagnostic co-development that includes a framework upon which decisions related to clinical trial design and optimal identification of patient

populations can be based, as well as a discussion of issues that may arise during the development of biomarker assays.

As we continue to better utilize our knowledge of genomics and more effectively apply it to biological processes, drug–diagnostic co-development will be useful in many other biomedical areas, including neurology, virology and infectious diseases. However, because oncology is currently one of the most active disease areas in drug–diagnostic co-development[1], we use examples from oncology in this article to illustrate current and future approaches towards the development of companion diagnostics.

Challenges in co-development

Sponsors face several diverse challenges during the co-development process, many of which will be based on the type of diagnostic test being used for biomarker assessment. The approach towards the development and approval process will need to be individualized for each disease and each therapeutic target as well as the type of biomarker and the type of assay that is needed to measure the biomarker (TABLE 1). Although the different types of diagnostic assays (such as sequencing, quantitative PCR, immunohistochemistry, quantitative real-time PCR and fluorescent *in situ* hybridization) that are used to identify patients for treatment encounter different issues and requirements for analytical validation, the core principles regarding clinical validation and clinical utility are similar. However, some biomarkers and types of assays will be more challenging for co-development than others (FIG. 1); for example, diagnostics consisting of bimodal biomarkers identified by sequencing will generally be less challenging than continuous biomarkers assayed by real-time PCR, which will be less challenging than multi-marker assays that use a biomarker signature (which could involve a combination of next-generation sequencing, immunohistochemistry and PCR).

This article highlights these common factors that sponsors may need to address throughout the drug–diagnostic co-development process. The FDA has reiterated the importance of communication with the agency early in the process[9], and flexibility from both sponsors and the FDA will be required throughout the development process[12].

Following the publication of the FDA's companion diagnostic draft guidance, sponsors and other stakeholders provided feedback to the agency on the guidance. Friends of Cancer Research solicited this input and

Table 1 | **Diagnostic assays used in oncology***

Biomarker	Platform or technology	Diagnostic test value type	Drug and diagnostic examples	Challenges
Mutation or mutations	Sequencing	Generally binary	No current examples	Test results depend on the percentage of cells with mutations (that is, there is a lower detection limit); may measure non-specific exons
Mutation or mutations	Quantitative PCR	Generally binary	Cobas 4800 BRAF V600 Mutation Test; TheraScreen K-RAS Mutation Kit	Test results depend on the percentage of mutant sequences, adequate specimen integrity and sufficient DNA to be detected
Protein expression	Immuno-histochemistry staining	• Generally continuous based on the intensity and proportion of cells with the given intensity; ordinal intensity scoring of currently approved tests • Could be treated as binary if the diagnostic readout is a complete loss of signal (for example, a H-score[‡] of zero)	Dako HercepTest (detects HER2 protein expression)	Generally semi-quantitative and non-automated evaluation; test results can depend on pre-analytical tissue processing factors
Gene expression	Quantitative real-time PCR	Generally continuous	No current examples	Manual macrodissection may be necessary for samples with low tumour cell content
DNA copy number	FISH or chromogenic in situ hybridization	Generally continuous; could be treated as binary if the diagnostic is a complete loss of copy number or high-level amplification	HER2 FISH pharmDx Kit; PathVysion HER-2 DNA Probe Kit; Her2 Dual ISH DNA Probe Kit	Relatively complex assay technology and interpretation
Fusion protein product	FISH	Threshold is set at specific percentage of cells; essentially a bimodal distribution	Vysis ALK Break Apart FISH Probe Kit	Relatively complex assay technology and interpretation
Gene signature	Next-generation sequencing	Could be treated as binary based on gene signature	No current examples	Complex assay technology and interpretation

ALK, anaplastic lymphoma kinase; FISH, fluorescence in situ hybridization; HER2, human epidermal growth factor receptor 2 (also known as ERBB2).*This table shows the types of diagnostic assays used in oncology for each type of biomarker and their potential challenges in determining what constitutes a 'biomarker-positive' or 'biomarker-negative' readout. ‡The H-score is a semi-quantitative intensity scale used to describe immunohistochemistry staining, and is calculated by the weighted combination of staining intensities of the cells and the proportion of cells stained at a given intensity.

requested any additional questions that may arise during the co-development process that were not addressed sufficiently.

Issues raised included the following questions: when should studies be restricted to diagnostically selected patients and how should this decision be made? If a drug label is intended for unselected patients, how can biomarker information be included in this label when the diagnostic is evaluated in the registration study? What are the possible approaches for utilizing assays that have traditionally been used as screening platforms, such as next-generation sequencing, in drug–diagnostic clinical trials that are intended for registration? How should summary measures (the biomarker signature or the composite 'score' of several biomarkers) in assays that measure more than one biomarker (multi-marker assays) be defined, clinically validated and thresholds reassessed?

And, for biomarkers that are not clearly binary: how can Phase III studies be conducted without pre-specifying a threshold cut-off value for biomarker selection? And how can the threshold value be readjusted from a pre-specified value and how may it be pre-planned?

In December 2012, the FDA released a draft guidance on enrichment strategies for clinical trials[11], which provided information on the use of predictive enrichment strategies to select those patients who are most likely to respond to a particular therapy. The enrichment draft guidance highlights several points of interest for drug–diagnostic co-development clinical trials. For example, the decision to conduct an enrichment trial is mainly a sponsor decision; prospective stratified randomization of patients with regard to the biomarker may not be essential; the classifier may be explored in the pivotal trial for regulatory filing purposes; and data from non-randomized and/or non-registration trials may be used to understand the extent of efficacy in biomarker-negative patients. The draft guidance on clinical trial enrichment also discusses, and comes to a similar conclusion on, many of the key issues

discussed in this article that are relevant to the drug–diagnostic co-development process, such as genomic strategies that could be used for predictive enrichment and choosing which patient populations to study.

Timelines of development
The July 2011 draft guidance states that a therapeutic and its companion diagnostic device would ideally be developed simultaneously, and the clinical performance and clinical significance of the companion diagnostic would be established using data from the clinical development programme of the corresponding therapeutic product[9]. Although the timelines for this co-development were not addressed in the 2011 guidance, suggestions have been made by several groups, including the FDA and the European Medicines Agency[12–14]. For the correct timing of co-development and the concomitant approval of both the therapy and the companion diagnostic, biomarker selection and quantification as well as analytical validation of an assay for
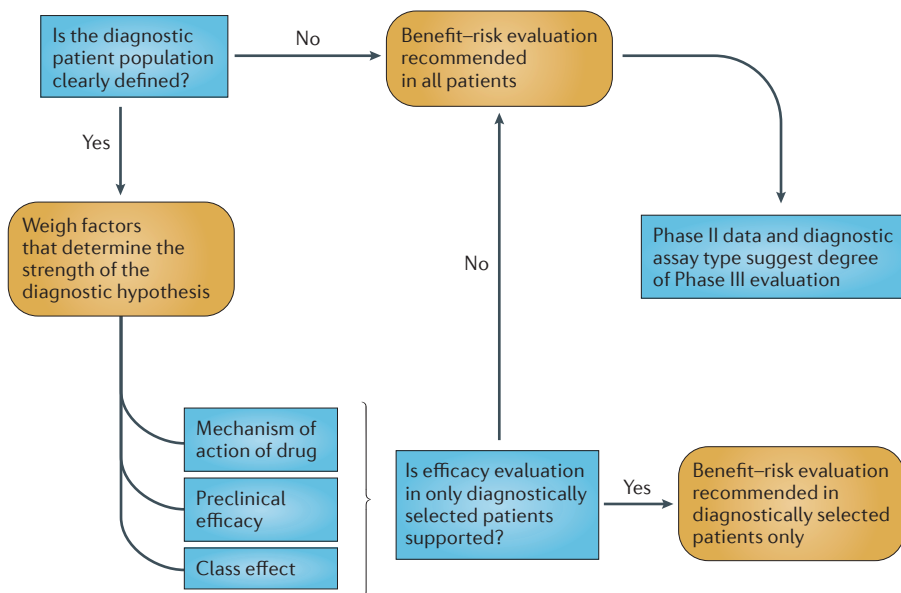
Figure 1 | **Decision tree for determining the evaluation of biomarker-negative patients.** This high-level flow chart summarizes the factors that affect the extent to which all patients, regardless of biomarker status, should be evaluated in the development programme of a drug and its companion diagnostic.

measuring the biomarker (all of which are separate but interconnected issues) would ideally be performed prior to conducting pivotal clinical trials. Moreover, the diagnostic assay would be locked down (that is, analytically and clinically validated and considered to be final) and available prior to the initiation of Phase III clinical trials[15]. Clinical validation of the companion diagnostic could then be carried out in the pivotal clinical trial or trials.

