D2.1 Report describing minimally acceptable performance criteria for diagnostic testing

LITMUS

Liver Investigation: Testing Marker Utility in Steatohepatitis

Grant Agreement No. 777377

WP2 – Analysis, Evaluation & Evidence Synthesis

Lead contributor	Patrick Bossuyt (P2 – AMC)
	p.m.bossuyt@amc.nl
Other contributors	Quentin Anstee (P1– UNEW)
	Hadi Zafarmand (P2 – AMC)
	Morten Karsdal (P13 – NB)
	Elisabeth Erhardtsen (P13 – NB)
	Jérôme Boursier (P16 – UA)
	Kevin Duffin (P39 –LLY)
	Julia Brosnan (P40 – PFE)
	Craig Hyde (P40 – PFE)
	Tim Bauer (P42 – SOMA)
	Richard Torstenson (P43 – NOVO)
	Morten Hanson (P43 – NOVO)
	Zsolt Bocskei (P44 – SAN)
	Daniel Margerie (P44 – SAN)
	Sergio Kaiser (P46 – NOVA)
	Steven Lewitzky (P46 – NOVA)
	Jaison Jacob (P46 – NOVA)

	Guido Hanauer (P47 – TAK)

Due date	30 April 2018
Delivery date	19 July 2018
Deliverable type	R
Dissemination level	PU

Description of Work	**Version**	**Date**
	2.0	19 July 2018

Document History

Version	**Date**	**Description**
V1.0	09 Feb 2018	First Draft
V1.1	16 Feb 2018	Second Draft
V1.2	27 Feb 2018	Third Draft
V1.3	30 March 2018	Fourth Draft
V1.4	01 April 2018	Fifth Draft
V1.5	04 April 2018	Sixth Draft
V1.6	16 April 2018	Seventh Draft
V1.7	04 June 2018	Eighth Draft
V1.8	11 July 2018	Ninth Draft
V2.0	19 July 2018	Version Delivered

Table of Contents

1 Publishable Summary

This document describes the possible context of use of biomarkers in LITMUS. It also describes the minimally acceptable performance criteria (MAC) for a biomarker for each context of use (COU). The COU is a concise statement that describes the biomarker's specified use in drug development. For the choice of terminology in the COU/MAC document we adhered to the BEST resource ("Biomarkers, EndpointS, and other Tools"), to promote effective and unambiguous communications. These COU/MAC document will guide the synthesis of the available evidence, and the evaluation of the performance of biomarkers within LITMUS in phase 1a (with registry data) and phase 1b (with longitudinal data).

We defined six intended contexts of use:

1. Diagnostic Biomarker – Screening (NASH/Fibrosis)
2. Diagnostic Biomarker – Inclusion A: Diagnosing NASH
3. Diagnostic Biomarker – Inclusion B: Diagnosing Fibrosis
4. Prognostic Biomarker – Enrichment (NASH/Fibrosis)
5. Prognostic Biomarker – Liver Events
6. Monitoring/Pharmacodynamic biomarkers (NASH/Fibrosis)

This COU/MAC document is meant to be a "living" resource that will be periodically updated with additional terms and clarifying information. The COU/MAC document will be accessible via the LITMUS website.

2 Introduction

Non-alcoholic fatty liver disease (NAFLD) represents a spectrum of disease occurring in the absence of excessive alcohol consumption that ranges from isolated hepatic triglyceride accumulation (steatosis, NAFL) through hepatic triglyceride accumulation plus inflammation and hepatocyte injury (non-alcoholic steatohepatitis, NASH) and ultimately progresses to fibrosis/cirrhosis and potentially hepatocellular carcinoma (HCC).[1]

Global prevalence of NAFLD in general population is estimated at 24%; the highest rates are reported from the Middle East (32%) and South America (31%), followed by Asia (27%), the USA (24%) and Europe (23%), and the lowest rate is estimated in Africa (14%).[2,3] The prevalence of NASH in the general population remains unknown, however indirect methods derived by calculating prevalence of NASH in NAFLD and prevalence of NAFLD in the general population is estimated a NASH prevalence of 21% in NAFLD and 3-4% in the general population in the USA.[2,3]

Liver biopsy remains the established but imperfect reference standard for the above-mentioned

spectrum of the disease, but it is invasive, resource intensive, prone to sampling error and carries a small but significant risk of complications.[4]

Consequently, there is an urgent need for regulatory approved biomarkers in NASH patients to support patient management and facilitate the evaluation of new drugs. Measuring biomarkers in the LITMUS project presents an option for getting data for regulatory submission and approval of a biomarker in EU and in the US.

