EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

9 June 2011
EMA/446337/2011
Committee for Medicinal Products for Human Use (CHMP)

Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection

Draft

Draft Agreed by Pharmacogenomics Working Party (PGWP)	**March 2011**
Adoption by CHMP for release for consultation	9 June 2011
End of consultation (deadline for comments)	25 November 2011

Comments should be provided using this template.The completed comments form should be sent to PGWPSecretariat@ema.europa.eu

Keywords	***Clinical trial designs, Enriched design, Genomic biomarkers, (GBMS), hybrid design, Predictive markers, Pharmacogenomics, Retrospective data analyses***

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu
An agency of the European Union

Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection

Table of contents

1. Introduction

The availability of techniques that facilitate study of the human genome has led to an exponential increase in investigation into genomic biomarkers (GBMs) for diagnosis of specific diseases, as a marker of response to treatment or of prognosis. Theoretically genomic BMs should offer the advantage of improved specificity and reduction of heterogeneity that is an integral part of phenotypic population grouping. This is very attractive in drug development because of their potential ability to reduce drug attrition and to reduce overall developmental costs, that are achieved through improved understanding of the mechanism of drug action, predict adverse events to individual drugs or as a group effect (e.g. CYP poor metabolisers), and use of novel development strategies in pre-clinical and clinical phases.

In clinical drug development, GBMs may aid and influence a wide range of areas: patient selection, stratification of treatment strategies or patient groups, early evaluation of treatment effect including adverse reactions, and prognosis. There is opportunity for the GBMs to be used for pre-defined subgroup analysis or to enable novel trial designs that might not be possible otherwise due to heterogeneity of clinical characteristics.[1,2] GBMs could also play a valuable role in the risk management strategies including risk minimisation by aiding *a priori* identification of patients susceptible to develop severe adverse effects (e.g. HLA B*- 5701 and use of Abacavir).

While a number of these aspects are discussed in many publications in the recent years, specific aspects relating to drug development and discussion on regulatory considerations have lagged behind. The intention of this paper is therefore to provide an evidence based consideration of GBM related issues from a regulatory viewpoint. Mention is also made of co-development of a GBM diagnostic test for use with a medicinal product.

The principles established in the reflection paper are based on the experiences gained from the evaluation of dossiers within the EU regulatory processes —including marketing authorisation applications reviewed by CHMP, the scientific advice documents and additionally, the voluntary genomic data submission meetings (briefing meetings) at the Pharmacogenomic Working party (PGWP) over the last several years. It is expected that these principles guide both industry and the assessors in the evaluation of such biomarkers in relation to the qualification process in the context of clinical development (BM qualification in EU) and the assessment of benefit: risk balance of medicinal products or selection of the relevant target population. The paper should also be read in conjunction with other relevant guidelines listed at the end of the documents under section “Other aspects”.

Development of GBMs and diagnostic tests may involve additional development of tests (companion diagnostics) or specific kits (platforms) to detect for the presence or absence of the GBM. Issues relating to these are outside the scope of this paper but a short discussion is included. The readers are referred to appropriate guidelines/ papers for details (see section on other aspects).

2. Scope and objectives

The objective of this reflection paper is to highlight key principles that should be considered by stakeholders[3] with focus on use of GBM in relation to patient selection and associated issues with trial methodology. Some of the controversial issues are highlighted. The principles are considered applicable to the development and validation of a GBM through the life cycle of a medicinal product, i.e., pre-authorisation and post-marketing stages. The main discussion will be related to drug

[1] EPAR_WC500049823.pdf
[2] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2383915/pdf/1477-7800-5-9.pdf

[3] Stakeholders include parties involved in biomarkers and drug development such as Pharma Industry, public-private partnerships, academia, patients and health care professionals.

development and use of the GBMS that predict drug response but many principles are applicable to GBMs that relate to prognosis as well. The document aims to highlight the main considerations related to use of GBMs based on the experiences of CHMP.

It is recognised that some of these principles may apply to non-genomic BMs in the context of drug development but will not be discussed here. Similarly, surrogate biomarkers (GBMs) are not discussed in this paper.

3. Features of genomic biomarkers (GBMs)

3.1. Classification of GBMs

While GBMS may be used to indicate many facets of a disease, two important roles are identified. In the context of this paper, the GBMs of interest are those that provide clues towards response (safety or efficacy or metabolic) to a particular therapeutic intervention, especially drug therapy (Predictive markers) or those that indicate disease prognosis (Prognostic Markers) that may not have an intrinsic relation to specific intervention, either drug therapy or otherwise. Some markers may play both roles.

Surrogate (pharmacogenomic) markers for clinical outcome are not addressed in this document as stated above.

There are situations where knowledge relating to a GBM might evolve both in its role as a single marker or part of a multimarker signature. Handling of such these are situation dependent and considered currently outside the scope of this paper. This also applies to increasing knowledge of the test. In both the above cases, regulatory decisions/ opinions will be based on available and advances in scientific knowledge.

3.1.1. Predictive GBMs

For the purposes of drug development, predictive GBMs occupy the highest area of interest. These should be pre-treatment characteristics that enable to determine whether a particular subject is a good candidate for treatment with a test agent. Commonly these tend to be binary or depend on classifiers (see section 3.2). Of note, these GBMs, in their simplest form could be a gene or point mutation. Alternatively, they could be based on expression levels of many genes where expression profiles of these genes are combined and evaluated in a predefined fashion. If the relationship between different genes or their expression levels are not predefined, but cut off points are generated using ROCs from one trial, then confirmation in a second trial would be expected. Evaluation of clinical utility of such predictive markers is facilitated by pivotal trials conducted in defined patient populations, selected and grouped based on the marker(s).

3.1.2. Prognostic GBMs

Prognostic GBMs (or markers) are those that correlate with outcome of disease in either untreated or heterogeneously treated patients. Development and evaluation of such GBMs are often based on a convenience sample of patients or subjects based on the availability of biological sample for assay of the GBM (blood or tissue). Thus prognostic BMs may or may not provide the basis for a clinical decision or influence the decision algorithm for treatment or intervention. However, studies evaluating prognostic GBMs may provide a scientific background of the natural history of the disease, facilitate development of additional other biomarkers (genomic or non-genomic) and contribute to drug development indirectly.