However, this is not always possible and several issues discussed in this article relate to the development status of the assay at the stage of Phase III trial initiation. When a validated assay is not available, it may not be possible to restrict the enrolment of patients to a diagnostically selected subset. In addition, if the assay is not ready or if the cut-off point defining diagnostically selected patients has not yet been determined, an analysis algorithm will need to be developed prospectively prior to designing a Phase III trial that includes both biomarker-positive and biomarker-negative patients.

Positive and negative diagnostic selection
Similar to the FDA's enrichment guidance, this article is generally structured to provide information on biomarker-positive patients; a biomarker-positive result indicates that a patient in the subset will be more likely to respond to a given drug. However,

diagnostics are also used to identify those patients who are least likely to show a favourable risk–benefit profile (for example, patients who are least likely to benefit from a treatment or most likely to develop adverse drug reactions) and to exclude them from treatment. Biomarker selection can improve treatment outcomes and patient care either by selecting the best candidates for therapy or by eliminating unsuitable candidates. Generally, the approaches towards patient selection and clinical trial enrichment will be similar in both circumstances, so the design of the clinical study must take into account whether patient selection is based on a biomarker-positive result or a biomarker-negative result.

Strategy for selected populations
One of the key questions facing a sponsor is whether to include both biomarker-positive and biomarker-negative patients in its clinical trials or to restrict clinical development to the diagnostic category that is considered to be most likely to benefit from the new drug (that is, biomarker-positive patients). If the clinical trial is restricted to biomarker-positive patients, then it must also be decided at what point or stage of development it is reasonable to make restrictions related to patient inclusion. There is potential added value in understanding the effect of a drug in all patients who are

stratified by their diagnostic status, including an assessment of potential safety effects in the biomarker-negative patients.

However, stakeholders — including the FDA — recognize the benefits associated with the efficiency of including only biomarker-positive patients. Therefore, the FDA does allow sponsors to limit the target patient population to biomarker-positive patients. Limiting a clinical trial to diagnostic-positive patients can be seen in the recent FDA approvals of two cancer drugs, which demonstrate that the FDA is amenable to this approach. Vemurafenib (Zelboraf; Roche/Plexxikon) was given full approval in patients with melanoma carrying the $BRAF^{V600E}$ mutation[16], and crizotinib (Xalkori; Pfizer) was given accelerated approval in patients with anaplastic lymphoma kinase (ALK)-positive non-small-cell lung cancer[17]. Both vemurafenib and crizotinib received approvals on the condition that post-marketing studies would be undertaken in patients for whom the treatment is currently not indicated; Roche has made a commitment to evaluate the efficacy of vemurafenib in patients with the $BRAF^{V600K}$ mutation who give a negative result for $BRAF^{V600E}$ with the companion diagnostic, and Pfizer has made a commitment to study crizotinib in patients who are ALK-negative with the companion diagnostic.

In general, decisions made during the development process are guided by the need to maximize the exposure of those patients who are likely to benefit from a drug and minimize the exposure of those patients who are not likely to benefit, or who are at an increased risk of developing serious adverse events. Full evaluation of drug efficacy and safety in both biomarker-positive and biomarker-negative patients ensures that the regulators, patients, physicians, payers and sponsors have sufficient data to make empirically based decisions on the drug's benefit–risk profile. Conversely, all drugs cause some toxicity and the exposure of patients to potential toxicity can only be justified if there is a reasonable potential for therapeutic benefit (that is, equipoise is maintained). If there is a sufficiently strong hypothesis based on data showing drug activity in diagnostically selected patients, the evaluation of biomarker-negative patients might even be considered unethical if other treatment options are available. We believe the decision of whether or not it is appropriate to include biomarker-negative patients in a clinical trial primarily depends on the strength of the science

that is available to support the diagnostic hypothesis that the drug is only active in biomarker-positive patients (including the drug's mechanism of action, its preclinical efficacy and, if known, its class effect), the potential for risk to patients, and the currently available clinical data.

A proposed decision-making strategy for determining whether and how the evaluation of diagnostically negative patients could be undertaken is shown in FIG. 1. Although the evaluation of diagnostically negative patients is not always required by the FDA, especially for cancer therapeutics, this method suggests including a variety of factors — such as mechanism of action and preclinical efficacy — in the decision-making process. The outcome from the flowchart may evolve as new data are developed to refine the diagnostic hypothesis. Regardless of which decision is made, the analytical performance of the diagnostic assay must be sufficient (or fit for purpose) to use in any clinical registration trial. If an acceptable assay is not available and sufficiently characterized prior to the registration trial, it is not possible to evaluate only biomarker-positive patients, and a clinical trial design that includes both biomarker-positive and biomarker-negative patients must be used.

The decision-making strategy illustrated in FIG. 1 provides a recommendation regarding the appropriate level of evidence that is needed before making the decision to not study biomarker-negative patients or to limit biomarker-negative patients to Phase I safety studies alone. TABLE 2 provides some examples of the key factors that influence the decision of whether to perform Phase II and more advanced studies in diagnostic-negative patients. Note that these categories should never be considered in isolation; rather, decisions should always take into account the totality of information. For example, although there is limited translation of results from preclinical efficacy studies to results from clinical trials in humans, these studies can help to inform the design of future trials, as described in TABLE 2.

Furthermore, it should be remembered that the scientific rationale or hypothesis upon which the use of a biomarker that is used for patient selection is based may not be correct and so the biomarker may not accurately predict therapeutic response. For example, it was assumed that EGFR expression would be a simple predictive biomarker that would predict response to the EGFR inhibitor cetuximab (Erbitux; Bristol-Myers Squibb/Eli Lilly). Initial clinical trials were conducted and therefore initial approval

Table 2	**Factors that influence the evaluation of biomarker-negative patients***		
Key factors	**Categories**	**Supports efficacy evaluation in diagnostic-selected patients only?**
Evidence supporting a clear definition of the diagnostically selected population	Compelling evidence supporting the threshold for the definition of a diagnostic-positive population (for example, mutation, complete loss of expression or high-level amplification)	Yes
	Threshold for diagnostic-positive population uncertain (for example, continuous or ordinal biomarker)	No
Mechanism of action	The diagnostic biomarker is an immediate drug target or a measure of the activation of a pathway of the drug target	Yes
	The diagnostic biomarker is less clearly related to the drug target	No
Preclinical efficacy (*in vitro* or *in vivo*)	The diagnostic has a high sensitivity and specificity for drug activity	Yes
	The diagnostic has a high sensitivity and lower specificity for drug activity	No
Class effect[‡]	Established class effect in diagnostic-selected populations	Yes
	First-in-class	No

*This table shows some examples of the multiple factors that influence the decision of whether to conduct efficacy evaluation in unselected patients. The terms 'efficacy' and 'benefit–risk' are used interchangeably.
[‡]Class effect refers to knowledge of the behaviour of existing drugs with similar mechanisms of action (or of targets, chemical structures or known pharmacological effects).

was given only in patients with metastatic colorectal cancer who had high EGFR expression[18]. Later studies demonstrated that EGFR expression was not the key determinant of response to cetuximab[19], which suggests that many patients who may have responded favourably to cetuximab were not given the opportunity to receive it. This example reinforces the necessity of a thorough decision-making process around the enrolment of biomarker-negative patients.