For the choice of terminology in this document we adhere to the BEST resource ("Biomarkers, EndpointS, and other Tools"), to promote effective and unambiguous communication among LITMUS project partners.[5]

Essential in the development of biomarkers is the timely identification of the context of use (COU). The COU is a concise statement that describes the biomarker's specified use in drug development. The COU includes two components: (1) the BEST biomarker category and (2) the biomarker's intended use in drug development (*Figure 1*).

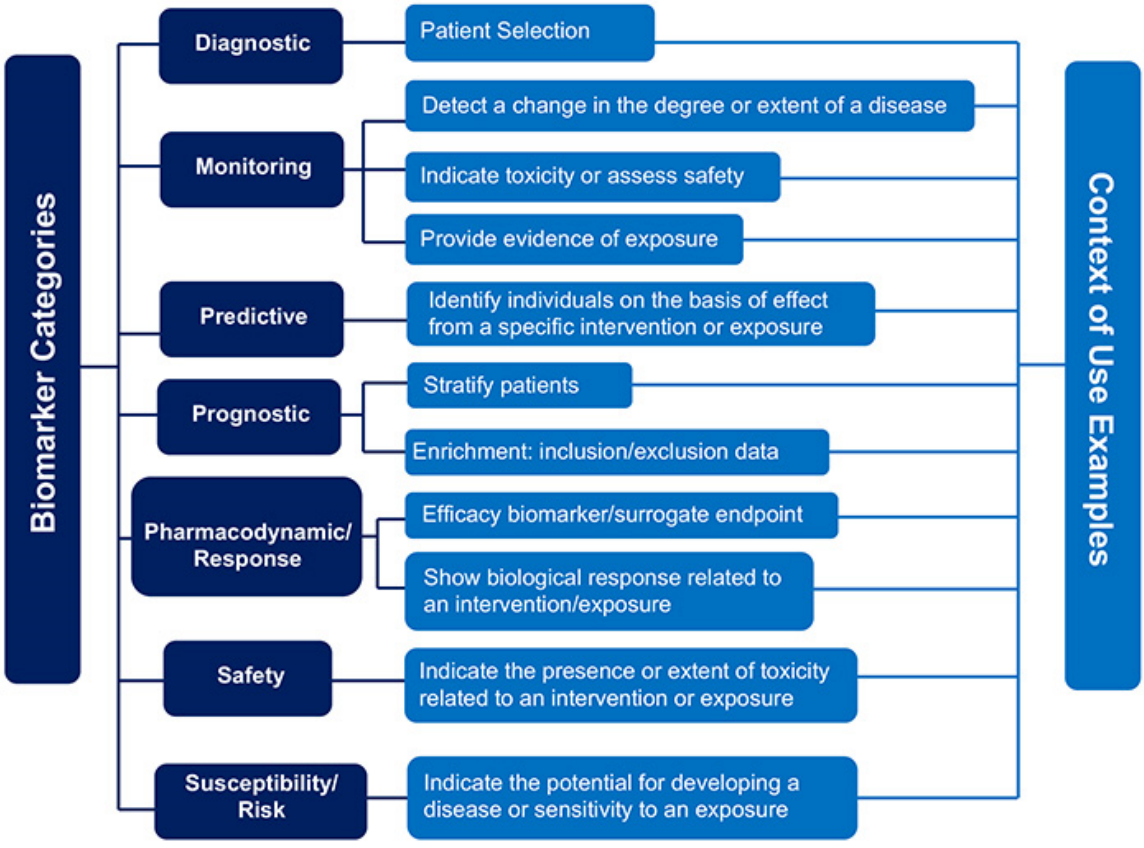

Figure 1: A graphical chart on context of use examples with respect to biomarker category[6]

Qualification of a biomarker means that the biomarker has been demonstrated to reliably support a specified manner of interpretation and application in drug development within a stated COU.