3.2. Selection of GBMs

Predictive GBMs may be indicators of efficacy (e.g. EGFR mutation status and use of geftinib) or safety (e.g., *HLA B* 5701* and abacavir hypersensitivity). This distinction may blur in certain situation and the data may provide opportunities for alternative interpretations. For example, the role of panitumumab (Vectibix) monotherapy in the third line indication in metastatic colorectal carcinoma is liable to interpretation as an efficacy marker while the combination with FOLFOX chemotherapy in the 2nd line indication suggests that mutant KRAS status may serve as a safety marker (potential for harm with the use of Vectibix + FOLFOX in those with mutated KRAS). GBMs may also serve as molecular targets for drug therapy (Her-2 receptor and trastuzumab). Therefore, the selection and evaluation of the GBM in any development programme (including design of the trials needed) will be dependent on the expected primary role of the GBM under consideration, the complexity of the relationship of the marker to the disease and, the mechanism of drug action. For example, while Her-2 receptor overexpression is an indicator of outcome in breast cancer[4], development of trastuzumab[5], a monoclonal antibody against HER-2 necessarily required modification of the trial designs that permitted evaluation of this intervention. It is important to consider that more than one marker may be linked to a particular disease and also influence the predictability of drug response either independently or simultaneously (e.g. ER and Her-2 in breast cancer[6], Her-2 and EGFR). Therefore, in exploratory studies, it is possible to evaluate a number of markers (or GBMs) among which one or more might eventually be selected for further evaluation depending on the situation, the drug in question and the mechanism or pathway of action. In such cases, the strength of association between each marker(s) and the relevant clinical endpoint will influence its subsequent development, clinical utility of the marker(s)[7] and the evidentiary standards needed to achieve clinical and regulatory adoption of the GBM. When a GBM or a panel of GBMs ("multimarker signatures/ gene signatures") are investigated within one or more exploratory studies, it is necessary to recognise that such studies are hypothesis generating and should include a set of classifiers[8] that translate the biomarker or the panel into a set of markers that predict clinical outcome.

Development and evaluation of multiple GBMs (as a simultaneous or sequential set) will present a different level of complexity than a single GBM, as each element (GBM) may have a different weight *vis a vis* the clinical impact of the overall panel. Warfarin genomics serve as an exemplar of this complexity with variable contributions from polymorphisms of *CYP2C19, VKORC1* and to a lesser extent *CYP4F2* gene or their different combinations. In cases with multiple GBMs or where a panel is evaluated, there is an inherent expectation that the relationships between the components of the panel are well established fairly early in the process[5] such as that late phase trials will provide confirmatory evidence. Ideally, the relative contribution of each GBM should be assessed independently and then of the combination as each marker may influence response to independent interventions or a complex interplay between markers and interventions is possible. The complex relation between HER2 and hormone receptors in breast cancer where response to hormonal treatment in ER+ patients is dependent upon simultaneous Her2+ receptor status[9] is one such multimarker example. Similarly, response to aromatase inhibitor (letrozole) was influenced by Her2/ EGFR *status*[10] in metastatic breast cancer.

---

4 Slamon DJ, et al *Science*. 1987; 235: 177-182.
5 Pegram M, Slamon D. *Semin Oncol.* 2000; 27 (suppl9): 13-19.
6 Daling JR et al: *Cancer.* 2001 Aug 15; 92(4):720-9.
7 Baulida J et al. *J Biol Chem* 1996, 271:5251-5257.
8 Simon R. *J Stat Plan Inference*. 2008 February 1; 138(2): 308–320 (A classifier is a mathematical function that translates biomarker value into a set of prognostic categories; it can also be defined as a marker that allows classification of patients.)
9 EPAR -/WC500049823.pdf
10 EPAR-PI-Tyverb-WC500044957.pdf

The considerations detailed above (classification and selection) will apply to both preauthorisation and post-authorisation studies with any medicinal product. Majority of the experiences with GBMs identified in the post authorisation period have been GBMs related to safety but this is not invariable. For example, HLA-alleles and hypersensitivity reactions to abacavir or carbamazepine are safety issues while the recent debate about tamoxifen and CYP2D6 polymorphisms relate nominally to efficacy(or lack of it). The timing of GBM generation is discussed subsequently.

3.3. Purpose of GBMs

3.3.1. Patient selection

In any drug development programme, among the many possible purposes of GBMs, the selection of patients is one of the commonest aims. The patient selection could be enhanced using a GBM in the following ways:

- *for better definition of the disease and/or its prognosis*: Identification of patients with a particular disease sub-type or disease severity as a target (e.g., Her-2 and breast cancer, or *Philadelphia chromosome* in chronic myeloid leukaemia).
- *for excluding patients at increased risk*: Identification of patients at increased risk of experiencing a serious adverse drug reactions for the purpose of excluding them from further clinical trials or treatment with that specific agent.(e.g., *HLA B* 5701* and abacavir use or carbamazepine and *HLA-B*1502*)
- *for prediction of drug response*: Identification of patients with high likelihood of experiencing benefit with a particular medicinal product with few or no safety issues/adverse events (trastuzumab in Breast cancer with Her-2 overexpression).

Whilst the above statements are usually applicable to particular medicinal products, it is possible that these are applicable and helpful in patient selection for combination of therapies or for sequential treatment algorithms (e.g., fields of oncology and HIV infection). These are discussed below.

3.3.2. Treatment algorithm allocation

GBMs may also be used for selection of treatment sequences whether in clinical trials or in clinical practice. In the context of clinical trials, treatment algorithms might be based on the presence of a single or a set of markers while maintaining the randomised comparison with standard of care (or placebo as appropriate). For example, in metastatic breast cancer trastuzumab could be used based on HER2 overexpression either in anthracycline pre-treated subjects (anthracycline + paclitaxel followed by trastuzumab) or in combination with docetaxel in anthracycline naïve subjects. Similarly, treatment strategies in breast cancer patients could differ based on tumour expression of estrogen receptors and HER2 receptors (use of trastuzumab+ anastrozole in post menopausal women who are ER+ and HER2+). Further examples of such marker determined strategies that influence treatment options or treatment durations are noted in HCV infections related to the viral genome (PEG-IFN + ribavirin) treatment duration of 48 weeks for genotype 1 and 4, and a duration of 24 weeks for genotype 2 & 3),[11] if the viral genotype is considered a GBM similar to tumour markers.

In these situations, the treatment allocation presupposes that the GBMs are predictive of response to treatment algorithm as a whole and may or may not predict response to individual agents within the scheme. In cases where GBMs (singly or in combinations) are used for selection of treatment strategies, it will be necessary to clearly define the treatment algorithm, the stratification and the eligibility criteria for subject entry into such studies. Within drug development programmes, it will be

[11] PEGINTF- EPAR_PI -WC500039195.pdf

crucial to define and detail the analysis plan including the criteria used to define a positive response prospectively.

3.4. Specific considerations for GBMs

Signal generation for genomic biomarkers may be slightly more complex than other types of BMs and depend on the type of material required (DNA, RNA, protein etc). The specific issues include consistency of sample collection, sample processing, assay methodology and opportunities for misclassification. There could be differences between different laboratories (central and local) in evaluation of the biomarker status. Such interlaboratory differences in the assay and misclassification of subjects could render the trial results less meaningful or even invalid when marker status was an entry criterion for the study or trial or for treatment allocation. Not surprisingly, such interlaboratory differences could also affect quantification and qualification of the genomic BM including estimation of their usefulness. Use of a single (central) laboratory may reduce the risk of different misclassifications but not necessarily guaranteed to avoid it altogether (as the same misclassification might occur repeatedly).The case of HercepTest highlights this aspect of a single technique and its concordance with the assay used in clinical trials[12].