Benefit–risk assessment of a bimodal or continuous biomarker. If a decision has been made to evaluate drug efficacy in biomarker-positive and biomarker-negative patients in a Phase I or Phase II clinical trial (a proof-of-concept study for a drug and diagnostic), it needs to be decided whether only a limited benefit–risk assessment or a full benefit–risk assessment of biomarker-negative patients is warranted in the pivotal Phase III study. In TABLE 3, we outline the three main benefit–risk scenarios that can be potentially observed in Phase II trials (or in Phase I trials if a Phase I–III strategy is being considered) based on a bimodal predictive biomarker (for example, the presence or absence of a mutation) or continuous predictor (for example, the level of protein

expression), and provide recommendations for the type of evaluation of diagnostic-negative patients that should be undertaken in a Phase III trial.

The identification of a threshold that defines the biomarker-positive patient population is simplified if there is a bimodal distribution of the molecular biomarker in the targeted patient population — for example, when assessing the mutation status. However, gene expression and protein expression are continuously distributed, making the definition of a predictive biomarker iterative. There are currently no available widely accepted statistical approaches for designing and powering a clinical study to estimate a relevant threshold for a continuous biomarker[20]. Clinical data obtained from multiple studies, at least one of which is likely to involve the validation of the threshold, will probably be required for the final definition of the biomarker-positive threshold. The identification of a biomarker-positive threshold for a biomarker with a continuous distribution may be particularly difficult if evidence of harm to a biomarker-defined subset of patients emerges during early development. This would make it potentially unethical to enrol biomarker-positive and biomarker-negative patients into a pivotal clinical trial.

Table 3 | **Recommended Phase III evaluation of diagnostic-negative patients**

Definition of diagnostic-selected population	Observed clinical benefit–risk in Phase II (or Phase I)*		
	Clinically meaningful benefit–risk in diagnostically selected patients; detrimental benefit–risk in diagnostic-negative patients	Clinically meaningful benefit–risk in diagnostically selected patients; no clinically meaningful benefit–risk in diagnostic-negative patients	Clinically meaningful benefit–risk in all patients; biomarker shows better effects in diagnostically selected patients
There is a clear binary definition of a diagnostic-positive population (for example, mutation, complete loss of expression or high-level amplification)	–	There is no or very limited benefit–risk evaluation of diagnostic-negative patients	Fully assess benefit–risk in all patients
The threshold for the diagnostic-positive population is uncertain (for example, continuous or ordinal biomarkers such as real-time PCR expression or immunohistochemistry staining)	Limited safety assessment of the patients in whom the biomarker value is borderline with the proposed diagnostic threshold, dependent on the severity of the detriment	If there is concern with the strength of the biomarker–outcome relationship, gated or full benefit–risk evaluation of diagnostic-negative patients should be conducted to define the best threshold	Fully assess benefit–risk in all patients

*The three main Phase II benefit–risk scenarios are given, with recommendations for the Phase III evaluation of diagnostic-negative patients based on the types of currently available diagnostic tests.

This situation may be even more challenging when the predictive biomarker is a composite of multiple continuous variables, such as gene expression signatures. In this case, flexibility in the prospective hypothesis and/or study design may be required to avoid delays in drug approval, and some of these scenarios are outlined below.

Determining degree of Phase III evaluation of biomarker-negative patients. For both bimodal and continuous classifiers that are used to select and stratify patients for therapy, we outline three possible outcomes from Phase II clinical trials. First, the investigational drug leads to a clinically meaningful benefit in biomarker-positive patients and a detrimental effect in biomarker-negative patients. Second, there is an observed clinically meaningful benefit in biomarker-positive patients and no benefit in biomarker-negative patients. Third, there is a clinically meaningful benefit in all patients, with biomarker values correlating with the degree of the benefit.

TABLE 3 provides recommendations for continuing the evaluation of biomarker-negative patients in Phase III clinical trials based on the biomarker type (binary versus continuous) and the outcome of the Phase II study. It should be noted that there may not be a clear distinction between each patient population based on the confidence intervals that are associated with clinical end points from Phase II trials. Other groups have suggested and evaluated various methods, such as decision analysis and portfolio management

concepts, for predicting which patient population should be used in a pivotal study[15,21–24]. However, it is difficult to always incorporate clinical and strategic considerations into these analytical methods.

Sample size limitations of Phase II trials will always lead to uncertainties when defining benefit and detriment boundaries, and patient safety is of paramount importance if the possibility for detriment exists, as in the scenario outlined in the first benefit–risk column of TABLE 3, where a clinically meaningful benefit is observed in biomarker-positive patients and a detrimental effect is seen in biomarker-negative patients. The final drug and diagnostic labels will be determined by the results that are observed in the Phase III trial in these benefit–risk outcome scenarios. The planned role of the diagnostic in the indication and in the clinical sections of the eventual label should be taken into careful consideration when designing the Phase III trial[20].

Two-staged clinical development. A two-staged clinical development plan refers to initially testing the effect of the drug in the biomarker-positive patient population in which the probability of treatment success is deemed to be the highest. Based on the outcome of this evaluation, testing in biomarker-negative patients may be warranted. Such a two-staged approach may be appropriate, for example, when — prior to conducting the first pivotal clinical trial — the treatment is expected to have activity in biomarker-positive patients but it is uncertain

whether this activity will be restricted to such patients. Even targeted agents can often exert some pleiotropic effects (for example, a tyrosine kinase inhibitor may be highly active against its target tyrosine kinase while also affecting other tyrosine kinases). Therefore, if the level of specificity is uncertain, it may be reasonable to suspect that biomarker-negative patients could also benefit from the treatment.

A staged clinical development plan may also be appropriate when there is a continuum of varying interpatient expression levels of the target, yet the level of expression of the biomarker that is required for a targeted agent to have activity is not known. In a staged approach, the initial clinical trial or trials, and initial regulatory approval, would be in the biomarker-positive patient population as those are the patients who are believed to be most likely to benefit from the treatment. This allows for quicker drug approval and availability as well as the use of the therapy in the patient population that is considered to be most likely to benefit. Subsequent studies in the biomarker-negative patient population would then be carried out depending on emerging clinical data and scientific understanding to ensure that a potentially effective drug in multiple populations is not missed owing to scientific assumptions. Following initial approval based on randomized trials in a specific patient population, testing in other smaller patient groups — with patient response rate as a measure of drug efficacy — could then be used to expand the label[25].

Biomarker information in the label when the Phase III trial is not diagnostically restricted. When Phase III clinical trials are restricted to diagnostic-positive patients, the target population is clear and labelling of the drug will often be relatively straight-forward. When the Phase III clinical trials include both diagnostic-positive and diagnostic-negative patients, a prospective statistical analysis plan is essential and labelling can be more difficult. The prospective analysis plan should ensure that the pre-specified level of type I error in the study (for example, a two-sided α of 0.05) is preserved across multiple component hypothesis tests that might be conducted (for example, the overall patient population and a diagnostic-positive subset of patients). The prospective analysis plan should also indicate which hypothesis tests are to be performed, the sequence in which these tests will be performed, and the way in which they will be used to determine the intended patient population in the eventual label.

