EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

2011-10-17

Submission of comments on 'Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection ' (EMA/446337/2011)

Comments from:

Name of organisation or individual
EFSPI [European Federation of Statisticians in the Pharmaceutical Industry www.efspi.org]

Please note that these comments and the identity of the sender will be published unless a specific justified objection is received.

When completed, this form should be sent to the European Medicines Agency electronically, in Word format (not PDF).

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

1. General comments

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	The guideline seems to focus on external validation – especially in a pharmaceutical development setting the necessity to perform independent (extra) studies may be a hurdle. Moreover, there are a series of statistical techniques to deal with the issues raised in this guideline through internal validation.	
	In general the reflection paper is of high quality and covers most of what is expected it should cover. However, there is one thing we want to address that the agency might want to consider to cover as well. That is how to evaluate the performance of a predictive GBM, e.g the ability of the GBM to predict who will benefit from the treatment. The paper does a good job describing the difference between prognostic GBMs and predictive GBMs in Section 3, but when it comes to evaluating the performance of GBMs we think it would help to make it clearer how to evaluate the performance of prognostic GBMs and predictive GBMs, separately. We are not sure if sensitivity, specificity or other related measures make sense in the evaluation of predictive GBMs as described in the reflection paper. To us it seems that evaluation of the performance of prognostic GBMs is covered only. Note that the FDA Drug-Diagnostic Co-Development concept paper from 2005 lacks the same clarity.	
	We like that the paper gives examples (e.g. HLA B*5701	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	and abacavir hypersensitivity). Has the agency considered to add references to the examples as well?	
	Many reflections in this EMA paper overlap with exsisting draft guidance from the FDA. As the number of documents with thoughts on this subject are immense it would maybe be worth to collect the basic thoughts and pninciples on the subject into a single ICH guidance document.	
	Comment on the term 'pharmacogenomic biomarkers'. 'Genomic biomarkers' might be a more appropriate term (Reflection paper on methodological issues associated with 'genomic biomarkers of relevance for pharmaceutical development').	
	This reflection paper provided comprehensive, clear and specific guidance to clinical development of pharmacogenomics biomarkers from the regulatory perspective. With increasing PHC (Personalised Health Care) component in Drug Development, it is also very timely. It’s one of the most useful guidance documents on the subject.	