The Biomarkers (GBMs) in the field of oncology present one additional consideration. In certain tumours, GBM expression may differ between primary and metastatic sites as well as in response to treatments.[13] In order to avoid such extraneous influences that may affect the outcome, effort should be made to establish a consistent pattern of sample collection, storage and evaluation. Where possible, GBM status of both primary and metastatic tumours should be evaluated during early development of the GBM. If a difference in marker status is noted between primary and metastatic sites in early studies, it may be necessary to define the relationship more clearly during late phase or pivotal trials. Where possible, stratification of subjects by tumour type (including histology, BM status and other factors) should be considered during the confirmatory trial(s).

3.4.1. Technical considerations for specific types of GBM

Variations in genomic DNA are robust molecular markers, usually easy to detect technically, using blood or tissue samples. Whilst the accuracy of genotyping is generally high, false positive or negative results do exist and may amount to a few percent per polymorphism in large samples (EU reflection paper on samples, test & data, handling). This however is dependant on the gene studied and the genotyping method used. Methods for detecting mis-genotyping at the population level have been described and should be utilised in any development programme to provide reassurance. Hence data quality assessment for genotyping should always be included in the study protocols.

On the other hand, mRNA biomarkers based on transcriptome studies are related to quantitative variations subject to both biological and experimental variability. Reproducibility should be tested in paired tissue samples. Results of transcriptome analysis should always be confirmed and extended for the selected genes by using other, independent methods for mRNA quantification, and/or protein quantification. Claims of physiological significance and clinical association should be based on stringent statistical procedures. Recommendations for generation and interpretation of transcriptome data are available and should be followed.[14,15]

GBMs related to tumour genome may have the additional consideration of stability both in vitro and in vivo. In a small percentage of subjects the marker status may change due to a number of reasons including change in clinical status or the instability within the tumour genome including the GBM.

[12] HercepTest IFU,- page 25. http://www.dako.com/uk/download.pdf?objectid=120856003
[13] Yonemori K, Tsuta K, Shimizu C et al.. J Neurooncol. 2008 Nov; 90(2): 223-8
[14] Villeneuve DJ, Parissenti AM. Curr Top Med Chem. 2004;4(13):1329-45)
[15] Georgitsi M, Zukic B, Pavlovic S, Patrinos GP. 2011.. Pharmacogenomics, in press

These however appear to be the exception than the rule. It is expected that this issue will be addressed during the qualification of the BM. Care should be taken to ensure that poor storage for prolonged periods of time do not alter the stability. Another aspect that will require consideration in relation to the markers and /or clinical trials in oncology is the consistency of marker occurrence in the tumour genome between primary tumour and metastatic lesions.

When genomic markers are used to define populations for treatment, misclassification of subjects is a risk (see foot notes 27 & 28, page 14) and this may impact the results and their interpretation. Care should be taken to evaluate reproducibility of the test used to avoid misclassification of subjects.

3.4.2. Timing of signal generation and impact on clinical development

GBM signals may be generated based on theoretical plausibility or based on an association noted in preliminary studies, and then confirmed empirically during the exploratory phase of development of the GBM. Alternatively, hypotheses may be generated during, immediately after, or a long time after the clinical development programme for a medicinal product. In some instances, a candidate biomarker may fit in to an existing body of knowledge (CYP450 polymorphisms or other drug metabolising enzymes) but in other cases novel GBMs may be investigated (or generated) within or outside of a drug development programme. While the evidentiary burden will differ between circumstances, it is expected that sound clinical and statistical principles are followed in all of these situations. In general, confirmation of findings obtained from early signal generating studies in a prospective pivotal clinical trial is expected and this more likely to be the expectation for markers of efficacy. The prospective clinical trials should provide a detailed analysis of the basis of the association, the interaction with the relevant therapeutic intervention, the predictive value of the marker and then its clinical utility. Occasionally, a GBM may be identified during or after the pivotal trials are completed through a retrospective or an exploratory analysis of the phase III trial due to the fact that the larger sample size in the pivotal trials provide greater opportunity both for defining the benefit more clearly or for identifying the low frequency ADRs, not evident in the exploratory clinical studies (for example Panitumumab & KRAS). The second scenario is more frequently true for GBMs related to safety. Indeed, markers (GBMs) related to safety may be identified after the medicinal product has been marketed, necessitating updates to the product literature. This sometimes is referred to as “retrofit” i.e., the product literature and clinical use may be modified from the original authorisation, once data relating to the GBM become available from post marketing studies or observations. This type of rescue strategy is not optimal. Evaluation of hypersensitivity reactions to abacavir serves as the exemplar of such a situation. Retrofit does not invariably imply retrospective analysis but may include such evaluation.

When a GBM is identified after a medicinal product is marketed for a certain length of time, there are certain specific aspects which assume importance in the development and evaluation of the GBM. These may also impact on the ability to achieve stringent evidentiary requirements in the post marketing (post authorisation) era and include; the frequency and severity of the safety event, the sponsor’s interest in evaluating the drug –event interaction, feasibility of conducting a prospective randomised trial for confirmatory evidence and ethical issues in case of serious or life threatening events. Often, there may be little interest from the industry / sponsor to pursue such development and other funding sources will need to be explored. A comparison between abacavir and carbamazepine highlights these points. For abacavir, the safety event was identified early in the post marketing period and thus retained the sponsor’s interest, was still within the patent period and a risk evaluation plan(risk management plan) could be generated relatively easily. The PREDICT-1 study used a randomised, double- blind design to assign treatment using abacavir with or without pre-testing for *HLA B*5701* allele, but in other situations data may become available through retrospective analysis or case control studies. The carbamazepine & HLA-B*1502 link with Steven Johnson syndrome provides a contrast in that it was identified late in the product life cycle based on a case-control design, the event rate was

lower (rarer) by an order of magnitude, there were a number of generic products on the market and there was ethic variability (specificity) with the data arising mainly from academic/clinical centres. There was little sponsor involvement in its evolution.

When a GBM is either identified or its predictive ability noted retrospectively after a pivotal trial is completed (e.g. panitumumab and *KRAS Wild type*), ideally such findings are expected to be confirmed in a prospective trial as detailed previously. However, as before, prospective studies may still not always be feasible due to many reasons including the need for a large trial population or sample size (e.g. SLCO1B1 polymorphism and rhabdomyolysis with statin use). Confirmatory evidence from well conducted case control studies, observational or epidemiological studies might also serve the purpose, the emphasis being on independent verification.

When prospective studies are precluded (reasons of rarity of event and ethical dilemmas in case of life threatening events or lack of funding support including commercial disinterest), there are two possible alternative scenarios or options for progressing the development of GBMs; one is to extrapolate from previous scientific knowledge and second is to obtain data from retrospective samples or analysis. Both approaches have significant limitations and extrapolation from prior scientific knowledge may not be possible for novel GBMs. It is recognised that findings from a retrospective analysis (association noted between GBM and drug response) replicated in an independent population or sample might provide supportive evidence and be sufficiently persuasive depending on the particular situation. Such an approach might be considered in cases where data from two completed but well conducted independent RCTs are available. Alternatively, the post hoc analysis could randomly define a testing sample and a validation sample within the same trial or a pooled dataset to investigate the association between GBM and the event assuming that the database (or trial) recorded sufficient number of events and the prevalence of the marker in the population is available. It is possible to hypothesize that a retrospective analysis of the existing database might be the preferable option to identify predictive GBMs when it relates to risk of particular toxicity (e.g. *HLA DRB1*07* or *DQA1*02,* Ximelagatran, liver injury and EXTEND study)[16]. The application for authorisation for Ximelagatran was withdrawn in EU and globally in 2006, based on the results from the EXTEND study.