This document describes the possible context of use of biomarkers in LITMUS. It also describes the minimally acceptable performance criteria (MAC) for a biomarker for each context of use. These criteria will guide the synthesis of the available evidence and the evaluation of the performance of biomarkers within LITMUS. This MAC document is meant to be a "living" resource that will be periodically updated with additional terms and clarifying information. The MAC document will be accessible via the LITMUS website.

2.1 Liver Biopsy Considerations

Diagnosis NASH versus NAFL and/or grading and staging of hepatic fibrosis is dependent on a good quality liver specimen in which an adequate sized biopsy is taken. The quality criteria for liver biopsy as the reference standard are: a biopsy of at least 15-20mm in length and 18-gauge (16-Fr gauge preferred) in caliber, and the presence of at least 8 complete portal tracts. If cirrhosis is suspected, a cutting rather than a suction needle is preferred.[7]

3 Biomarkers in LITMUS

Below is an initial description of the context of use, the rationale, and the minimally acceptable performance criteria for each category of biomarkers.

3.1 Diagnostic Biomarker – Screening

3.1.1 Context of use

A test, measured at baseline, used as a supportive diagnostic tool to identify subjects at high risk of NASH (with or without fibrosis) or fibrosis (with or without NASH) for inclusion in clinical trials for drug development, after further confirmation of diagnosis with biopsy.

3.1.2 Rationale

Participants in ongoing NAFLD drug trials need to have biopsy proven NASH with a specific grade of fibrosis to qualify for enrolment. A non-invasive screening biomarker would identify patients at high risk of having NASH (with or without fibrosis) or fibrosis (with or without NASH). Planning liver biopsy

selectively in patients at high risk would therefore increase the efficiency of screening liver biopsies, by increasing the biopsy positivity rate (number of eligible patients in those undergoing biopsy).

3.1.3 Required Clinical evidence

A triaging biomarker will not replace liver biopsy, but will increase the efficiency of liver biopsies, by only planning liver biopsies in biomarker positive patients. It should therefore have sufficiently high negative and positive predictive value relative to histology in the intended use population.

Diagnostic accuracy compared against liver biopsy as the clinical reference standard.

3.1.4 Target conditions

- NASH (with or without fibrosis)
- Fibrosis
    - Capture the liver fibrosis granularity
      (i.e. a five-point scale and a seven-point scale)
    - Primary dichotomy: F0 – F2 versus F3 – F4

3.1.5 Minimally acceptable performance criteria

The performance (predictive values (PPV, NPV), and number needed to test (NNT)) should exceed that of current screening tools, such as FIB4 score and liver enzymes. For instance, for screening advanced fibrosis (F3-F4) in NAFLD patients[8], the new biomarker should have a positive predictive value of at least 75% and a negative predictive value of at least 85%. At a 19% prevalence[8‡], this would correspond to a positive likelihood ratio of 13 and a negative likelihood ratio of 0.8. This could be achieved by various combinations of sensitivity and specificity.

‡Based on the McPherson 2010 Study in 145 NAFLD patients, in which 63% had NASH and 19% had advanced fibrosis on liver biopsy.

3.2 Diagnostic Biomarker – Inclusion A: Diagnosing NASH

3.2.1 Context of use

A test, measured at baseline, used as a diagnostic tool to identify subjects with NASH for inclusion in clinical trials for drug development.

3.2.2 Rationale

Participants in ongoing NAFLD drug trials need to have biopsy proven NASH (with or without fibrosis) to qualify for enrolment. A non-invasive biomarker that could replace liver biopsy in evaluating potentially eligible patients would greatly facilitate recruitment for future trials.

3.2.3 Required Clinical evidence

A biomarker to replace liver biopsy should not be inferior to liver biopsy considering the level of information yield, risks and benefits. It should therefore have high agreement with histology (i.e. diagnostic accuracy) considering the expected benefit threshold over the liver biopsy for the required target condition(s).