Some of the recommendations in the FDA draft guidance[9] are currently unclear, such as those regarding the inclusion of information that describes how clinical benefit relates to the biomarker status (for bimodal markers) or to the calculated numerical biomarker value (for continuous biomarkers) in the clinical or the mechanism of action sections of the label. One approach might stipulate that such an inclusion could be considered provided that pre-specified primary end points are met (for example, in diagnostic-positive patients and all patients), there is a clinically meaningful difference in the predicted benefit for diagnostic-positive patients compared to diagnostic-negative patients, and there are no observed detrimental effects in diagnostic-negative patients.

The relationship between the biomarker and the treatment effect (for example, the degree of benefit that the treatment provides based on the resultant biomarker assay) should be described both graphically and analytically in the label, and maximally informative statements should be provided to assist patients and health-care providers in their prescription choices. In instances where the biomarker is naturally bimodal (such as the *KRAS* mutation), the difference in the benefit could be described as the difference between the estimates of treatment effect in the diagnostic-positive and diagnostic-negative patient subsets with confidence intervals of the benefit in the diagnostic-positive and diagnostic-negative patients. When the biomarker is continuous or ordinal (for example, when gene expression of one or multiple genes is used to determine a score that can have different values, or when there are different levels of immunohistochemistry staining), parametric or non-parametric statistical approaches could be used (for example, parametric modelling of the biomarker–drug effect relationship or a non-parametric estimation of the drug effect at different percentiles of the biomarker distribution), but the planned approach needs to be specified in advance in the statistical analysis plan[26].

Patient selection using Phase III data

In this section, which applies only to labels that include information about the co-developed diagnostic, we discuss two situations in which a sponsor may need to define or redefine the diagnostically selected patient population in the label after analysis of the primary clinical trial has been completed. One situation is when there is a need to readjust the biomarker threshold that was prospectively specified and used in the clinical trial. The other situation is when the Phase III data will be evaluated for a biomarker that was not prospectively specified in the protocol. In both instances, we outline the motivation and need for redefining the diagnostic-positive patient population and highlight the approaches that are necessary to ensure an unbiased effect estimate of the benefit and uncompromised control of the type I error.

Threshold readjustment of the existing biomarker. Based on the results of a diagnostic assay, health-care providers and payers need to be able to determine whether or not a patient will benefit from a particular treatment. This determination will be more complex for those biomarkers that have a continuous relationship with the expected benefit (see FIG. 2; HER2 overexpression as predictive of breast cancer response to trastuzumab) as opposed to a bimodal relationship (echinoderm microtubule-associated protein-like 4 (EML4)–ALK fluorescence *in situ* hybridization assay as predictive of lung cancer response to crizotinib). In the former case, the process is more complex, whereas for the latter the process is relatively less complicated. When readjusting the threshold, one must also consider whether the diagnostic is being used as a positive or negative selection test. The bimodal expression of 6-*O*-methylguanine DNA methyltransferase (MGMT) in glioblastoma can be used as a positive selection test (low expression predicts a good response to temozolomide) or as a negative test (high expression predicts a poor response to temozolomide) to identify a population of unmet medical need in which to develop a new therapy[27]. To maximize benefit risk assessments in these settings, different cut-offs would probably be selected.

Although the scientific rationale, preclinical data and epidemiological data may be suggestive of a threshold for companion diagnostics that measure continuous biomarkers, the relevant clinical data that are required to establish a reasonable threshold are typically very limited prior to the initiation or analysis of the first pivotal clinical trial. For example, for first-in-class drugs in oncology, the completion of the pivotal clinical trial is typically the first time that a substantial amount of clinical data becomes available. Although basing the final labelling of the drug on a threshold that was selected on the basis of limited data from a Phase II trial may keep the analysis simple, this will not be in the best interest of patients. So, a sponsor — in collaboration with the regulators — may need to propose refining the diagnostic subset of patients indicated in the label by adjusting the biomarker threshold based on the availability of the Phase III data. Some examples of when and how to evaluate alternative thresholds to the one that has been pre-specified are shown in TABLE 4.

Proposal for planned threshold readjustment. When readjusting the threshold or conducting a Phase III clinical trial without a pre-specified threshold, the protocol should prospectively specify an analysis plan to ensure that the study-wise type I error is preserved and that the size of the treatment effect in the diagnostic-selected population is estimated in an unbiased way[26]. In this article, 'unblinding' of the study refers to the time when the primary analysis of the study is conducted and patient treatment assignments are known. Not every registration study will be blinded, but the guidance is general and the primary analysis will be the first opportunity for the sponsor to compare the biomarker data with the data on the efficacy of the drug and patient treatment outcomes.

There are several statistically sound ways of ensuring that the study-wise type I error (for example, two-sided α of 0.05) is preserved, including hierarchical approaches and split-α approaches for testing biomarker hypotheses[28–31]. Moreover, pre-specifying a biomarker threshold prior to the initiation of the Phase III trial may not be necessary,
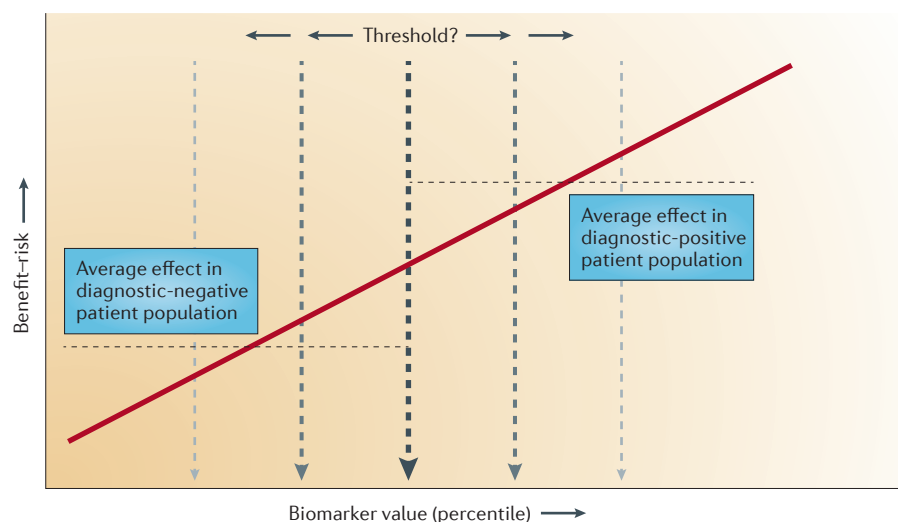
Figure 2 | **The relationship of biomarker prevalence versus expected treatment.** This hypothetical graph demonstrates the continuous relationship between the biomarker and the expected benefit of the drug, normalized to risk. Moving the diagnostic threshold (indicated by the dotted line) will affect both the average benefit of the treatment and the size of the patient population receiving it. For example, moving the threshold to the right will result in the diagnostic-selected population showing higher average benefit but including fewer patients, whereas moving the threshold to the left would increase the size of the patient population receiving therapy but show lower average benefit.

as valid statistical algorithms exist that use data from all patients in Phase III trials to first conclude whether the treatment is active in some or all patients and then to identify the patients who may not benefit meaningfully[32].

The use of an independent data set is one method of testing the updated threshold. If two clinical trials are being conducted close in time, the first trial could be used to redefine the threshold value, whereas the corresponding effect size would be estimated based on the data set from the second clinical trial using that threshold. The planned primary analysis of the second data set could be conducted with either the original threshold value or a redefined threshold, provided this has been specified prior to the primary analysis of the second clinical trial.