In situations akin to those detailed i.e. when the evidence is primarily retrospective, certain requirements could be envisaged for the evidence to be persuasive: i) the strength of the association should be high; ii) the biological plausibility for the interaction should be strong; iii), the marker status of the majority of the subjects in the dataset should be known to avoid bias and iv) the diagnostic performance of the marker for the measured outcome should be of acceptable level. There may be additional elements such as the temporal relationship. For example, in the case of ximelagatran a direct thrombin inhibitor, the liver injury developed after exposure had been completed and routine monitoring limited to the duration of the trial only would not detect or mitigate risk of liver injury. Therefore it is important to evaluate and define the temporal relationships with adequate follow-up of subjects. This delayed occurrence of liver injury precluded a subsequent prospective study but offered the opportunity to revisit cases post-hoc (after exposure was completed).

3.4.3. Reduction of BIAS

Bias or confounders may play a considerable role in selection and validation of GBMs. These may be relevant only in certain circumstances and some are only noted with particular trial designs. Of the various types of bias, selection bias and measurement bias are of importance in the development of a GBM in addition to confounders. *Bias* is easier to minimise in prospective studies and is likely to be reduced by proper design and execution of the study including appropriate blinding and randomisation. Selection bias could impact retrospective analysis significantly, in particular because not all relevant trials will be accessible (publication bias) and those that are might not be described comprehensively,

[16] Agnelli G, Eriksson BI, Cohen AT et al. Thromb Res 2009; 123(3) 488-97

focussing instead on more favourable aspects of the trial results. A larger sample size may increase precision but does not remove bias and is not limited to retrospective trials. Additional considerations (for retrospective studies) include bias arising out of incomplete outcome data due to any of the following; exclusions, attrition, and/or reporting or publication bias. *Measurement bias* is an important consideration in relation to GBMs in a retrospective analysis and is likely to occur when different instruments or methodologies are used for measurement, especially in a meta-analysis of studies, the common thread being the GBM. A centralised measurement laboratory technique or test for the GBM with well defined assay sensitivity and specificity is likely to aid in reducing this, both retrospectively and prospectively. Moreover, careful selection of the studies included in the metaanalysis and pooled dataset with predefined criteria for selection is also helpful in avoiding the introduction of some types of bias.

3.4.4. Multiplicity

Regardless of whether the investigations are prospective or retrospective, the problem of multiplicity (increased false positive error rate due to multiple comparisons being made) will need to be addressed in the development of a GBM. Multiplicity in this context encompasses two distinct aspects; one is the use of multiple GBMs or a panel attempting to identify which have sufficiently strong associations with outcome. When multiple potential GBMs are examined in a development programme, the number of GBMs examined will depend on the signal generation approach which may investigate potential associations across the entire genome or fewer potential GBMs if the basis for exploration is more targeted.[17, 18] These issues assume greater significance in common multifactorial diseases where a single GBM might not be sufficiently predictive and multiplex testing might offer advantages. The main purposes of control of multiplicity here is for the company or investigator to follow reliable leads only and, for evaluation by both company and the regulators of the strength evidence for the association identified.

The second is the issue around multiple testing within the clinical trial. For GBMs to be investigated in prospective studies, the sponsor will wish to consider issues around multiple testing in the analysis plan and, if properly implemented, this should control the regulatory risk from multiple testing. For retrospective evaluations this cannot formally be controlled and inference has therefore to be particularly cautious. However, a number of potential corrections have been proposed in the literature, including those by Bonferroni, Benjamini-Hochberg or Sime's. From a methodological perspective, a statistical procedure that protects against false claims of significance while addressing the correlated nature of multiple testing for genetic interaction is reasonable. While Bonferroni's correction is criticised for being conservative, from a regulatory perspective associations that retain statistical significance even under the more extreme correction methods might be more persuasive, in particular when evidence comes from retrospective trials. Reference is also made to the CHMP guideline on multiplicity issues (CHMP/EWP/908/99).

4. Development of GBMS

4.1. Exploratory development

4.1.1. Non-randomized [cohort, case-control or single arm] studies;

Frequently GBMs are identified as an exploratory parameter in non-randomised cohort or single arm studies (within or outside of drug development programmes). These GBMs may be prognostic for disease severity, outcome etc, or predictive of a particular response to single or combination therapies.

[17] Yang Q, Khoury MJ, Botto L et al. *Am J Human Genet*. 72 : 636-649, 2003.
[18] Janssens CA, Pardo MC et al. *Am J Hum Gent.* 74:585-588, 2004

Such studies for identification and development of GBMs are likely to vary widely in their designs, especially in the early stages. These exploratory studies tend to be poorly selected convenience cohorts of limited sample size, and often lack sufficient rigor to establish the predictive value of the GBM and to quantify its sensitivity and specificity. Many studies lack pre-defined (clearly established) biomarker related end points or analysis plans. In some studies, the eligibility criteria may have been independent of the biomarker status at the time of entry. While this may be equivalent to and has the advantages of an unselected study design, the lack of a GBM based treatment allocation is a limitation and therefore does not provide true validation of the marker.

Genome wide association studies (cross sectional investigations of an association), often serve as useful tools for identification of a genomic marker when a large variability of phenotype exists but with a single common characteristic of interest. They serve as a search strategy rather than specific developmental design. When retrospective association studies (GWAS) provide the initial evidence of a link between the GBM, the disease and drug response, they often suffer from limitations similar to those stated above. In any retrospective analysis, it is important to consider the population (sample) size where the association was established as this largely depends on the availability of biological sample (blood, tissue or other) from a large majority of the subjects to avoid selection bias and other common potential biases that impact on the representation of the population identified by the GBM (see also sections 3.4.3 and 4.2.2).

Case control studies may provide useful information where the number of cases is limited, although the overall population from which the cases and controls are derived might be considerable but they may not provide definitive evidence. The main points for consideration in case control studies are the definitions applied to cases and controls, the ability to extrapolate the findings to the general population and any differences that exist in handling of the two groups including therapeutic interventions. There could be selection bias where GBM is used to define the disease or risk associated with particular treatment (applies to any retrospective exercise). Case control studies are retrospective evaluations that may limit the utility of the therapeutic intervention or its assignment and ability to determine the true, unbiased impact of the intervention on the natural history (for example discussion regarding tamoxifen in breast Ca; genotype based warfarin dosing). In contrast to case control studies, cohort studies could be prospective or retrospective and provide incidence and natural history of the disease but rarely of drug response because of the absence of a concurrent control arm. One important aspect of these cohort studies is that the patient selection may not be based on the marker but other, clinical parameters. They may have limited value in developing a genomic biomarker predictive of drug response but provide clues towards a marker of interest with a defined outcome. This however is limited by external influences or confounding variables. The genome wide association study evaluating the link between SLCO1B1 polymorphisms, high dose statin use and myopathy (SEARCH study[19]) in the background of a randomised outcome study is of interest and highlights some of the confounding factors.