Given the known limitations of liver biopsy (imperfect sensitivity and specificity) there is an upper limit on the diagnostic accuracy that can be achieved for diagnostic biomarkers for NASH.

3.2.4 Target conditions

- Discriminating NAFL (steatosis) vs. NASH
- Assessing grade (activity) of NASH.

3.2.5 Minimally acceptable performance criteria

Non-inferior sensitivity and specificity relative to histology considering the expected benefit threshold; the biomarker should have a sensitivity and specificity exceeding 80%.

3.3 Diagnostic Biomarker – Inclusion B: Diagnosing Fibrosis

3.3.1 Context of use

A test, measured at baseline, used as a diagnostic tool to identify subjects with fibrosis for inclusion in clinical trials for drug development.

3.3.2 Rationale

Participants in ongoing NAFLD drug trials need to have biopsy proven fibrosis (with or without NASH) to qualify for enrolment. A non-invasive biomarker that could replace liver biopsy in evaluating potentially eligible patients would greatly facilitate recruitment for future trials.

3.3.3 Required Clinical evidence

A biomarker to replace liver biopsy should not be inferior to liver biopsy considering the level of information yield, risks and benefits. It should therefore have high agreement with histology (i.e. diagnostic accuracy) considering the expected benefit threshold over the liver biopsy for the required target condition(s): fibrosis.

Given the known limitations of liver biopsy (imperfect sensitivity and specificity) there is an upper limit on the diagnostic accuracy that can be achieved for diagnostic biomarkers for fibrosis.

3.3.4 Target conditions

- Capture the liver fibrosis granularity
  (i.e. a five-point scale and a seven-point scale)
- Primary dichotomy: F0 – F2 versus F3 – F4

3.3.5 Minimally acceptable performance criteria

Non-inferior sensitivity and specificity relative to histology considering the expected benefit threshold. The new biomarker should have a sensitivity and specificity exceeding 80%.

3.4 Prognostic Biomarker – Enrichment

3.4.1 Context of use

A prognostic biomarker, measured at baseline, used to identify subjects more likely to develop substantial and/or timely worsening in their condition, which would create a clinical trial patient population more likely to benefit from treatment.

3.4.2 Rationale

By including patients more likely to experience subsequent worsening of their condition, enrichment strategies can increase the absolute magnitude effect of treatment (if any), which can lead to increased precision in the treatment effect estimate, smaller study groups, or both.

3.4.3 Required Clinical evidence

A biomarker for prognostic enrichment should be associated with the outcome of interest in measurement-based clinical treatment trials.

3.4.4 Target conditions

Future worsening in NAFLD activity score (NAS) or the Activity component of the "Steatosis, Activity, and Fibrosis" (SAF) score, transition from NAFL to NASH, and/or progression of fibrosis stage.

3.4.5 Minimally acceptable performance criteria

The performance will be evaluated within the typical time windows for treatment trials.
We will dichotomize the outcome measures and define the minimally acceptable criteria for sensitivity and specificity, using Receiver Operating Characteristic (ROC) curves to express the accuracy for changes in the positivity cut-off for continuous markers.

3.5 Prognostic Biomarker – Liver Events

3.5.1 Context of use

A prognostic biomarker used to identify subjects likely to develop the liver event of interest, e.g. hepatic decompensation, variceal haemorrhage, development of HCC (and/or liver-related/all-cause mortality) in the future.

3.5.2 Rationale

At present, fibrosis of the liver is the best predictor of patient-important liver-related events. Using a biomarker that is a stronger predictor of liver-related events as an outcome measure could lead to treatment trials that are more feasible and more valid than biopsy-based trials. Replacing histology as the outcome measure with biomarkers that are at least as strongly associated with longer-term clinical outcomes (and sensitive to meaningful changes, see monitoring) would replace the need for repeated biopsies while providing sound evidence about the effectiveness of treatment.

3.5.3 Required Clinical evidence

A prognostic biomarker for liver-related events should be non-inferior to liver biopsy (considering the expected benefit threshold) in predicting liver-related events.