Even when an independent test data set is used to test the threshold, as described above, it is always helpful to pre-specify the procedure that will be used to adjust the threshold — for example, the threshold that corresponds to the largest subset of diagnostic-selected patients with a clinically meaningful benefit. Several empirical or model-based approaches could be used, including pre-specifying specific threshold quartiles (for example, the 25th, 50th and 75th percentiles of biomarker distribution),

to test for a continuous biomarker (for example, the expression of a biomarker that is quantified by RT-PCR) or to test for ordered categories for an ordinal biomarker (for example, the expression of a biomarker that is quantified by immunohistochemistry staining).

A split data set is one approach that can be used to obtain an unbiased estimate of the effect. If an independent data set is not available, great care needs to be exercised during statistical modelling to avoid overfitting (developing a statistical model that describes random error or noise instead of the underlying biological relationship of the data) and the bias that results from this[33]. A careful statistical analysis plan for potential threshold adjustment needs to be fully outlined prospectively prior to the unblinding of the study. Many approaches with well-understood properties have been developed for this purpose, such as bootstrap aggregating and adaptive designs for the prospective development of gene-expression-based classifiers[34–36]. One application of such an approach is outlined in Supplementary information S1 (box). In this approach, data obtained from patients in Phase III trials are re-sampled to obtain an unbiased estimate of an underlying distribution of the threshold and corresponding effect size.

The operational characteristics of any threshold adjustment procedure need to be carefully assessed using simulations. Special care needs to be taken to ensure that there is an unbiased estimate of treatment effect and correct type I error control when testing treatment effects overall and in biomarker-specified subsets.

Timing of threshold pre-specification. It may not always be possible to pre-specify a meaningful threshold on which the primary analysis can be based prior to the initiation of the registration study. For example, a pivotal trial in a given indication could be initiated contemporaneously with the proof-of-concept trial for the drug and the diagnostic. Alternatively, the final commercial diagnostic test may not be available for the start of the Phase III trial. In these cases, it will not be possible to prospectively stratify patients based on biomarker expression prior to randomization. Generally, the lack of prospective stratification using the biomarker does not have a meaningful effect on the conclusions of the trial, and its potential impact can be addressed via well-described sensitivity analyses[37,38].

However, if a diagnostic is being used to determine whether a patient is eligible for enrolment into the trial, then the analytically validated diagnostic test for which the sponsor is seeking approval as a companion diagnostic should be available at the start of the Phase III clinical trial. In some cases, substantial modifications will be made during the clinical trial to the diagnostic assay that is being used to determine patient eligibility, or it will be known prior to initiating the trial that the assay used in the clinical trial is not going to be the final assay to be marketed. In these situations, bridging studies between the clinical trial assay used at enrolment and the final diagnostic assay upon which the primary analysis is based will be needed to obtain approval.

New biomarker specification after trial initiation or completion. New molecularly defined disease subtypes might emerge — but they will not be well established — prior to the primary analysis of the trial data. Alternatively, it may not be possible to use a putative biomarker owing to the lack of an appropriate assay prior to the start of the trial or prior to trial completion. However, the statistical validity of the clinical trial is not impaired by the delayed specification of the biomarker if the diagnostic-positive patient population is defined prior to the primary analysis of the data from the Phase III trial.

Table 4 | Factors involved in diagnostic threshold readjustment*

Reason	Readjustment proposal	Conditions required for proposal	Effect as compared with using the original threshold per statistical plan
New treatment options have emerged while the trial was underway that result in a change in how the disease is molecularly classified and treated; as a result, higher efficacy may be needed for approval	Readjust the threshold to ensure that the diagnostic-selected population has a clinically meaningful benefit based on currently available knowledge	Meeting the primary end point in the pre-specified diagnostic subset (using original threshold)	Higher estimated expected benefit in indicated diagnostic-positive patients; lower prevalence of the selected population
There is a statistically significant benefit in the patient population selected with the pre-defined threshold, but the estimate of the benefit is lower than was expected			
There is an observed benefit in the patient population selected with the predefined threshold as well as in the subset of patients with diagnostic values that were not captured by the predefined threshold	Adjust the threshold to be more inclusive of biomarker values to include all patients with relevant clinical benefit	Meeting the primary end point in the pre-specified diagnostic subset (using original threshold), as well as in the diagnostic-negative patients or all patients, as pre-specified in the analysis plan	Lower estimated expected benefit in indicated diagnostic-positive patients; higher prevalence of the selected population

*Examples are given here of the reasons why the diagnostic threshold might need readjustment, the resulting proposal, required conditions for proposal implementation and the effect on the label.

The analysis should be based on information that was collected externally to the Phase III trial, and the protocol should specify in detail how the single to-be-determined biomarker will be used in the primary analysis.

In some cases, a prospective–retrospective analysis of a biomarker that was identified following the primary analysis of the Phase III trial[39] will be needed using data obtained from the registration trial. Considerations for these situations have been discussed in the literature (see REFS 32–34,40–42). The recent approval of the Qiagen TheraScreen KRAS mutation test as a companion diagnostic for Erbitux (cetuximab) is the first case of a companion diagnostic for an approved oncology drug that relied on a prospective–retrospective analysis and serves as an effective case study for the applicability of this approach. However, it is unknown how a putative biomarker can be clinically validated when the original clinical trial was negative. In this case, it may be necessary to conduct a new clinical trial as it may be unlikely that approval for a drug or diagnostic would be granted on the basis of subset analysis from a failed study.

One option is to plan for a pre-specified but not yet actionable biomarker by prospectively specifying that part of the type I error will be reserved for a diagnostic subset that will be specified in the future (for example, present primary end points are tested at two-sided $\alpha = 0.04$, whereas $\alpha = 0.01$ will be left for future testing). In this proposal, many of the considerations described in REF. 34 will apply[42]. For example, archived

biopsy samples must be available for the majority of patients in the trial, an analytically validated assay must be used and a focused analysis plan must be prospectively defined that describes how the single new biomarker will be analysed using the Phase III trial data.

It is unlikely that a sufficient portion of the 0.05 type I error will be reserved for the evaluation of more than one yet-to-be-determined biomarker. In addition, when a hypothesis about the new biomarker is tested on the data set from the pivotal Phase III trial, there must be strict assurances that no information about the new hypothesis has previously been made available during the generation of the initial pivotal Phase III trial data and that this is the only hypothesis being tested.

Multi-marker diagnostic development
Until now, we have been discussing the use of a single biomarker. However, several biomarkers may be needed to identify the patients who are likely to benefit from a therapy. In the straightforward case of mutations in oncogenes and tumour suppressor genes, the diagnostic would need to capture both common and less common biologically equivalent mutations to ensure that the broadest group of patients who may benefit from a therapy is identified using only one diagnostic platform. More generally, aberrant activation of a growth factor signalling pathway may be due to four factors (FIG. 3): first, overexpression of the growth factor; second, the presence of activating mutations in the growth factor receptor or oncogenes that

lie downstream; third, a loss-of-function mutation in downstream tumour suppressor genes; and fourth, activation by a compensatory pathway. A biomarker that accurately identifies tumours with the associated pathway alteration would qualify as a potential predictive diagnostic; however, in order to test for all of these factors associated with pathway activation, it may be necessary to use more than one diagnostic assay. Alternatively, a downstream readout of pathway activation or a gene signature, as measured by either multiplex RT-PCR or an enzyme-linked immunosorbent assay (ELISA), could be used.

Clinical validation of the summary measures. The use of multiple biomarkers will make the development and qualification of companion diagnostics even more complex. If multiple biomarkers are needed to identify patients who may benefit from a drug, then the composite biomarker — for example, a combination of gene expression values — needs to be evaluated rather than the individual components. Moreover, it is the composite biomarker that needs to be validated as 'fit for intended use' or 'fit for purpose' based on the Phase III trial.