On occasion, preliminary information relating to the GBM might arise from previous observations on other drugs of the same class or drugs with a shared characteristic (e.g. increased rate of adverse events in *CYP2D6* poor metabolisers [PMs] might span across drug classes that are substrates for *CYP2D6*). Therefore, for a new agent it is appropriate that confirmation of the relative importance of that particular GBM in man is obtained early (e.g., role of *CYP2D6* polymorphims on the effects of a new *CYP2D6* substrate drug), prior to registration (or approval) of the agent. In cases where data regarding the GBM become available after registration (or approval) or even patent expiry subsequent

[19] The SEARCH Collaborative Group— A Genomewide Study NEJM, volume 359: 789-799; Aug 21, 2008.

clinical trials that are planned and executed in a targeted population may be needed. Such a development programme is likely to involve both cohort studies and prospective RCTs. The cohort studies in this context are likely to provide background information on the marker while prospective RCTs will evaluate the true effect by reducing impact of confounding variables. The ongoing debate about the role of *CYP2D6* polymorphims and the use of tamoxifen is an example that highlights some of the difficulties when data become available in the post marketing phase.[20] Schroth et al examined the impact of CYP2D6 polymorphism in a retrospective cohort study in 1325 patients while Wegman[21] and colleagues evaluated this in ~220 subjects of a group of 680 patients. The studies differed in the context of patient groups included, treatments considered and availability of tumour tissues for genotyping. Other studies[22] have evaluated additional GBMs and emphasized the interaction between markers and the complexities in evaluating the importance of markers retrospective exercises.

4.1.2. Randomised control studies (RCTs -prospective or retrospective evaluation);

Exploratory investigation of GBM (hypothesis generation) through randomised clinical trials is often possible where preliminary information regarding the value of a predictive GBM is based on published literature or from early studies within a development programme. These could be new prospective RCTs or retrospective analysis of data from a completed trial or trials. Use of a prospective RCT for identification (and validation) of GBMs would be ideal but for certain constraints; they are expensive, time or effort intensive, and often need significant preliminary evidence to demonstrate either association or biological plausibility prior to the RCT. Designs applicable in such instances would be similar to pivotal trials for validation and are discussed in section 4.2 of this document. Alternatively, a retrospective analysis of a completed RCT (comparing two different drugs or treatment strategies) could act as the hypothesis generator. For such retrospective exploration or validation, certain elements are critical: that data should be available from well conducted RCTs, GBM data from sufficiently large number of subjects within the trial should be available to avoid selection bias; the analysis plan should be pre-defined. The panitumumab experience is a case in point. In the pivotal Phase III study, EGFR status was an inclusion criteria and therefore GBM data were available in all randomised subjects and thus reduced the possibility of selection bias and the analysis by KRAS mutation was pre-specified, albeit as an exploratory investigation. Analysis of data obtained from two (or more) independent and well conducted RCTs provide the strongest evidence. It is anticipated that in majority of such cases, confirmatory evidence from a pivotal RCT will be available.

4.2. Confirmatory development

The confirmatory step for establishing the role of a GBM assumes that a single GBM or GBM signature (panel of GBMS) has shown promise in early development with sufficient rigor to be taken forward to obtain clinical validity. A GBM with high positive and negative predictive value in exploratory studies would be one of interest although the level of stringency to be applied for selection must be determined on a case-by-case basis and cannot be specified here.

[20] Schroth W et al. JAMA, 2009; 302 (13): 1429-36.
[21] Wegman P et al. Br cancer Res 2005, 7 (3): R 284-90.
[22] Kiyatoni K et al. J Clin Oncol, 2010, 28 (8): 1287-93.

4.2.1. Trial designs for prospective validation:

The trial designs used for confirmatory development are likely to be influenced by factors that vary between markers such as: the pathway or marker involved; the mechanistic or biological relation between the marker, the disease and the planned intervention; the prevalence and inheritance pattern of the marker in the population; the hypothesized effect size; influence of ethnicity and gender; and the analyses planned including any stratification utilised. The analytical validity available for the GBM at the time of inception of the trial is also likely to be an important factor.

RCT is the preferred design for the pivotal/confirmatory trials for prospective validation of biomarkers (especially predictive markers, the main focus of this document) as stated before. Several forms of RCT are possible; unselected, enriched or targeted, hybrid and adaptive designs, the latter three being more specific in terms of the population enrolled and final analysis. Some aspects of the designs are discussed below. It should be noted that when the prevalence of the marker is rare, it is best to seek additional advice as none of the scenarios discussed below might best fit.

Unselected design RCTs

In general, trials using the unselected designs are likely to be most useful when eligibility for entry into the trial is not based on biomarker status. The unselected RCTs can be broadly classified into a) sequential testing strategy designs, b) marker-based designs, or c) hybrid designs, which are differentiated from each other by the protocol specified approach. The primary analysis will be dependent on the strategy adopted. For example, in the sequential strategy design, the response to the treatment in the overall population could be the primary analysis with the marker dependent response as the secondary analysis but modification of this analysis plan is possible i.e., the marker dependent response as the primary analysis and response in the overall group as the secondary analysis. The sample size requirement for such a design is likely to be larger (than other designs), and a clear demonstration of benefit in the prespecified GBM based analyses will be expected. This does introduce a level of difficulty in decision making when the overall trial shows no clear benefit but GBM based analysis does. This may be overcome by pre-specifying the GBM based analysis. It is important to consider the requirements for an application based on a single pivotal trial in these situations. Frequently the results of the secondary GBM based analyses are likely to need further confirmation in a second trial of sufficient power that may use alternative designs. The trial evaluating use of panitumumab (20020408) in metastatic colorectal cancer for a third line indication used this design to recruit subjects who all had have EGFR+ tumours as a primary inclusion criterion. The response to KRAS-WT or mutant KRAS served as the pre-planned secondary (or exploratory) analysis. As the selection of subjects into the trial was not related to KRAS mutation status, for this analysis (WT vs mutated KRAS), the trial behaved as an unselected design trial.

Enriched design RCTs (targeted design);

Enriched or targeted designs are those in which marker status forms the critical eligibility criterion, i.e., subjects are included based on the presence or absence of the marker. If enriched or targeted design RCTs are used, strong biological plausibility linking the GBM and disease and persuasive preliminary evidence of association between GBM & drug response are necessary. As this is a GBM defined population, the reasons for exclusion of subjects outside of the GBM defined population will need to be clearly defined. Targeted enriched designs are most applicable when the GBM either forms or influences the therapeutic (drug) target directly. The most popular (successful) example of enriched design studies are the trials evaluating response to trastuzumab combined with paclitaxel in Her-2 positive post surgical patients after combination therapy with doxorubicin plus cyclophosphamide, where, trastuzumab produced a ~25% reduction in the hazard ratio for DFS (disease free survival).