3.5.4 Target conditions

Primary endpoint events include death (liver-related or all-cause); model of end stage liver disease (MELD) score ≥15; liver transplant; hepatocellular carcinoma (HCC); large oesophageal/gastric varices; ascites; increase in Hepatic Venous Pressure Gradient (HVPG) >10 mmHg; histological progression to

cirrhosis; hospitalization (as defined by a stay of ≥24 hours) for onset of: variceal bleed, hepatic encephalopathy, spontaneous bacterial peritonitis.

Time-window will extend several years after measurement of the prognostic biomarker.

3.5.5 Minimally acceptable performance criteria

A measure of association that depends on the study design (such as Cox proportional hazard or logistic regression models) reporting the relative hazard or odds ratio for liver-related events per unit or standard deviation of change in the biomarker should be compared against the comparable evidence for fibrosis on liver biopsy.

3.6 Monitoring/Pharmacodynamic biomarkers

3.6.1 Context of use

A biomarker measured serially and used to detect a change in the degree or extent of disease.

3.6.2 Rationale

Trials in NAFLD/NASH currently rely on repeated liver biopsies study to evaluate the effectiveness of treatment. A monitoring/pharmacodynamic biomarker that is also diagnostic (corresponds to liver biopsy) and/or prognostic (predicts liver-related events), when measured at baseline and after a treatment period, could be used as an alternative outcome measure in trials of the effectiveness of treatment.

3.6.3 Required Clinical evidence

A monitoring/pharmacodynamic biomarker would have to be sensitive to short term changes.
In addition: see evidence required for diagnostic biomarker and prognostic biomarker (liver-related events).

3.6.4 Target conditions

Changes in the biomarker value would indicate a recent or an impending change in the individual's condition of NASH/Fibrosis, or (pharmacodynamic) a beneficial or adverse effect of a drug or an intervention.

3.6.5 Minimally acceptable performance criteria

Sensitivity to meaningful changes in histological and/or clinical status.

See also criteria for diagnostic biomarkers and for prognostic biomarkers (liver-related events).

4 Summary

In Table 1, a summary of the possible context of use of biomarkers in LITMUS, the associated target condition(s), the required study design to provide the required clinical evidence, and the reference standard test is provided.

Table 1: The possible context of use of biomarkers in LITMUS and its related target condition(s), required evidence, and reference standard

Category of Biomarkers	Context of Use	Target	Required Evidence	Reference Standard
Diagnostic	Screening (NASH/Fibrosis)	NASH (with or without fibrosis)/ Fibrosis	Cross-sectional	Liver biopsy
Diagnostic	Inclusion (diagnosing NASH)	NASH	Cross-sectional	Liver biopsy
Diagnostic	Inclusion (diagnosing Fibrosis)	Fibrosis	Cross-sectional	Liver biopsy
Prognostic	Enrichment (NASH/Fibrosis)	Worsening in NAS or SAF score, NASH, or fibrosis	Longitudinal	Serial biopsy
Prognostic	Enrichment (Liver Events)	Liver-related events	Longitudinal	Events
Monitoring/ Pharmacodynamic	Monitoring/Response (NASH/Fibrosis)	Meaningful changes	Longitudinal	Changes

5 Notes

In the EU, the qualification of the biomarker must be at the level required in the CE marking. However, the current European regulations on medical devices and on in vitro diagnostic medical devices (IVD) will be replaced in year 2020 and 2022, respectively, and the content is not yet fully defined.[9] In the US, a new biomarker must be evaluated according to the Global Laboratory Standards (Clinical and Laboratory Standards Institute, CLSI) guidelines, and it is expected that the new regulation in EU will be along the same lines. Because of these changes, biomarker qualification should be according to CLSI. For Magnetic Resonance Diagnostic Devices, the devices should be evaluated according to the Guiding Document issued by FDA on November 18, 2016.[10] This will also ensure that biomarker qualification will be valid in both the EU and the US.