2. Specific comments on text

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
33,34		Non-randomized, randomised: please be consistent in spelling	
52		"reduce overall development cost" doesn't seem to be a realistic prospect. Development of predictive biomarker often involves greater cost, not less. Surrogate markers could help to reduce cost in early development but that's not within scope of this document Please change "reduce overall development cost" to "reduce failure rate in development".	
68-76		It would also be helpful to acknowledge that the development of pharmacogenomics biomarkers is an area of active research. There are some limitations in the principles summarised here. Most examples used were on targeted therapies in oncology. Some GBMs are related to the pharmacokinetics of the drug predictive of the safety outcomes. In other areas, while the same principles generally apply, there is greater challenge in the development.	
74		Please remove ":" between benefit and risk	
82-76		The term GBM and pharmacogenomics are not clearly defined. It would be helpful to define them/clarify these terms in the scope or in the next section. For example, many examples are genetic markers. While protein markers are mentioned later in the text, it's not immediately clear that they are within the scope.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
82-92		It's also helpful to mention that only pre-treatment baseline markers are considered here. Some post treatment protein markers could be used to monitor the treatment and predict future response. They may help physicians to decide whether to continue the treatment. They are outside of the scope.	
103		Please re-phrase ("Handling of such these are …")	
109		"pre-treatment" may be the most desirable, but this is not necessary. Also (early) changes during treatment may be good markers . Pleasedelete "pre-treatment"	
114		Please define ROCs as Receiver Operating Characteristics.	
115		Confirmation in a second trial is desirable, but is in many cases unfeasible. Please change "would be expected" to "would be desirable".	
126		Proposed change: for example –novel drug target identification	
130-132		Meaning unclear ("The role of panitumumab…is liable to interpretation as an efficacy marker"). It is not clear what the efficacy marker is in the panitumumab example.	
151		"should include a set of classifiers" is misleading. Consider change to "should include a prediction / classification rule"	
157 - 159		This sentence is difficult to interpret. Is there an expectation that the biological rationale is provided? This may not be realistic at an early stage. Please rephrase.	
170-172		Proposed change: and following example relates to classifying safety issue- for elevated bilirubin levels, patients with A allele of rs887829 (a polymorphism in UGT1A1) are predisposed to	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Gilbert's syndrome an inherited but benign condition	
215-223		Please add the following text somewhere: Even within the same lab, batch effect can occur due to the experimental factors such as machines, operators, lot number of reagent etc. Co-development diagnostic assay is necessary to ensure that the assay produce consistent classification of subjects prior to the start of a confirmation trial.	
220 ff		The use of a single laboratory precludes the assessment of reproducibility	
239-240		Strongly agree that the data quality assessment should always be included in the protocol. However, it should be restricted to genotyping. For gene expression and protein data, the quality assessment is even more important. Proposed change (if any): recommend putting the sentence "data quality assessment should always be included in the protocol; the specific type of quality assessment needed is dependent on the type of GBM and tissue".	
257		"see foot notes 27&28, page 14" – We do not find those foot notes	
258		"Care should be taken to evaluate reproducibility of the test used to avoid misclassification of subjects" Consider to take this a step further, i.e. consider to add some guidance, or an example, how this can be done.	
275		Please define ADRs as Adverse Drug Reactions.	
276		Please define KRAS.	
300		"ethic variability" should probably be replaced with "ethnic	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		variability".	
326-330		This is extremely relevant and helpful. Proposed change (if any): reference section 4.2.2	
338-355		Important: One key area of bias for GBMs not covered in this section or document as a whole is with respect to endpoint ascertainment. If this is not fully objective, this can also significantly dilute the strength of the link between a GBM and event of interest. For example, for HLA-B*5701 and abacavir hypersensitivity reaction, diagnosis of HSR is imperfect. A skin patch test was developed to help refine case ascertainment within clinical trials but this will not always be feasible for other situations.	
338-355		Meta-analysis posed many other challenges, in addition to bias. Typically, meta-analyses are only used in post marker researches, not in the clinical development of GBM. It's not mentioned elsewhere in the document. The bias mentioned here is well known. It would be simpler to exclude meta-analysis issues from the scope of this document.	
338-355		Proposed change: mandatory sampling is strongly recommended when feasible to avoid selection bias.	
338-355		Fewer patients taking part in clinical trials consent to participate in genetic/genomic analysis. Proposed change: For development/validation of GBM in clinical trials, often different studies across a compound are pooled to generate sufficient sample size, and if not carefully controlled, a big source of bias often accrues due to variability of consent by patients	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		which can be included in a GBM analysis.	
340		"measurement bias" -- in this paragraph the statistical concept bias is to some degree mixed up with concepts such as measurement error and accuracy. Please consider a more clear formulation.	
369-380		This section could be reduced in size as the CHMP guideline on multiplicity issues is already referenced. Bonferroni is indeed very conservative. Bonferroni-Holm could be applied – this is statistically valid, it would also address the regulatory concerns, and it is less conservative. Proposed change: Use Bonferroni-Holm, or refrain from providing specific advice.	
398		The abbreviation GWAS is used without having been introduced earlier. Proposed change: Consider to add a list of abbreviations.	
400-400		This sentence is not very clear. I have to read it many times to understand what it's trying to say. The selection bias mentioned here is different from the selection bias mentioned before. Regardless of sample size, one has to check whether the identified GBM is confounded with other variables that characterise the patient population (e.g., region, ethnicity). A more serious issue with GWAS in a small sample size is the lack of power. Due to the burden of multiple testing, an exceptionally large sample size is required to detect a GBM, unless the effect size is very large. This point is worth mentioning here.	