The enrichment design presupposes that the assay accuracy and reproducibility are very well established and, that there is little opportunity or possibility for misclassification of subjects (as GBM+ or GBM-ve) as misclassification might compromise the integrity of the trial and the actual benefit questioned (see also section 3.4.1)[23,24]. This design in its many forms is only powered to detect differences in outcomes in the group randomised to the marker defined treatment and provides no information on the remainder of the diseased population. Therefore it only validates the positive benefit: risk ratio of the treatment in the selected (marker based) population. This design is likely to be most valuable when the treatment benefit in the overall population is modest but with an unacceptable level of risk (in a maker defined population, GBM+ or GBM-). It is important to note that if the difference in response rate between investigational agent Vs placebo is the same as the difference in response rate between GBM+ vs GBM-ve subjects treated with placebo (even when evaluated separately), the predictive value of the enriched trial is likely to be rendered uninformative.

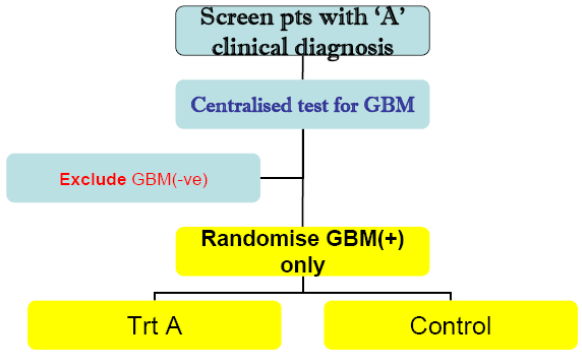

Fig-1; Enriched / Targeted Design

In general, enriched designs may be most useful where therapies have modest benefit with significant toxicity in the unselected population or when unselected design might be ethically not possible.

Marker based designs;

There are a number of examples where marker based designs have been adopted for drug development or validation in the context of a binary marker. These could be marker by treatment-interaction design or marker based strategy design. The marker by treatment interaction design uses the marker as a stratification tool and patients are assigned to treatments within each subgroup. The main advantage of this is that sample size is prospectively defined within each subgroup and also that it is equivalent to two RCTs.

In the marker-based strategy design, patients are randomly assigned either based on or independent of the marker status. In the latter case (not shown in the figure), the overall detectable difference in outcomes is reduced and the sample size becomes larger.

[23] Perez EA, J of Oncology 2006, 24: 3032-3038
[24] Paik S et al. J Clin Oncol 2007, 25:511

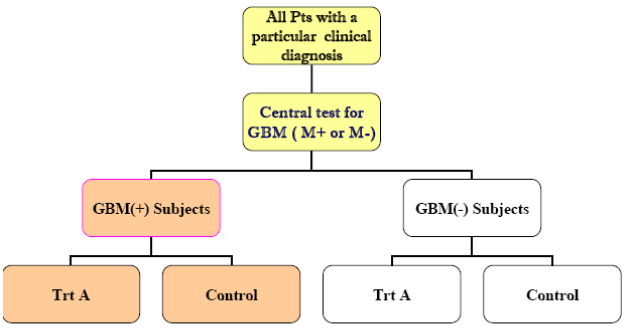

Fig-2a: Marker by treatment interaction Design

549

550 *Hybrid design RCTs;*
551

552 In the hybrid design (as explained here), only a subgroup of GBM defined subjects are randomly
553 assigned to the treatment under investigation based on the marker status while the other GBM defined
554 subjects are assigned to standard care therapy(s). Although the trial is powered similar to the enriched
555 design, such a strategy could add additional value. This design is most useful when treatment involves
556 multiple agents or strategies with compelling evidence of efficacy with certain treatments. The
557 standard care therefore will include the previously defined treatment while the experimental arm and
558 the overall trial will provide additional information relating to subsequent or additional treatment
559 options for GBM defined subjects.

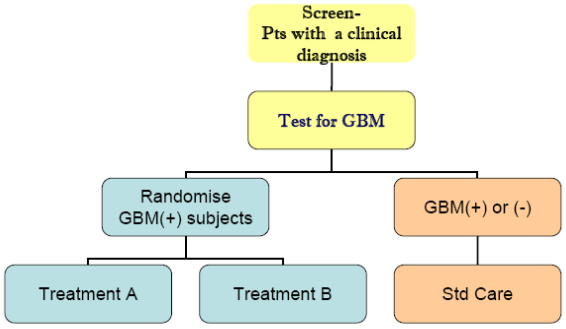

Fig 2B: Hybrid design (one example)

560

561 The perceived advantage of the hybrid design is that there is potential for incremental efficacy over
562 standard care and subsequent comparisons but it requires that samples for GBM assay obtained at
563 screening are stored for future testing for other prognostic markers. A modified version of the hybrid
564 design was used in Predict -1 study to evaluate abacavir hypersensitivity. Subjects with a clinical
565 diagnosis of HIV were randomised to genetic test group and standard care group. The former group
566 were administered abacavir only after testing negative for HLA-B* 5701 excluding patients who tested
567 positive, while the standard care group were not tested for HLA-B*5701 before exposure to abacavir.
568 Both groups were monitored for hypersensitivity reactions. While this is a classical case of safety
569 interaction, it could also be interpreted as a trial evaluating utility of HLA-B* 5701 marker.

570

571 *Adaptive designs;*
572

573 There is increasing interest in adaptive designs in recent years. These could be adaptive threshold
574 (statistical analysis) design, adaptive accrual design or adaptive randomisation based on outcome.

Combinations of these are also possible. The adaptive threshold design permits two methods of analysis; one, a pre-specified threshold level of significance for the overall comparison with a different level of significance for subsequent comparisons, depending on the alpha spending or the second type of analysis which assumes effectiveness only in GBM+ subjects and tests for this. Such a design would also be useful when there is need to test of effect of treatment and prospective validation of a cut-off point for the chosen marker. The adaptive randomisation scheme permits modified accrual into a specified treatment group based on an interim testing for futility. Few successful adaptive design examples are available at this point in time in the regulatory context and indeed few such designs have been tried in a clinical trial, but the potential exists.

4.2.2. Comparison of different designs (pros & cons)

There is abundant literature comparing different designs of the trials used for validation of GBMs. When there is a true predictive marker with high biological plausibility based on available evidence and scientific background, the enriched design is likely to be the most efficient but this has two prerequisites; one, there should be a cut off point for determining the marker status and second, there is need to ensure that misclassification does not occur in order to avoid the trial losing its value. In cases where the marker cut-off point is not established, but the marker prevalence is high, the unselected design or its modifications offer the most suitable option, but may suffer from a need for larger sample sizes. In comparison, the targeted design requires fewer randomly assigned patients and indeed fewer patients screened when compared to the unselected design, although this is dependent on assay accuracy, reproducibility, and marker prevalence. The regulatory acceptability of excluding GBM-ve patients from trials will depend on the strength of evidence (plausibility, scientific rationale and clinical data) provided for the lack of effect in these patients.