Provided there is appropriate preclinical evidence and a scientific rationale, the clinical validation of each biomarker separately should not be required. This is applicable to both single- (for example, an RT-PCR panel) and multi-platform diagnostics (for example, immunohistochemistry staining and mutations), but different practical considerations may apply.
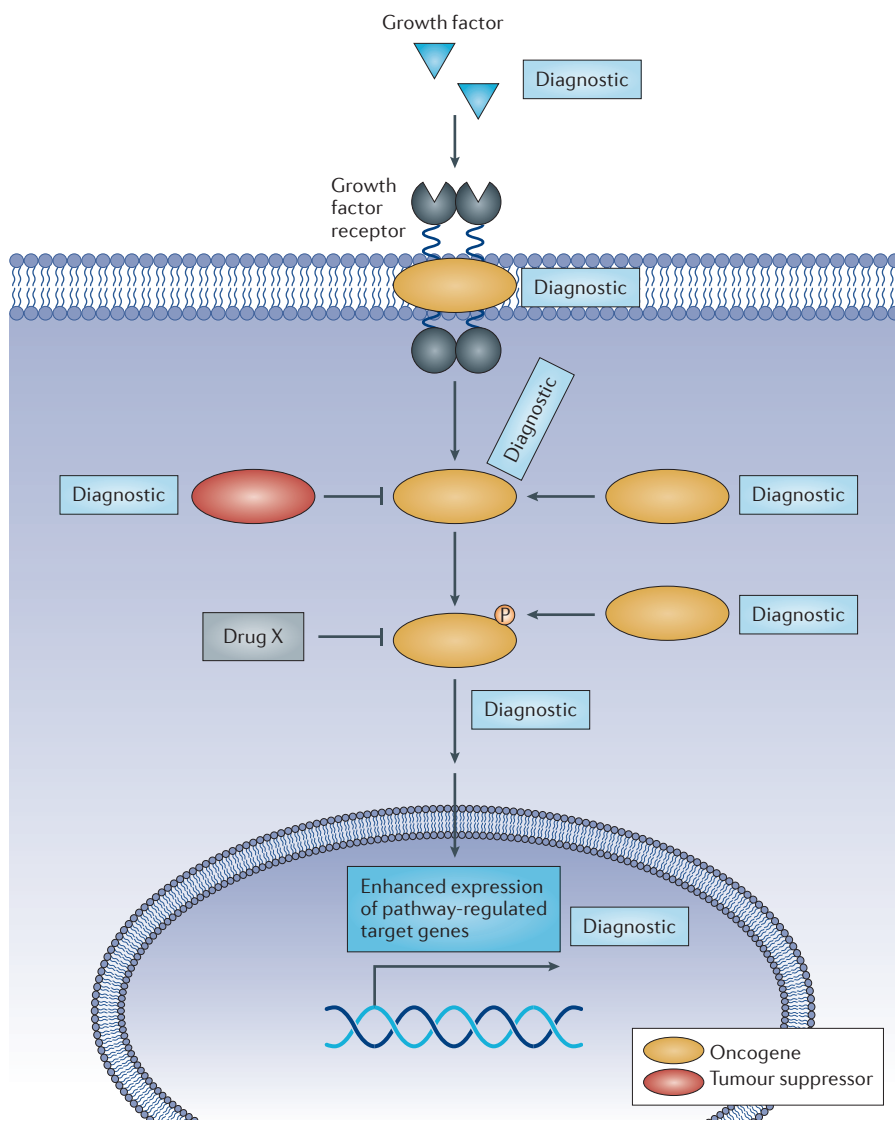
Figure 3 | Identification of a biomarker in tumours with an associated pathway alteration. Aberrant activation of a growth factor signalling pathway may be due to activating mutations in the growth factor or downstream oncogenes or due to a loss-of-function mutation or mutations in downstream tumour suppressor genes. In such a pathway, a biomarker that accurately identifies tumours with the associated pathway alteration would qualify as a potential predictive diagnostic for a drug that is downstream of activation.

Adaptive definition of the diagnostic-selected population. In 2011, a panel of representatives from the FDA, the US National Institutes of Health and the pharmaceutical industry recommended that the FDA should approve the adaptive signature approach for the identification of diagnostic-selected patient populations[38]. In this approach, a training set of data from the Phase III study is used to define the diagnostic-selected patient population, and data from the remaining patients are used for evaluating the effect of the treatment in that subset.

Redefining the threshold of the existing summary measure. Similar to when a single biomarker is used, one could consider pre-specifying the refinement possibilities of the threshold summary measure. Identical considerations would apply as the summary measure is univariate, but care will be needed to ensure that an unbiased estimate of the effect is obtained if the threshold is to be adjusted.

DNA sequencing and diagnostics
Although FDA approval of companion diagnostics has been limited to 'one drug–one diagnostic' test pairs, technological

advancements may soon alter that approach. The diagnostic work-up of patients who are newly diagnosed with cancer is quickly advancing towards testing for somatic mutations in a large panel of genes. Next-generation sequencing and screening platforms are being used to characterize tumours and determine the best courses of treatment for patients with specific mutations. Multi-biomarker screening (which may be distinct from the type of multi-marker screening discussed above in the section entitled "Multi-marker diagnostic development") has several advantages over the current 'single-test, single-drug' paradigm. Below, we discuss some suggestions that are relevant to the use of screening multiple mutations, such as with next-generation sequencing, which are in clinical trials for FDA approval[43].

Potential path to the FDA approval of a multi-marker screening platform. As mentioned earlier, a diagnostic that is used to support the registration of a drug in a diagnostically defined subpopulation of patients must be analytically validated. No platform screening technology system has yet been approved by the FDA for a specific drug, and data using such technology have never been submitted to the US Center for Devices and Radiological Health (CDRH) for drug approval.

One possible way to obtain approval of a screening platform as a companion diagnostic could be to obtain an investigational device exemption for the entire platform prior to the initiation of clinical testing. The presence of a particular predictive biomarker in the platform could enable the entire platform to be considered as adequate for selecting patients for treatment with a new compound that was targeted against the biomarker. If the new drug demonstrated acceptable clinical benefit and safety in the patient population that had been selected by the specific biomarker from the platform, this would demonstrate that the biomarker had clinical validity and clinical utility, and the biomarker could be reviewed and given FDA clearance for diagnostic use (together with approval for the drug).

Analytical validation of classes of mutations. Often, the group of patients who are predicted to benefit from a molecularly targeted drug is defined as those patients who have tumours that contain any of a pre-specified class of DNA alterations. If the results of next-generation sequencing are used to determine whether a patient has a specific class of DNA alteration, then the

next-generation sequencing platform should be analytically validated for assessing the specific set of DNA variants that are included in the definition of the diagnostic for appropriate usage. In these situations, the platform only needs to be validated for the specific alteration or alterations being evaluated.

Although it should not normally be necessary to individually validate each DNA variant or alteration, appropriate methods of determining the equivalence of DNA alterations within a class of mutations will have to be decided by the FDA. Rigorous preclinical studies might be used for mutation assessment and class assignment, if they are deemed acceptable by the agency. The drug company might also be required by the FDA to make a post-approval marketing commitment to work with device manufacturers and health-care payers and providers to generate a registry of mutation data that allows the duration of therapy to act as a readily obtainable surrogate outcome for determining clinical benefit. Analytical validation studies should be sufficiently extensive to characterize the sensitivity and specificity of representative alterations and to characterize the kinds of alterations and the genomic contexts (such as short deletions in homopolymer regions) that are problematic.