The purpose of analytical method validation is to demonstrate that a specific method for quantitative measurement of an analyte (in this case a biomarker) is reliable and reproducible for its COU. A full validation is important when developing and implementing a bioanalytical method for the first time. A partial validation is a modification of already validated bioanalytical methods and is appropriate for bioanalytical method transfers between two laboratories, a change in the biological matrix of interest (e.g. plasma to urine). The key parameters for analytical validations are: (analytical) accuracy, precision, (analytical) sensitivity, reproducibility and stability.[11]

Clinical evidence supporting the use of a biomarker is most often reviewed in two stages. In stage one (Phase 1a, in the case of LITMUS), evaluations are performed in smaller clinical studies to identify/select biomarkers, while in stage two (LITMUS Phase 1b) the selected biomarkers are confirmed in independent clinical studies which are powered to evaluate the performance of the biomarker in the target population. In this context clinical validation is not a synonym of demonstration of clinical utility. However, for each COU in LITMUS the specific processes of clinical evaluations will be provided later.

6 Discussion

As a collaborative work between WP2, WP7 and LITMUS project leads, we prepared a document describes the possible context of use of biomarkers in LITMUS. It also describes the minimally acceptable performance criteria for a biomarker for each context of use.

Due to importance of this document for evidence synthesis, evaluating performance of biomarkers in phase 1a (with registry data) and phase 1b (with longitudinal data), we organised a face-to-facet input

meeting hosted by Sanofi in Paris on 10th April 2018 to discuss a more aligned version of the document (version 1.5) to get input from all WP2/WP7 partners.

The document (ver. 1.7) was also discussed during the LITMUS Steering Committee meeting in London on 8th -9th May 2018. The COU/MAC document was finalized after final approval by the LITMUS Project Executive in the TC on 11th July 2018 and by the Project leads.

As the document needed to be aligned with WP2, WP7, and the Project leads and had to be developed in agreement with requirements of the qualifying agencies in both the EU and the US, the planned due date on 30th April 2018 was achieved on 19th July 2018.

7 Conclusions

The performance of biomarkers in LITMUS has to be evaluated against well-defined criteria for acceptability. These criteria include a definition of the context of use, the appropriate target condition(s), acceptable reference standard, required clinical evidence, and minimum levels of accuracy measurements.

As summarised in Table 1, we have defined six different contexts of use in the three categories of biomarkers (namely, diagnostic, prognostic, and monitoring/pharmacodynamic biomarkers) in the course of the LITMUS project. These contexts of use include screening (NASH/Fibrosis), inclusion A: diagnosing NASH, inclusion B: diagnosing liver fibrosis, enrichment (NASH/Fibrosis), enrichment (Liver Events), and Monitoring/Response (NASH/Fibrosis).

These performance criteria will be used in the synthesis of the available evidence, in the analysis of the available data in Phase 1a and in the analysis of the data collected in the prospective recruitment of patients with histologically characterised NAFLD in Phase 1b. Markers unlikely to meet these minimum criteria may not advance to regulatory qualification.

The context of use/minimally acceptable performance criteria document is meant to be a “living” resource, further modified, if needed, in the course of the LITMUS project and will be periodically updated with additional terms and clarifying information.

8 References

1. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol 2013;10:330-44.
2. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73-84.
3. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018;15:11-20.
4. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128:1898-906.
5. BEST (Biomarkers, EndpointS, and other Tools) Resource, FDA-NIH Biomarker Working Group. (Accessed March 14, 2018, at https://www.ncbi.nlm.nih.gov/books/NBK326791/pdf/Bookshelf_NBK326791.pdf.)
6. Biomarker Qualification Program Submission, U.S. Food and Drug Administration. (Accessed February 27, 2018, at https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/ucm535398.htm#2.)
7. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67:328-57.
8. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. Gut 2010;59:1265-9.
9. Regulatory framework, European Commission, DG Internal Market, Industry, Entrepreneurship and SMEs. (Accessed March 14, 2018, at https://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework_en.)
10. Guidance for Industry and Food and Drug Administration Staff: Guidance for the Submission of Premarket Notifications for Magnetic Resonance Diagnostic Devices. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health. (Accessed March 14, 2018, at https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM454613.pd0f.)
11. Guidance for Industry: Guidance for the Bioanalytical Method Validation, U.S. Food and Drug Administration. U.S. Department of Health and Human Services, Food and Drug Administration,

Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM). (Accessed March 14, 2018, at https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm368107.pdf.)