Limitations of the enriched design include: enriched design does not validate the GBM itself but only the benefit of the treatment in question in the specified population. The results might be irrelevant if, the difference (Drug - Placebo) noted in the GBM based enriched design study is same as the difference between GBM+ and GBM- subjects when treated with placebo. Assay Accuracy (used for classification of GBM+ or GBM- subjects) influences the unselected and targeted designs differently. If there is misclassification, in the unselected trial only inferences about the marker might be affected while in the enriched trial, this may compromise the overall integrity and the result of the trial in addition to inferences about the marker.

The marker based designs offer advantages in particular situations. In the context of a binary marker or multimarker signature that could be crystallised to a binary classification, smaller sample sizes and higher event rates (or larger event rate difference between groups) are likely with the marker by treatment interaction design compared with the unselected design. The marker based strategy design has a potential disadvantage; there is overlap of patients treated with the same regimen on both the marker-based and the non–marker-based arms. One caveat of note is that experience is predominantly in the field of cancer therapeutics and their applicability in other fields remains to be confirmed.

Is Retrospective validation possible? (confirmation);

When new prospectively designed trials are not feasible due to variety of reasons, the possibility to test the predictive ability of a marker using data from previously well conducted randomized controlled trials (RCTs) comparing therapies could be considered in certain circumstances (retrospective validation). For any retrospective validation crucial elements such as data from one or more well conducted prospective RCTs and availability of GBM status from a large number of subjects to avoid

selection bias are important. In addition, the hypothesis to be investigated and the plan for analysis should be documented before the retrospective evaluations begin. As discussed previously, for retrospective validation, use of one or more independent data sources or RCTs may provide the necessary evidence. The designs of studies included in the retrospective validation are likely to be similar to the prospective validation trials but likely to have a preponderance of unselected designs as regards the GBM. One point of difference between the two routes is that in a retrospective analysis of a previously completed RCT, eligibility for entry into the trial may not be based on the marker status (i.e. unselected design) and this may help validation of the marker. The study identifying relation between wild type KRAS in metastatic CRC and improved progression free survival (PFS) after panitumumab (Vectibix) provides one such example of retrospective validation.[25] In this instance, a differential effect of panitumumab between carriers of wild type and mutated KRAS suggested by the post hoc GBM analysis formed the basis of conditional authorisation in Europe, along with a biological plausibility for the association derived from trials of cetuximab. The authorisation stipulated that further data should be generated prospectively. The consideration that biomarker (GBM) status information should be available for majority of subjects was met in the trial 20020408. In this instance data was obtained from a RCT, analysis was prospectively defined and GBM status of majority of subjects could be determined avoiding any selection bias.

In comparison (or contrast), the interaction between EGFR FISH and/or EGFR mutation status, with Geftinib (Iressa in EU) was evaluated in several studies but as an exploratory objective in patients with non-small cell lung cancer. The studies included plausibly diverse patient populations (ISEL in Asians), INTEREST in all comers (with Caucasian preponderance), and IPASS in a mixed group. Whilst the studies were prospective, only INTEREST study included EGFR FISH + based difference as the co-primary objective, rendering these effectively to a post-hoc (retrospective) analysis. The differential response rates noted in these studies might have been influenced by differences in ethnicity, in other clinical features or prior therapy. The differences in the number of subjects with known/ identified marker status, for each of the biomarkers (EGFR FISH status, EGFR mutation status and EGFR protein expression) may also have a played a role. Notwithstanding the disparate results, the pooled analysis suggested benefit from geftinib therapy only in case of EGFR mutation positive tumours because of the directional concordance between various comparisons and the replicated interaction between EGFR mutation status and response to geftinib. Of note, based on the results of these multiple studies and pooled analysis, both the CHMP and the expert advice group concluded that while a broad indication for geftinib in NSCLC was not agreeable, the response to geftinib is influenced by the EGFR mutations status and a restricted indication was accepted. One criticism of the geftinib development programme is the lack of information relating to the biomarkers from all subjects included in various trials and this could be designed and organised better. This example highlights two important aspects of retrospective evaluation of GBMs: replication of the GBM – drug response interaction in different studies and populations; and secondly, the need for maximising the GBMs status information from all subjects in the analysis.

Overall therefore, retrospective validation or acceptance of retrospective data in the regulatory/ scientific context might be possible if the following aspects are fulfilled: data from conducted RCTs; availability of marker status information from majority of the subjects in those RCTs; a predefined hypothesis as well as analysis plan; a statistically compelling association having adjusted for multiple testing; and finally replication of the results in one or more independent samples.

[25] Vectibix EPAR-2007. http://www.ema.europa.eu/humandocs/PDFs/EPAR/vectibix/H-741-en6.pdf

5. Diagnostic performance of the marker

It will be important to characterise the diagnostic performance of the GBM or biomarker and to explain how this performance would be affected by the disease characteristics, effect of other therapeutic intervention and any epidemiological differences in gender or ethnicity in the expression of the GBM. For example, the relation between *HLA-B* 1502* and Steven Johnson syndrome when exposed to anti-epileptic agents demonstrates certain ethnic variability and this could impact the performance when used in a broad unselected population.

The standards of diagnostic performance should conform to the general standards for qualification and validation. Specific aspects such as repeatability, reproducibility and precision estimates should be evaluated.[26] The diagnostic performance relates essentially to evaluation of sensitivity and specificity.

The value of a GBM (or other traditional markers) is often specific to the population tested and utility would need to be demonstrated in populations that reflect clinical use. Any extrapolation to other populations will need to be adequately justified taking into account the prevalence of the GBM, the established phenotype in the populations tested and extrapolated, including estimated levels of risk for safety GBMS.

The points raised below may in general be applicable to both markers and in some instances tests, but a detailed discussion on the diagnostics is beyond the scope of this document.

5.1. Sensitivity, specificity, NPV, PPV

Irrespective of the type of trials used for exploration, it is important to collect information on the performance of the biomarker(s) for predictability and clinical validity. Clinical validity of a marker is a complex interplay of sensitivity and specificity (of the test/ marker) and the penetrance of the genomic abnormality or mutation. The developmental studies in addition to hypothesis generation, should aim to indicate whether further evaluation of the GBM is feasible and worthwhile. In this aspect they would serve equivalent to phase-II clinical studies but are likely influenced by the limitations of the exploratory data sets (in cohort or single arm studies). It is expected that some form of validation of the data generated in these studies will be available and it is necessary to perform this on data not used for generation of the GBM itself. The process should gather information relating to the predictive performance of the GBM(s) and this should include positive and negative predictive values that are evaluated further at the confirmatory stage.