Conclusions and future considerations

In this article we have discussed some of the most important challenges that sponsors and regulators currently face during the companion diagnostic development process. We have focused on the restriction of diagnostic-selected patient populations, on threshold adjustment for diagnostic selection, and on the use of multi-marker assays and next-generation sequencing in biomarker screening. We hope that this article serves as a guide to potential approaches for dealing with these challenges.

Following the publication of the FDA's co-diagnostic draft guidance[9], stakeholders provided feedback and suggested issues that may arise during the co-development process and that were not addressed sufficiently in the July 2011 guidance document. Owing to the rapidly evolving nature of genomics and molecular diagnostics, there are many remaining issues that are not discussed in this article. These include the emergence of new scientific evidence following the approval of a therapy and companion diagnostic; and defining a patient population or indication using a biomarker rather than by classical histological tumour classification.

Glossary

Bootstrap aggregating
A type of model averaging that improves the stability and accuracy of the algorithms that are used in biomarker studies, and is typically conducted by repeatedly re-sampling data points from a given data set.

Bridging studies
Studies in which clinical samples that were tested initially with an assay used in a clinical trial are re-tested with another assay to support the approval or clearance of that assay.

Classifier
An algorithm (or statistical rule) that can be used to predict prognosis or the responsiveness of patients to a given therapy and thereby used to select and/or stratify patients for therapy in clinical trials. The inputs to the algorithm are the values obtained from one or more predefined biomarkers.

Diagnostic platform
A form of molecular diagnostic testing that provides patient-specific information using parallelized platform sequencing technology.

Effect size
An estimate of the treatment effect relative to the control (or any other parameter of interest).

Hierarchical approaches
Sequential approaches to the testing of multiple hypotheses where a given null hypothesis can only be tested if all null hypotheses that are ranked higher are rejected.

Investigational device exemption
A regulatory submission that allows a medical device to be used in a clinical study without full approval from the US Food and Drug Administration in order to collect data on the safety and effectiveness of the device.

Laboratory-developed tests
A class of *in vitro* diagnostic tests that are currently not regulated by the US Food and Drug Administration.

Next-generation sequencing
A set of technologies that enable the rapid generation of enormous amounts of DNA or RNA sequencing data.

Notified body
An organization that has been accredited by a member country of the European Union to determine whether a product meets certain predetermined standards.

Split-α approaches
Approaches that are undertaken for the testing of multiple hypotheses to maintain the study-wise type I error at the intended 0.05 level by splitting the threshold for declaring significance (that is, α) among the hypotheses to be tested.

Statistical analysis plan
The pre-specified analyses that will be applied to the data generated from a clinical trial.

Summary measures
The mathematical combination of values produced by one or more biomarkers, resulting in a single value that can be used for making decisions about drug treatments.

Training set
A data set that is used for the development of a statistical model and all of its parameters. Another data set known as the test set is then used to test the accuracy of the model.

Type I error
The chance of falsely rejecting the null hypothesis in favour of the alternative (the probability of the false positive); for example, falsely claiming that a relationship exists between treatment effect and biomarker value in the absence of a true relationship.

Unbiased effect estimate
An estimate of the treatment effect in which the expected value (based on hypothetical repetitions of the study) equals the true value of the effect.

Although this article addresses the questions that arose following the FDA approval of companion diagnostics as medical devices, it does not explicitly address the issue of the FDA approval of laboratory-developed tests and the use of these tests for off-label prescription. Within the drug development community, it is recommended that diagnostic assays or laboratory-developed tests conducted in Clinical Laboratory Improvement Amendments (CLIA)-accredited laboratories should be considered as companion diagnostics that are regulated by the FDA. In addition, some recent drug labels for targeted therapies (such as vemurafenib) have specified the use of an FDA-approved test but they do not specifically name the FDA-approved diagnostic assay.

However, many diagnostic companies find laboratory-developed tests fairly controversial as there are currently fewer validation standards for laboratory-developed tests than for FDA-approved diagnostic tests. Laboratory-developed tests continue to be developed and used for prescribing drugs, so another large challenge will be how to determine the equivalence of multiple different laboratory-developed tests and assays to measure the same biomarker. Many discussions among regulators, sponsors and researchers are currently underway to define the appropriate use of laboratory-developed tests.

As companion diagnostics emerge as a key part of personalized medicine, sponsors are discovering that some of the challenges in the companion diagnostic pathway relate to the different and sometimes disparate development processes and timelines required for therapeutics and diagnostics. Furthermore, there are cultural differences between the two industries[44], especially

regarding the expectations around the characterization of precision. For instance, in therapeutics it must be shown that the investigational therapy is superior to the standard therapy, and estimates of any given measure of clinical benefit — within a fairly broad range — are acceptable. In diagnostics, there is an expectation of a very high degree of precision and near-perfection of the operating characteristics of the assay (even though such tests are ultimately applied to therapies that have an unpredictable benefit in the biomarker-defined population). In addition to increased communication between the FDA and the sponsors, communication and flexibility will be of utmost importance for building the relationship between companies developing therapeutics and those developing diagnostics (and, increasingly, also within the same company). Just as increased communication between sponsors and the FDA has proved to be fruitful, early and increased interactions between companies developing therapeutics and those developing diagnostics will be beneficial. Indeed, some companies, including Roche, have embedded members of the diagnostics team within the pharmaceutical development team in order to improve project planning and the development of companion diagnostics[44].

Additional guidance from the FDA will also be beneficial, especially as new technologies rapidly advance and are introduced into clinical practice. For example, recommendations on the use of bridging assays or updating components of existing assays would be helpful, as will clarification of the optimal processes that are required for the clinical validation of multi-marker tests.

Jane Fridlyand is at Genentech, 1 DNA Way, MS 441b, South San Francisco, California 94080, USA.

Richard M. Simon is at the US National Cancer Institute, 9000 Rockville Pike, Bethesda, Maryland 20892-7434, USA.

Jessica C. Walrath, Jeff D. Allen and Ellen V. Sigal are at Friends of Cancer Research, 1800 M Street NW, Suite 1050, South Washington, DC 20036, USA.

Nancy Roach is at Fight Colorectal Cancer, 1414 Prince Street, Suite 204, Alexandria, Virginia 22314, USA.

Richard Buller is at Pfizer, 10555 Science Center Drive, San Diego, California 92121, USA.

David P. Schenkein is at Agios Pharmaceuticals, 38 Sidney Street, Suite 2, Cambridge, Massachusetts 02139, USA.

Keith T. Flaherty is at Massachusetts General Hospital, 55 Fruit Street, Yawkey 9E, Boston, Massachusetts 02114, USA.