412-415		Please re-phrase. Not clear how a case control study can limit	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		utility of an intervention or its assignment…	
462-464		Does this mean that two RTCs would be ideal but at least one is required? Please clarify.	
469-471		"A GBM with high positive… must be determined on case-by-case basis and cannot be specified here". Consider to explain why that must be done on a case-by-case basis, and perhaps mention examples of what can kind of thing that will affect the level of stringency.	
483		Please substitute "rare" -> "low" for prevalence	
491-496		This appears to describe analysis strategy rather than design as such. Please consider moving to separate analysis strategy sub-section	
492-496		This suggests that in sequential multiple testing there is a free choice on the order going either from overall to subset or from subset to overall. Is this generally accepted or is it only accepted in this case?	
530		maker should be marker	
530-533		This seems to be a confusion of prognostic and predictive effects of the GBM. Even if the predictive effect (diff in response rate between active and placebo) is of the same magnitude as the prognostic effect (GMB+ vs GMB- on placebo) this may well be clinically relevant.	
535-536		This sentence is a duplicate of the sentence on lines 528-530.	
550 ff		This section fails to clearly describe the merits of this type of design. The text and figure do not correspond. For example, the text states that some GMB+ subjects are randomised to	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		investigational treatment and others to Std Care. This is not clear from the figure.	
550 ff		It is difficult to understand the purpose of the hybrid design, or even how it works, eg how do you decide which subset of GBM+ve gets randomised and is the experimental arm the same as the standard of care arm.	
559-560 (Fig. 2B: Hybrid design)		As drawn, this design only provides valid comparisons of treatment A vs treatment B for GMB+ subjects unless subjects are randomized to biomarker based treatment or not (in which case it could also provide comparisons bewteen treatment A or B versus Std Care).	
561-562		What dose this sentence mean? ("potential for incremental efficacy over std care and subsequent comparisons")	
567-569		In order to evaluate utility it would not seem reasonable to compare all those randomized to genetic testing group to all those randomized to standard care group (without excluding GBM- pts in the genetic testing group). Otherwise the comparison is biased.	
571 ff		That section on adaptive designs is quite vague (as expected as there are limited practical examples on that in this field so far). However, I think it would be interesting to hear the agency's current view on adaptive design that is adaptive in the population, e.g. you start with a broad population but might end up in a targeted population based on interim analysis, e.g. in a phase III trial.	
579		re-phrase ("…is need to test of effect of treatment…")	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
598-600		See previous comment to lines 530-533	
606		Proposed change: "crystallised" to "dichotomised".	
614		This section is very interesting and helpful. It's a common belief that regulatory agencies will not accept retrospective validation of GBMs, while scientifically, there is no reason to not allow retrospective validation when the retrospective validation analysis is well conducted and the evidence is sufficiently compelling. While it's recognised that evaluation of retrospective validation is much more difficult and universal rules cannot be easily set, further clarification on this issue is much appreciated.	
615-617		Even when prospectively designed trials are feasible, a pharmaceutical company may be unwilling to do it due to the cost and the timeline of such studies, when there is sufficient retrospective data to provide compelling evidence of the effect of GBM. "in certain circumstance" is vague. It's sufficiently cautious to state that "the possibility could be considered". Any limitation of retrospective validation can be stated within the paragraph. If retrospective validation would only be considered under extreme circumstances, it's good to state it explicitly. Proposed change: When sufficient data is available from previously well conducted RCTs comparing relevant therapies in the relevant patient population, the possibility to test the predictive ability of a marker using such data could be considered.	
658-662		This summary still leaves doubt on when retrospective	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		validation would be accepted. It's also not clear how to document evidence that the analysis is fully pre-specified. The RCTs would have already been unblinded, but the GBM are not yet tested. Proposed change: add the following sentence to the paragraph: If a company wishes to use retrospective validation to gain approval for a GBM, early interaction with health authority is recommended to obtain guidance on such a strategy for the particular circumstance.	
661-662		Does this include the possibility of a data-split? Anyway, internal validation should be at least addressed!	
665 ff		The diagnostic performance section pre-supposes that a true diagnostic outcome exists, enabling sensitivity and specificity statistics. In some cases this might not be applicable. For instance if a sub-segment of a disease population is identified by a simple GBM or a multivariate GBM, this in itself may define the population. Sensitivity and specificity have no meaning here. Some discussion of such a scenario could be useful.	
667		"..disease characteristics…" – this could lead to small subgroups. How is the multiplicity issue to be addressed here? Performance estimates in small subgroups tend to be unreliable.	
684-719		Important: In addition to the GBM, other covariates may be important in predicting whether or not the event of interest occurs. This is particularly relevant when the event diagnosis is imperfect (eg abacavir hypersensitivity). In this instance,	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		adjusting for these other covariates in the statistical analysis allows removal of some of the noise introduced from the diagnosis imperfections (eg concurrent NNRTI or PI use for abacavir hypersensitivity). One limitation of sensitivity, specificity, NPV and PPV is that it is not possible to adjust for those additional covariates prior to calculating these values - which then limits their usefulness.	
695		Please define PPV and NPV.	
726		See previous comment to line 665+	
734		What is meant by "link" the specific test method and the value of the GBM? Please be more specific.	
743-747		Important: It would be extremely helpful for sponsors to provide some guidance/discussion on what "similar" concordance would be defined as. But this may be viewed as beyond the scope of the document.	
779-809		Please add definitions for sensitivity and specificity.	

Please add more rows if needed.