The PPV and NPV are likely to be dependent on the prevalence of the marker and the phenotypic expression in the population of interest. A GBM with high positive and negative predictive value in exploratory studies would be of particular interest although the level of stringency to be applied is unclear especially in the regulatory context and therefore can not be exactly specified. However, in certain situations even 95% specificity might not be acceptable and the national cancer institute uses the example in screening for ovarian cancer with low prevalence.[27] The definitive assessment of a marker is often a multivariate analysis especially when other markers or features are available. When multimarker panels or signatures are used, the results of a multivariate analysis are dependent upon certain limitations; the cut off adopted for the GBM, whether continuous or categorical, whether all other existing markers were coded appropriately and how the variables were modelled. Alternative methods based on likelihood ratios, concordance index or change in concordance index have been proposed and might be appropriate depending on the situation. Yang et al used the likelihood ratio defined as the probability that a patient with the disease has observed test result, compared with the

[26] Biomarker qualification- renal biomarkers. http://www.ema.europa.eu/pdfs/human/biomarkers/28329810en.pdf
[27] http://www.cancer.gov/cancertopics/understandingcancer/moleculardiagnostics/page34

probability that a patient without disease has the same result.[14] The concordance index is the probability that of two randomly selected patients, the one with the worse outcome is in fact predicted to have a worse outcome.[28] The concordance index is similar to the area under the receiver operating characteristic (ROC) curve.

In a binary model (with values such as 0 & 1 or above and below a defined cut off), the ROC curve (receiver operating characteristic curve) provides the tools to select optimal models independent of the class distribution. The ROC curve also permits use of different cut points in continuous data sets, where the numerical value of the marker is relevant, provides the ability to calculate the likelihood ratio from the slope at any given point and the AUC of the curve is a measure of test accuracy. The markers for renal damage evaluated in the predictive consortium used this successfully for several markers (KIM-1, albumin, cystatin-c etc)[29] although there are very few examples of successful application of this to GBMs in the regulatory context.

6. Devices / diagnostic Kits for GBM assessment

The GBM and an associated assay to detect the GBM in other diseases or situations may be available on the market and in such instances the validation of the test is limited to the disease and treatment under evaluation as an acceptable assay or methodology might already be available (for example HLA alleles). For newly identified or specific GBMs, development of a specific assay/KIT might be necessary, parallel to the drug development. Often there is co-development of 'companion diagnostics' to a drug and issues relating to specificity and sensitivity of the assay of the marketed test need to be considered. Similar considerations apply to other assays or tests that might be available to detect the GBM (such as those commonly referred to as 'home brew tests'). The intention of this paper is primarily to discuss the methodological aspects related to the GBM and not the performance of the commercial test or kit. However, some observations are made below;

- In the context of drug development it might be possible to include a companion diagnostic within the development programme. When a particular diagnostic kit or methodology is employed in the pivotal trial and such test is specific to the identification/quantification of the GBM, it may be necessary to link the specific test method and the value of the GBM. Identification and quantification of HER-2 overexpression using immunohistochemistry (IHC) or FISH testing is one such example. In such instances the drug label will provide this additional information as part of the description of the results. Other such examples exist (Dako test and cetuximab, Monogram trofile assay and maraviroc).

- When the test used for identification of the GBM is of a more generic nature, i.e. not developed specifically for GBM (e.g., identification of CYP2D6 polymorphism) such a discussion on the characteristics of the companion diagnostic test is unlikely to be included in the drug label. It is also not anticipated that a specific diagnostic kit be made available.

- There may be situations where in the assay / test used in the clinical trial might not be available but replaced with a new or a different test. When such is the case concordance between the clinical trial assay and the marketed test for measuring the GBM marker status would be expected. Similar concordance between other commercially available tests is also an important consideration but a discussion of these is beyond the current scope of the paper.

[28] Kattan MW. Clin Cancer Res. 10: 822-824. 2004
[29] Nephrotoxicity BMs. Final conclusions of EMA/FDA VXDS experience. EMEA/679719/2008 Rev.1

7. Potential external influences on GBM evaluation

It should be recognised that the relation between a marker and the therapeutic effect could also be influenced by other genomic or non-genomic factors that are associated with the phenotype and determine an individual's response to the agent in question. This may be independent of the target disease or condition being investigated. For example, use of CYP450 substrates or inhibitors such as paroxetine administered simultaneously or concomitantly with tamoxifen has been argued to influence treatment outcomes in ER positive breast cancer when this is not controlled within the pivotal trial. This is similar to the effect proposed for certain *CYP2D6* gene polymorphisms that influence tamoxifen metabolism (*CYP2D6* substrate).

Similarly, polymorphisms that relate to drug receptors may influence response to one or more agents independent of the underlying disease or diseases. The response to the drug may also be affected by polymorphisms of other metabolising enzymes. For example, interaction of *CYP2C9* polymorphisms on response to warfarin is influenced by simultaneous inheritance of polymorphisms of *VKORC1* gene. The overall effect therefore would be a composite of the two influences and clarity regarding the relative contributions should be sought during evaluation.

A detailed discussion of evaluation of pharmacogenomic GBMs in early studies is available in a separate document (EMA/CHMP/37646/2009 - see below).

8. Other aspects

This reflection paper should be read in conjunction with the following notes for guidance:

- Position paper on terminology in Pharmacogenetics (EMEA/CPMP/3070/01)
- Reflection paper on pharmacogenomic samples and data handling (EMEA/CHMP/201914)
- ICH Topic E 15: Establish definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories (CHMP/ICH/437986/2006)
- Reflection paper on the use of pharmacogenetic in the pharmacokinetic evaluation of medicinal products (EMA/641698/2008)
- EMEA/CHMP/SAWP/72894/2008 Corr[1]. Qualification of Novel methodologies for Drug Development: Guidance to Applicants.
- EMA/CHMP/ICH/380636/2009; Genomic biomarkers related to drug response: context, structure and format of qualification submissions.
- EMA/CHMP/37646/2009; Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products;

9. Glossary

Accuracy

Reflects the degree of closeness of measurement of a quantity to its actual value.

Analytical Validity

Analytical validity in the context of genomic biomarker describes the ability of a particular test to measure accurately (and reliably) the genotype (marker) of interest. This evaluates the test performance.

Biomarker

A characteristic that is measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacological responses to a therapeutic intervention.

Classifier

A classifier is a mathematical function that translates biomarker value into a set of prognostic categories; it can also be defined as a marker that allows classification of patients.

Clinical Utility

Clinical utility of usefulness of the marker or test is the likelihood that the test(marker) will lead to improved outcome with a given intervention. *In other words*, it is the value that the marker (GBM) provides in predicting drug response or prognostic evaluation of a marker defined population over and above other standard clinical features.

Clinical Validity

Refers to the accuracy with which a test predicts the presence (or absence) of the clinical disease or phenotype.

Genomic biomarker

A measurable DNA and /or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes and /or response to therapeutic or other intervention.

(Biomarker) Qualification

Qualification is a conclusion that, within the stated context of use, the results of assessment with a Biomarker can be relied upon to adequately reflect a biologic process, response or event and support the use of biomarker during drug or biotechnology development.

Reproducibility

Refers to the ability of a test or experiment to be accurately reproduced or replicated by an independent worker/researcher/laboratory.