Howard I. Scher is at Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

Correspondence to J.C.W.
e-mail: jwalrath@focr.org

1. Naylor, S. & Cole, T. Overview of companion diagnostics in the pharmaceutical industry. *Drug Discovery World* [online], http://www.ddw-online.com/personalised-medicine/p92845-overview-of-companion-diagnostics-in-the-pharmaceutical-industry.spring-10.html (2010).
2. US National Cancer Institute. Targeted cancer therapies (Fact Sheet). *US National Cancer Institute* [online], http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted (2012).
3. Savage, D. G. & Antman, K. H. Imatinib mesylate — a new oral targeted therapy. *N. Engl. J. Med.* **346**, 683–693 (2002).
4. Demetri, G. D. *et al.* Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N. Engl. J. Med.* **347**, 472–480 (2002).
5. Garnock-Jones, K. P., Keating, G. M. & Scott, L. J. Trastuzumab: a review of its use as adjuvant treatment in human epidermal growth factor receptor 2 (HER2)-positive early breast cancer. *Drugs* **70**, 215–239 (2010).
6. Hudziak, R. M. *et al.* p185HER2 monoclonal antibody has antiproliferative effects *in vitro* and sensitizes human breast tumor cells to tumor necrosis factor. *Mol. Cell. Biol.* **9**, 1165–1172 (1989).
7. Kulke, M. H. Systemic therapy for advanced pancreatic neuroendocrine tumors. *Semin. Oncol.* **40**, 75–83 (2013).
8. Motzer, R. J. *et al.* Sunitinib in patients with metastatic renal cell carcinoma. *JAMA* **295**, 2516–2524 (2006).
9. US Food & Drug Administration. Guidance for industry and Food and Drug Administration staff — *in vitro* companion diagnostic devices (draft guidance). *US Food & Drug Administration* [online], http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm262292.htm (2011).
10. Teutsch, S. M. *et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet. Med.* **11**, 3–14 (2009).
11. US Food & Drug Administration. Guidance for industry: enrichment strategies for clinical trials to support approval of human drugs and biological products (draft duidance). *US Food & Drug Administration* [online], http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM332181.pdf (2012).
12. European Medicines Agency. Reflection paper on co-development of pharmacogenomic biomarkers and assays in the context of drug development. *European Medicines Agency* [online], http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/07/WC500094445.pdf (2010).
13. US Food & Drug Administration. Drug-diagnostic co-development concept paper (draft — not for implementation). *US Food & Drug Administration* [online], http://www.fda.gov/downloads/drugs/scienceresearch/researchareas/pharmacogenetics/ucm116689.pdf (2005).
14. Moore, M. W., Babu, D. & Cotter, P. D. Challenges in the codevelopment of companion diagnostics. *Per. Med.* **9**, 485–496 (2012).
15. Beckman, R. A., Clark, J. & Chen, C. Integrating predictive biomarkers and classifiers into oncology clinical development programmes. *Nature Rev. Drug Discov.* **10**, 735–748 (2011).
16. Ravnan, M. C. & Matalka, M. S. Vemurafenib in patients with *BRAF V600E* mutation-positive advanced melanoma. *Clin. Ther.* **34**, 1474–1486 (2012).
17. Tanizaki, J. *et al.* Activation of HER family signaling as a mechanism of acquired resistance to ALK inhibitors in EML4-ALK-positive non-small cell lung cancer. *Clin. Cancer Res.* **18**, 6219–6226 (2012).
18. Brand, T. M., Iida, M. & Wheeler, D. L. Molecular mechanisms of resistance to the EGFR monoclonal antibody cetuximab. *Cancer Biol. Ther.* **11**, 777–792 (2011).
19. Chung, K. Y. *et al.* Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. *J. Clin. Oncol.* **23**, 1803–1810 (2005).
20. Fridlyand, J. *et al.* An industry statistician's perspective on PHC drug development. *Contemp. Clin. Trials* http://dx.doi.org/10.1016/j.cct.2013.04.006 (2013).
21. Freidlin, B., McShane, L. M., Polley, M. Y. & Korn, E. L. Randomized phase II trial designs with biomarkers. *J. Clin. Oncol.* **30**, 3304–3309 (2012).
22. Kaiser, L., Becker, C., Kukreti, S. & Fine, B. Decision making for a companion diagnostic in an oncology clinical development program. *Drug Inform. J.* **46**, 294–302 (2012).
23. Mandrekar, S. J. & Sargent, D. J. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J. Clin. Oncol.* **27**, 4027–4034 (2009).
24. Mandrekar, S. J. & Sargent, D. J. All-comers versus enrichment design strategy in phase II trials. *J. Thorac. Oncol.* **6**, 658–660 (2011).
25. Simon, R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per. Med.* **7**, 33–47 (2010).
26. Zhan, S. *et al.* Defective neuropeptide processing and ischemic brain injury: a study on proprotein convertase 2 and its substrate neuropeptide in ischemic brains. *J. Cereb. Blood Flow Metab.* **29**, 698–706 (2009).
27. Spiegl-Kreinecker, S. *et al.* O6-methylguanine DNA methyltransferase protein expression in tumor cells predicts outcome of temozolomide therapy in glioblastoma patients. *Neuro Oncol.* **12**, 28–36 (2010).
28. Brannath, W. *et al.* Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat. Med.* **28**, 1445–1463 (2009).
29. Chen, C. & Beckman, R. A. Hypothesis testing in a confirmatory phase III trial with a possible subset effect. *Stat. Biopharm. Res.* **1**, 431–440 (2009).
30. Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–802 (1988).
31. Wang, S. J., O'Neill, R. T. & Hung, H. M. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm. Stat.* **6**, 227–244 (2007).
32. Jiang, W., Freidlin, B. & Simon, R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl Cancer Inst.* **99**, 1036–1043 (2007).
33. Christine, M. M., Sharly, J. N. & Gilbert, S. O. (eds) *Evolution of Translational Omics: Lessons Learned and the Path Forward* (The National Academies Press, 2012).
34. Breiman, L. Bagging predictors. *Machine Learn.* **24**, 123–140 (1996).
35. Freidlin, B., Jiang, W. & Simon, R. The cross-validated adaptive signature design. *Clin. Cancer Res.* **16**, 691–698 (2010).
36. Freidlin, B. & Simon, R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin. Cancer Res.* **11**, 7872–7878 (2005).
37. Kaiser, L. D. Stratification of randomization is not required for a pre-specified subgroup analysis. *Pharm. Stat.* **12**, 43–47 (2013).
38. Scher, H. I., Nasso, S. F., Rubin, E. H. & Simon, R. Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clin. Cancer Res.* **17**, 6634–6640 (2011).
39. Simon, R. M., Paik, S. & Hayes, D. F. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J. Natl Cancer Inst.* **101**, 1446–1452 (2009).
40. European Medicines Agency. Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection. *European Medicines Agency* [online], http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf (2011).

41. Goldberg, P. KRAS finding changes oncology practice but poses profound regulatory dilemma. *The Cancer Letter* **35**, 1–8 (2009).

42. Patterson, S. D. *et al.* Prospective-retrospective biomarker analysis for regulatory consideration: white paper from the industry pharmacogenomics working group. *Pharmacogenomics* **12**, 939–951 (2011).

43. Simon, R. & Roychowdhury, S. Implementing personalized cancer genomics in clinical trials. *Nature Rev. Drug Discov.* **12**, 358–369 (2013).

44. Park, R. Companion diagnostics: a blossoming field. *IVD Technology* [online], http://www.ivdtechnology.com/article/companion-diagnostics-blossoming-field (2011).

45. Gaffney, A. Device group launches offensive against pending EU legislation. *Regulatory Focus* [online], http://www.raps.org/focus-online/news/news-article-view/article/2977/device-group-launches-offensive-against-pending-eu-legislation.aspx (2013).

46. European Commission. *In vitro* diagnostic medical devices: Directive 98/79/EC. *European Commission* [online], http://ec.europa.eu/enterprise/policies/european-standards/harmonised-standards/iv-diagnostic-medical-devices/ (1998).

47. Stynen, D. Revision of Europe's IVD Directive 98/79/EC. *IVD Technology* [online], http://www.ivdtechnology.com/article/revision-europe's-ivd-directive-9879ec (2011).

48. European Commission. Revision of the medical device directives. *European Commission* [online], http://ec.europa.eu/health/medical-devices/documents/revision (2012).

49. Study Group 1 of the Global Harmonization Task Force. *Principles of In Vitro Diagnostic (IVD) Medical Devices Classification* [online], http://www.imdrf.org/docs/ghtf/final/sg1/procedural-docs/ghtf-sg1-n045-2008-principles-ivd-medical-devices-classification-080219.pdf (Global Harmonization Task Force, 2008).

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

SUPPLEMENTARY INFORMATION
See online article: S1 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF