DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Considerations from the IQ Induction Working Group in Response to Drug-Drug Interaction Guidances from Regulatory Agencies: Focus on CYP3A4 mRNA *in vitro* response thresholds, variability, and clinical relevance

Jane R. Kenny, Diane Ramsden, David B. Buckley, Shannon Dallas, Conrad Fung, Michael Mohutsky, Heidi J. Einolf, Liangfu Chen, Joshua G. Dekeyser, Maria Fitzgerald, Theunis C. Goosen, Y. Amy Siu, Robert L. Walsky, George Zhang, Donald Tweedie and Niresh Hariparsad.

Genentech, South San Francisco, California (JRK); Boehringer Ingelheim, Ridgefield, Connecticut (DR); Sekisui-XenoTech LLC, Kansas City, Kansas (DBB); Janssen R&D, Spring House, Pennsylvania (SD); Vertex Pharmaceuticals, Boston, Massachusetts (CF, NH); Eli Lilly and Company, Indianapolis, Indiana (MM); Novartis, East Hanover, New Jersey (HJE); GlaxoSmithKline, King of Prussia, Pennsylvania (LC); Amgen Inc., Cambridge, Massachusetts (JGD); Sanofi, Waltham, Massachusetts (MF); Pfizer Global Research and Development, Groton, Connecticut (TCG); Eisai, Andover, Massachusetts (YAS); EMD Serono R&D Institute, Inc., Billerica, Massachusetts (RLW); Corning Life Sciences, Woburn, Massachusetts (GZ); Merck & Co., Inc., Kenilworth, New Jersey (DT)

Running title:

IQ considerations on regulatory guidances for induction: part 2

Corresponding Author:

Jane R. Kenny, PhD.

Department of Drug Metabolism and Pharmacokinetics, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA.

Phone + 1 650 467 6027

email: kenny.jane@gene.com

Number of Text Pages: 33

Number of Tables: 4

Number of Figures: 8

Number of References: 38

Number of words in Abstract: 250

Number of words in Introduction: 822

Number of words in Discussion: 2099

Number of Supplementary Text pages: 2

Number of Supplementary Tables: 9

Number of Supplementary Figures: 5

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Abbreviations:

AUC, area under the curve; AUCR, area under the curve ratio; CAR, constitutive androstane receptor; C_{av}, average concentration; $C_{av,ss}$, average concentration at steady state; C_{max}, maximum concentration; $C_{max,ss}$, maximum steady state concentration; $C_{max,ss,u}$, unbound maximum steady state concentration; Cmpd, compound; CRO, contract research organization: Ct, cycle time; ΔCt, delta cycle time is the change in Ct for gene of interest relative to housekeeping gene; ΔΔCt, delta delta cycle time is the change in ΔCt for test compound relative to vehicle control (i.e. fold induction); CYP, cytochrome P450; DDI(s), drug-drug interaction(s); DME, drug metabolizing enzymes; DMLG, drug metabolism leadership group; DMSO, dimethylsulfoxide; $EC_{10,}$ concentration achieving 10% of maximal induction; EC_{50}, concentration that supports 50% of maximum response; EMA, European medicines agency; E_{max}, maximum fold increase (or induction) minus baseline of 1-fold; F2, the concentration achieving 2-fold induction; FDA, food and drug administration; fmCYP, fraction metabolized by cytochrome P450; f_{up}, fraction unbound in plasma; GAPDH, glyceraldehyde 3 phosphate dehydrogenase; Ind_{max}, maximal fold induction; IVIVE, in vitro in vivo extrapolation; IQ, innovation and quality consortium; IWG, Induction Working Group; LC-MS/MS, liquid chromatography tandem mass spectrometry; LoB, limit of blank; LoD, limit of detection; NME, new molecular entity; PCR, polymerase chain reaction; PPB, plasma protein binding; PMDA, pharmaceutical and medical devices agency; PK, pharmacokinetics; PXR, pregnane-X receptor; RT-PCR, reverse transcription polymerase chain reaction; t_{max}, time after dosing maximal concentration is reached; QD, one dose per day; RIS, relative induction score; SD, standard deviation; TDI, time dependent inhibition; UWDIDB, University of Washington drug interaction database; %CV, percent coefficient of variation.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

ABSTRACT

The IQ induction working group presents an assessment of best practice for data interpretation of *in vitro* induction, specifically, response thresholds, variability, application of controls and translation to clinical risk assessment with focus on CYP3A4 mRNA. Single concentration control data and E_{max}/EC_{50} data for prototypical CYP3A4 inducers were compiled from many human hepatocyte donors in different laboratories. Clinical CYP3A induction and *in vitro* data were gathered for 51 compounds, 16 of which were proprietary. A large degree of variability was observed in both the clinical and *in vitro* induction responses, yet analysis confirmed *in vitro* data are able to predict clinical induction risk. Following extensive examination of this large dataset, the following recommendations are proposed. (a) CYP induction should continue to be evaluated in three separate human donors *in vitro*. (b) In light of empirically divergent responses in rifampicin control and most test inducers, normalization of data to percent positive control appears to be of limited benefit. (c) Two-fold induction, with concentration dependence, is an acceptable threshold for positive identification of *in vitro* CYP3A4 mRNA induction. (d) To reduce the risk of false positives, in the absence of a concentration dependent response, induction ≥ 2-fold should be observed in more than one donor to classify a compound as an *in vitro* inducer. (e) If qualifying a compound as negative for CYP3A4 mRNA induction, the magnitude of maximal rifampicin response in that donor should be ≥10-fold. (f) Inclusion of a negative control adds no value beyond that of the vehicle control.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

INTRODUCTION

Regulatory agencies have issued guidelines and guidances for the conduct of drug-drug interaction (DDI) studies with specific sections focusing on human cytochrome P450 (CYP) induction. The European Medical Agencies (EMA) 2012 guideline (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/07/WC500129606.pdf), the Pharmaceutical and Medical Devices Agency (PMDA) 2014 guidance (Drug Interaction Guideline for Drug Development and Labeling Recommendations (MHLW, 2014), updated 2017, English translation not yet available from PMDA) and the Food and Drug Administration (FDA) 2017 draft guidance (http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm292362.pdf) specify that *in vitro* CYP induction assessment be conducted in human hepatocytes from three different donors using mRNA as the primary endpoint. All three agencies consider a 2-fold increase in mRNA the threshold for a positive *in vitro* induction signal. The EMA and PMDA also specify that this increase must be concentration-dependent. The FDA states that a ≥2-fold increase and a response of ≥20% of positive control are interpreted as a positive finding. The EMA and PMDA state that an *in vitro* induction response of <100% (i.e. <2-fold) is only negative if it is also <20% of the positive control response. The agencies agree that evaluation should adequately explore clinically relevant drug concentrations for the maximum therapeutic dose, although the exact definition differs. EMA calls for 50-fold the mean steady state unbound C_{max} for hepatic and 0.1 x dose/250 mL for intestinal induction assessment. The PMDA requests at least 10-fold steady state unbound C_{max}. The FDA asks that, if solubility allows, at least one concentration should be an order of magnitude greater than unbound steady state C_{max}, with the caveat that, if protein binding is >99%, the fraction unbound in plasma ($f_{u,p}$) be capped at 0.01. All three agencies agree that the *in vitro* donor providing the most sensitive, "worst-case" positive response, be used to determine the clinical induction risk.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Once an *in vitro* induction assessment has been deemed positive, the agencies provide recommendations for subsequent assessment of whether a clinical DDI study is warranted. This step involves the use of mathematical models to predict the DDI risk based on the relevant clinical concentration and *in vitro* E_{max} and EC_{50} values. Risk assessment falls into three general categories: 1) basic models or R values; 2) correlation methods, where extensive *in vitro* calibration is performed (Fahmi and Ripp, 2010); or 3) mechanistic models that use either static or dynamic concentrations of inducer to predict AUCR. The latter two approaches use the clinical definitions of bioequivalence for DDI to flag induction risk, namely a victim drug AUCR of 0.8 or less. The simplest calculation or R value approach (see equation in Table 1A), is recommended as a first step by the FDA and PMDA but not the EMA, where F2 is considered the basic method (Table 1B). Interestingly, the 2017 FDA draft guidance added a 10-fold multiplier to unbound drug concentration and changed the threshold from R<0.9 to R<0.8 as a trigger for further evaluation of DDI risk (Table 1C). Common to all three agency recommendations are the static mechanistic model (Einolf, 2007; Einolf et al., 2014; Vieira et al., 2014) that considers induction at both the hepatic and intestinal level (for CYP3A inducers) in relation to the fraction of victim drug that is metabolized by a specific CYP (fmCYP) (Table 1D) and a correlation method, the Relative Induction Score (RIS) (Fahmi and Ripp, 2010) (Table 1E) that relies on calibration to known clinical inducers in that human hepatocyte donor. Notably, the FDA and PMDA (but not the EMA) guidances include an option of dynamic mechanistic assessment, such as PBPK, for induction DDI. Finally, when a test compound has both *in vitro* CYP induction and inhibition (either reversible or time-dependent), both the FDA and EMA caution against risk assessment of induction and inhibition in a combined approach.

The International Consortium of Innovation and Quality in Pharmaceutical Development (IQ) Induction Working Group (IWG) recently highlighted several areas of regulatory recommendations that would benefit from further evaluation (Hariparsad et al., 2017).

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Recommendations from the IWG were provided on the evaluation of down-regulation*, in vitro* assessment of CYP2C induction and the use of CITCO as a positive control for CYP2B6. Two other areas were highlighted by the IWG for further evaluation, namely, *in vitro* data interpretation, and induction time course. This manuscript focuses on data interpretation; specifically, what constitutes a positive *in vitro* induction signal and how to assess whether this induction signal is clinically relevant.

IQ member companies shared blinded clinical induction data for proprietary compounds along with the corresponding *in vitro* data. The literature reports of clinical induction are dominated by CYP3A, with very few examples of CYP1A2 (Gabriel et al., 2016) and CYP2B6 (Fahmi et al., 2016). The dataset gathered reflected this and all data, with the exception of one clinically relevant CYP1A2 DDI, were for CYP3A4. Therefore, the following evaluation of *in vitro* CYP induction data interpretation, namely response thresholds, variability, application of controls and translation to clinical risk assessment, and the subsequent recommendations are focused on induction of CYP3A4.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

MATERIALS & METHODS

Proprietary inducer data from within IQ member companies.

To allow for an assessment of induction by proprietary compounds from IQ consortium member companies, a template (https://iqconsortium.org/initiatives/working-groups/induction/) was developed to collate the necessary data and supplementary information.

The survey was distributed by the IQ Secretariat to representatives of IQ Consortium member companies. It was stipulated that responses should be reflective of the company as only one response was permitted from each company. Surveys were returned to the IQ Secretariat who then blinded the data as unnamed Company and compound, for example "Company A compound 1". This was then streamlined to compound (Cmpd) for Cmpd1 through Cmpd16. Compound identity was further blinded by requiring both *in vitro* and *in vivo* data in molar concentrations and withholding the molecular weight. Companies were asked to provide regulatory quality data rather than discovery screening data and, where available, to include data for positive and negative controls that were run in the same assay as the test compound. The template was built to be relatively exhaustive and to collect the majority of the data generated in an *in vitro* induction study. As with any survey, limitations do exist, including the expectation that all information requested in the template would not be provided by every company (Hariparsad et al., 2017). Different assay designs, and especially data from studies before the 2012 EMA and FDA regulatory guidances, would often result in less comprehensive data sets. Companies were also asked to provide any evidence of time-dependent inhibition and/or auto-induction, *in vitro* and *in vivo.*

In vitro parameters collected included time of incubation, cellular overlay (i.e. matrigel), plate layout (e.g. 96-well), media used, supplements added, any additional protein in the media, any viability method and viability cut-off values for cytotoxicity assessment, housekeeping gene

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

used, method of mRNA analysis, probe substrates for CYP activity, enzyme(s) involved in the compounds metabolism and estimation of $f_{m,CYP}$ (fraction of dose eliminated by a specific CYP).

Clinical data requested included C_{max}, C_{av} and AUC, at both single dose and multiple doses of the proprietary compound, along with blood to plasma ratio and fraction unbound in plasma. For the DDI study, companies provided the identity of the probe drug, dosing regimen, AUC, C_{max}, and t_{max}, pre- and post-administration of the potential inducer to steady state.

Prototypical inducer data from literature.

In vivo DDI data used for this analysis were also gathered from the University of Washington drug interaction database (UWDIDB; www.druginteractioninfo.org). The objects (hereafter called victim drugs) included in this assessment were those recommended by the FDA (https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm#table3-1). In addition to collecting the CYP3A clinical induction studies by considering the substrates recommended by regulatory agencies (designated as CYP3A sensitive), a second-tier data collection was employed. Here the focus was to collect all positive and negative clinical induction studies for the perpetrators in order to build knowledge around the thresholds for true *in vitro* and *in vivo* negatives. When CYP3A was determined to contribute to the overall metabolism of the victim drug, the clinical study was included as part of the "all data" or complete analysis. Additionally, to account for perpetrators which exhibited both *in vitro* induction and inhibition mechanisms (reversible or time-dependent), positive and negative clinical inhibition studies were also collected from the UWDIDB and sorted in the same manner as described for the clinical induction studies. A minimum of five-days repeat dosing was selected as the threshold to include clinical studies, since this would likely establish steady state conditions, accounting for the half-life of both the clinical inducer and CYP3A enzyme (reported to be 23-87 h) (Ramsden et al., 2015). The clinical dataset collected for rifampicin

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

was limited to a dose level of 600 mg QD, which is the therapeutically relevant dose resulting in maximal in vivo induction (Kozawa et al., 2009). Additionally, the dose level for ritonavir was restricted to >100 mg QD to reflect both its clinical use as a boosting agent and earlier therapeutic doses (Ruane et al., 2007). Clinical induction data were collected for compounds with existing *in vitro* data made available from member companies and focused on identification of compounds with mild or no clinical induction. Therefore, not all clinically relevant inducers are captured within this dataset (e.g. modafanil, avasimibe).

Median as well as worst-case clinical AUCR values were used to evaluate the ability of the *in vitro* parameters to predict the observed clinical effect. (The median is preferable to the mean in representing the center of a population because it is less susceptible to bias when non-normality or outliers are present.) In the case of the *in vitro* parameters, both the worst-case donor and median induction parameters were used for modeling purposes. Using the complete set of *in vitro* data to fit a 3-parameter sigmoidal dose response model (a common Hill function model used in pharmacology, Table 1F), correlation approaches were established using the slope and RIS. The RIS model was used as described (Fahmi and Ripp, 2010) by fitting the data using the unbound $C_{max,ss}$ of inducers to generate a curve against known clinical induction response and then inputting the unbound $C_{max,ss}$ of test compounds in order to predict the % change in AUC. The estimated portal concentration in the RIS model was also applied, as recommended in the EMA guideline. In the case of literature compounds, the gut concentration was estimated for evaluation of the F2 value (Table 1B) and for inclusion into the mechanistic static models. The mechanistic static model was evaluated with input concentrations by using the estimated portal concentration and the estimated gut concentration, as recommended by regulatory guidances. In addition, the unbound $C_{max,ss}$ was used for the hepatic portion and the calculated hepatic portal concentration was used as the input for the gut portion. The concentration resulting in 2-fold induction (F2) was used, as described in the EMA guideline, by considering

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

30- and 50-fold unbound $C_{max,ss}$ as the inducer concentration. The R3 model, as described in the FDA DDI guidance from 2012, was evaluated using multiple approaches; total and unbound $C_{max,ss}$ with a cut-off value of 0.9 and a d-value of 1 (R3 = 0.9 (total and unbound)); total $C_{max,ss}$ and a cut-off value of 0.8 (R3 = 0.8, d =1); gut concentration as the input (gut), cut-off of 0.95 and the unbound $C_{max,ss}$ (R3 = 0.95); applying a universal scaling factor value of 0.3 determined from empirical fitting of the full dataset to varying d values with the goal of increasing the quantitative accuracy (R3 = 0.9, d = 0.3, with $C_{max,ss}$ total as input); slope value with the total and unbound $C_{max,ss}$ as inputs (R3 = 0.9, slope (total), R3 = 0.9, slope (unbound)); the average unbound or total concentration (average unbound, average total); and lastly, limiting the maximum plasma protein binding (PPB) to 1% (f_{up} > 0.01). In addition, the recommended approach in the draft FDA and PMDA DDI guidance documents from 2017, was evaluated by using the R3 equation as described with a 10-fold multiplier for inducer concentration. Additionally, a 50-fold multiplier for inducer concentration was used to explore the impact on the number of false negative induction DDI predictions.

Culture of cryopreserved human hepatocytes for induction.

The *in vitro* data presented encompasses data from member companies for proprietary and well-known or prototypical inducer compounds, data from literature and data generated by the IWG. Different conditions were employed by laboratories (Hariparsad et al., 2017) which reflect general protocols for generating *in vitro* induction data. Various lots of human cryopreserved hepatocytes, from both males and females of different ages and racial origin, were obtained from several commercial vendors; CellzDirect (Durham, NC), Bioreclamation In Vitro Technologies (Baltimore, MD), Corning Life Sciences (Woburn, MA) and XenoTech LLC, (Kansas City, KS). As detailed in previous publications (Fahmi et al., 2010; Sane et al., 2015), cryopreserved human hepatocytes were thawed in hepatocyte thawing medium and were seeded in collagen I coated 24- or 96-well plates at cell densities of 0.5-1 × 10^6 viable cells per

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

well in hepatocyte plating medium. Viability, as determined by trypan blue exclusion or other methods, was 85% or better when cells were plated. The cells were initially maintained overnight at 37°C in a humidified incubator, with 95% atmospheric air and 5% CO_2, in hepatocyte incubation media. Following overnight incubation, the cells were either treated with compounds or were overlaid with matrigel to form sandwich cultures, maintained for an additional 24 hr under incubated settings and then treated with compounds. Compounds were dissolved in DMSO and added to the culture medium at various concentrations (final DMSO concentration, 0.1% or 0.5%). After daily treatment for 2-3 days, the medium was removed, and the cells were washed with PBS. The cells were lysed in lysis buffer and prepared for RNA isolation. Cell viability was assessed by visual inspection of the monolayer, checking for confluency and morphology. Different companies used different plating conditions and a representation of the conditions is shown in Supplementary Table 1.

mRNA Preparation and Analysis.

Following the isolation of RNA with commercially available kits, cDNA was synthesized using standard PCR protocols. Designated CYP enzymes and an endogenous probe (e.g. GAPDH) mRNA levels were quantified by real time PCR. The gene-specific primer/probe sets were typically obtained from Applied Biosystems Incorporated (Foster City, CA). The relative quantity of the target cDNA compared with that of the house keeping gene, was determined by the $\Delta\Delta$Ct method (Livak and Schmittgen, 2001). This relative quantification measures the change in mRNA expression in a test sample, relative to that in a vehicle control sample (final DMSO concentration, 0.1% or 0.5%). In order to reduce variability, Ct values >32 were excluded from the analysis, since this is indicative of low expression.

CYP3A Enzyme Activity.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Midazolam 1'-hydroxylase or testosterone 6β-hydroxylase activities were measured *in situ* with methods similar to those described by (Zhang et al., 2010). Briefly, following the treatment period cell culture medium was removed, hepatocytes were rinsed, and marker substrate reactions were started by the addition of either midazolam (30 µM) or testosterone (200 µM). Following a 30 minute incubation at 37°C, marker substrate reactions were stopped by removal of an aliquot from each well and combining with acetonitrile containing internal standard (deuterated metabolite). Metabolite formation was quantified by LC-MS/MS.

***In vitro* human hepatocyte induction assay for clinically weak inducers.**

In vitro induction data for clinically weak inducers (defined as eliciting a clinical AUCR of 0.5 to 0.8 for a victim drug) were available for most compounds from literature resources or IWG member companies. *In vitro* induction parameters were generated for felbamate, rufinamide, oxcarbazepine, flucloxacillin and lersivirine, using four human hepatocyte donors in four labs, since no published or IWG derived values were available. The human hepatocyte donors were obtained from different commercial vendors; including Triangle Research Laboratories (Durham, NC), Bioreclamation In Vitro Technologies (Baltimore, MD), Corning Life Sciences (Woburn, MS) and XenoTech LLC, (Kansas City, KS). The tested compounds were purchased from Sigma-Aldrich (St. Louis, MO) or MedChem Express (Monmouth Junction, NJ). The member companies followed their internal induction protocols to generate the data. Two companies used sandwich cultured hepatocytes and two used monolayer cultured hepatocytes. Top test concentrations were selected to cover the estimated gut exposure (0.1 x Dose/250 mL) and 50-fold the unbound $C_{max,ss}$, with consideration of solubility and cytotoxicity limits. Compounds were dissolved in DMSO and added to the culture medium at seven or eight concentrations (final DMSO concentration, 0.1% or 0.5%).

***In vitro* reversible and time-dependent CYP inhibition for prototypical inducers.**

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Using the UWDIDB, a literature review was conducted to evaluate whether the *in vitro* inducers were also *in vitro* reversible or time-dependent inhibitors. In cases where inhibition parameters were available from literature, the data were scrutinized to ensure that the methodology for deriving the parameters was sound. Where information on the inhibition potential was not available, the inhibition potential was evaluated by the IWG and used to determine whether mixed mechanisms of DDI (inhibition and induction) could impact the IVIVE (see Supplementary Text for methods).

Analysis of basal enzyme levels and single point data of vehicle, negative and positive controls.

Member companies were invited to submit historical *in vitro* induction datasets obtained from multiple repeated experiments, with single concentration negative and positive control inducers. Given the limited application of negative controls across participating labs, flumazenil was selected for further evaluation as a negative control. An additional consideration was the availability of *in vitro* CYP3A datasets with sufficient size to perform statistical analysis. Specifically, statistical analysis of intra-donor variability was performed on CYP3A4 mRNA from flumazenil treated hepatocyte donors where there existed a minimum of 20 repeated experiments. Based on this selection criterion, subsequent data analysis was performed on 10 individual hepatocyte donors from a single participating laboratory.

For datasets with positive control inducers using single concentrations, data analysis was performed on 15 individual hepatocyte donors, from two participating laboratories, where a minimum of 10 repeated experiments were available. Both CYP3A4 mRNA and CYP3A enzyme activity were analyzed.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

The intra-donor variation in rifampicin fold induction response was further interrogated in three hepatocyte donors, namely H2, H4 and H12, which were selected based on variability observed in rifampicin-CYP3A4 mRNA response to be representative of low, mid and high intra-donor variability with large sample sizes. Where available, additional gene expression (RT-PCR) data for CYP3A4 and the relevant housekeeping gene (18S or GAPDH) from the vehicle control (DMSO), positive control (rifampicin) and negative control (flumazenil) treatment groups were also analyzed. These datasets included cycle time (Ct), delta cycle time (ΔCt) and fold induction (delta delta cycle time (ΔΔCt) values. Similarly, these laboratories supplied additional data for CYP3A activity, including enzymatic rates (midazolam 1'-hydroxylation or testosterone 6β-hydroxylation) for the vehicle control (DMSO) and rifampicin treated.

Data normalization as percent positive control.

Test compound maximum fold induction data were expressed as percentage (%) of positive control rifampicin response, where the "total signal" is the signal from the positive control (e.g., 10 μM rifampicin) and the "blank signal" is the signal from the solvent-treated wells (or 1-fold) (see Table 1G for equation) (Sinz et al., 2006b). To maximize the available data for analysis, several sources of *in vitro* induction data were combined: IWG generated data for weak clinical inducers, IWG gathered member data for prototypical and proprietary compounds, and data published by Zhang et al (Zhang et al., 2014). Data were normalized to the rifampicin fitted Ind_{max} rather than the response at a given concentration (e.g., 10 μM rifampicin) since this was not available for all datasets.

***In vitro* data analysis: curve fitting, E_{max} and EC_{50}, F2 and slope analysis.**

In vitro concentration-induction response data were collected from literature or provided by IWG member companies. The data selected for analysis was determined to meet quality criteria if

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

the tested concentration range included adequate points to define a baseline (no response) and maximal effect response, prior to fitting. Ideally, typical sigmoidal concentration response data span no-to-full effect, with a minimum of 5 to 6 data points. Nonlinear regression analysis has been recommended for fitting concentration-dependent induction response, as described previously for a typical physiology or pharmacology response (Meddings et al., 1989). To remove data fitting as a source of variability, collated induction data were re-fit using the sigmoidal model described above using GraphPad Prism versions 6.0 and 7.0. Induction parameters were determined by plotting the *in vitro* fold induction data (mRNA and enzyme activity normalized to the control) against the nominal *in vitro* concentration using GraphPad Prism and two concentration-response models (Table 1F and H). The baseline was set to 1 assuming that the vehicle control represents no change and equals a fold induction of 1. The best fitting model was determined based on a sum of squares F-test and Akaike's information criteria results. Note that for IVIVE, the maximal fold induction (Ind_{max}) was converted to E_{max} by subtracting the baseline of 1-fold. In the case of atypical or bell-shaped concentration-response curves, where the higher concentration gave a lower response than the preceding concentration by more than 20%, the higher concentration data were excluded from the fitting. In most of these cases cytotoxicity was a plausible explanation for decreased induction response at higher concentrations. Note that assessment of cytotoxicity was defined by the laboratory that generated the data, a summary of these methods was provided in a previous IWG publication (Hariparsad et al., 2017). No other data exclusion criteria were applied. The initial slope was also determined by fitting the data using linear regression in GraphPad Prism, as a surrogate for full induction parameters in the cases where solubility or cytotoxicity may limit the ability to estimate the clinical risk from the *in vitro* data.

Rifampicin CYP3A4 mRNA concentration-induction response data, generated in 38 human hepatocyte donors and over a concentration range of 0.01 - 30 μM, were collated from IQ

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

member companies using their preferred conditions. The data were fit in Graphpad Prism v7.0 using a 3-parameter log(agonist) vs. response equation, as detailed in Table 1F, to determine the fitted EC_{50} and E_{max} values.

A similar exercise was undertaken to summarize the fitted EC_{50} and E_{max} parameters for CYP3A4 mRNA compound data for the following: troglitazone (10 donors from three laboratories, concentration range of 0.01-20 μM), pioglitazone (12 donors from five laboratories, concentration range of 0.05-150 μM), ritonavir 18 donors from four labs, concentration range of 0.01-100 μM), nifedipine (21 donors from six laboratories, concentration range of 0.05-300 μM), phenobarbital (21 donors from seven laboratories, concentration range of 0.9-3000 μM), carbamazepine (25 donors from seven laboratories, concentration range of 0.01-500 μM), rosiglitazone (26 donors from seven laboratories, concentration range of 0.05-300 μM), and phenytoin (28 donors from seven laboratories, concentration range of 0.1-1000 μM). Mean, standard deviation, median, minimum, maximum and %CV values for each compound data set were calculated using GraphPad Prism, version 7.

To evaluate intra-donor variability, three laboratories provided data for nine donors, where data was available from at least three separate experiments to determine EC_{50} and E_{max} values, on different days in the same donor, using standard company methods. Mean, standard deviation, median, minimum, maximum and %CV values for each donor were calculated using GraphPad Prism, version 7.

Clinical risk assessment.

The clinical relevance of *in vitro* induction was assessed by considering the recommendations in regulatory guidance documents as described by equations A to E in Table 1. Since a degree of variability was observed in the clinical induction response, the median and worst-case *in vitro*

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

induction parameters were compared with both the median and worst-case AUCR values. In addition, the substrate specificity was considered by binning clinical trials according to the contribution of CYP3A to the overall clearance. In cases where the magnitude of clinical induction was substrate dependent (e.g. for ritonavir), additional information on the metabolic pathways was obtained by a literature review (Supplementary Table 2). This review was helpful for evaluating whether the maximal induction response could be mediated through a co-regulated induced enzyme (other than CYP3A), especially in cases where there were mixed mechanisms of DDI observed. Where the plasma free fraction was reported to be <1%, both the reported value and 1% (as recommended in regulatory guidances) were used to estimate the unbound $C_{max,ss}$ in the equations. All the *in vitro* induction parameters were fit using each equation and the worst-case and median donor data were used to evaluate the IVIVE. The rates of false positive and negative predictions were used to assess the utility of the various IVIVE methods. The equations are described in Table 1J to M. Additionally, the ability of the equation to result in quantitative predictions was assessed by comparing the predictions from *in vitro* parameters with the clinically observed AUCR.

Statistical analysis.

Evaluation of normality

Normal quantile plots in the Distribution platform of the JMP® 12.0.1 software (SAS Institute Inc., Cary, NC) were employed to evaluate normality of per-donor distributions of fold induction of negative controls and positive controls. The distributions of negative controls were not systematically non-normal, therefore probability estimates for negative controls assume that the data are normally distributed. The majority of distributions of positive controls were positively skewed, necessitating a log transformation of the positive control data prior to estimation of probabilities. Indeed, both the FDA (2001; https://www.fda.gov/downloads/drugs/guidances/ucm070244.pdf) and EMA (2010;

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf) guidances on bioequivalence recommend a log transformation prior to data analysis. Datasets with a normal distribution were graphed on an arithmetic scale, whereas those exhibiting a non-normal distribution were graphed with a log scale y-axis.

Limit of blank (LoB) and limit of detection (LoD).

Calculations of LoB and LoD were adapted from equations published by Armbruster and Pry (2008). Briefly,

$$LoB = Mean_{blank} + 1.645(SD_{blank}) \text{ and,}$$

$$LoD = LoB + 1.645(SD_{low\ concentration\ sample}),$$

where blank is the negative control (flumazenil) and variation (SD) of the low concentration sample is assumed to be equal to the variation in the blank response. The LoB represents the fold induction value for which there is a 95% probability that a blank, or negative control response, falls below. The LoD represents the fold induction value for which there is a 95% probability that a response above this value is true positive response (i.e. 5% Type I or Type II error).

Estimation of probability of exceeding X-fold induction (per donor).

For negative controls, the mean and standard deviation of the fold induction values for each donor (intra-donor) were calculated by Excel 2010, and then the probability for that donor to exceed X-fold induction was estimated by the Excel function

1-NORM.DIST(X,Mean,StDev,True)

where “X” is the fold induction of interest, “Mean” and “StDev” are the empirical intra-donor mean and standard deviation of the fold induction data of each donor, and the flag “True” instructs the NORM.DIST function to provide the corresponding cumulative normal probability. For positive controls, each fold induction value of each donor was first transformed by the

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

natural logarithm function (LN) in Excel 2010, and then the mean and standard deviation of the log-transformed values of each donor calculated by Excel. Finally, the probability of exceeding X-fold induction for each donor was estimated by the Excel function

1-NORM.DIST(LN(X),Mean(LN induction),StDev(LN induction),True)

where the terms within the NORM.DIST function are as defined above, but now applied to the log-transformed induction data of each donor.

Monte Carlo simulation of the probability that 0, 1, 2, or 3 of three randomly selected donors exceed X-fold induction.

The variability observed in the 10 negative control donors and 15 positive control donors were assumed to be representative of their respective populations. For negative control donors, the @Risk 7.5.1 software (Palisade Corporation, Ithaca, NY) was employed, with an Excel worksheet, to randomly select three donors at a time from among the 10 available donors, and, for each selected donor, to simulate a fold induction value from a normal distribution possessing that donor's fold induction mean and standard deviation. From each set of three donors, the number (0, 1, 2, or 3) of donors exceeding X-fold induction was counted and logged by @Risk 7.5.1. This process was repeated 100,000 times to determine the probability that 0, 1, 2, or 3 donors, among three randomly selected donors, would exceed X-fold induction.

For positive control donors, the same calculation process was employed and repeated 100,000 times, except that, for each donor, a log-transformed fold induction value was simulated from a normal distribution possessing that donor's log-transformed fold induction mean and standard deviation. For a positive control donor, X-fold induction is exceeded when the simulated log-transformed value exceeds LN(X).

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

RESULTS

Establishing a threshold for a positive vs. negative *in vitro* CYP3A4 mRNA induction response.

To evaluate potential thresholds for positive or negative *in vitro* induction response, the variability in *in vitro* human hepatocyte induction experiments was interrogated by analyzing CYP3A4 mRNA and activity data generated with a negative control compound, namely, flumazenil repeated under the same experimental conditions. Fold induction data for flumazenil were collected and analyzed from 10 hepatocyte donors, where data from ≥20 repeated experiments were available in each donor for CYP3A4 mRNA expression. In total, data were collected from 314 individual experiments for CYP3A4 mRNA (range 23-54 experiments/donor) and from 111 individual experiments for CYP3A activity (range 4-24 experiments/donor) (Table 2).

Individual flumazenil data for CYP3A4 mRNA and CYP3A activity, across the 10 hepatocyte donors, are illustrated in Figures 1A and 1B, respectively. Summarized data from statistical analyses are presented in Table 2. As mRNA is the recommended primary endpoint in most CYP induction experiments, subsequent data analyses focused on the variability observed in the CYP3A4 mRNA data sets. Flumazenil-CYP3A4 mRNA data demonstrated a normal distribution and, therefore, were plotted on an arithmetic y-axis (Figure 1A and 1B) and calculation of means and standard deviations were performed without log transformation (Table 2). The majority (300/314; 95.5%) of individual experimental data points for flumazenil-CYP3A4 mRNA were within 2-fold (0.5 – 2-fold) of the vehicle control, DMSO. The mean fold induction values for flumazenil-CYP3A4 mRNA ranged from 1.01- to 1.53-fold (overall mean 1.20-fold) which tracked closely with the vehicle control (represented by 1-fold change) as expected with a true negative control, however, there was notable intra-donor variability. In five of 10 donors examined (50%), there were no reported responses outside the 2-fold range (i.e. <0.5 or >2-

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

fold). In the other five donors, one or more values were outside the 2-fold range (1 <0.5-fold; 13 >2-fold). The calculated probabilities of a flumazenil-CYP3A4 mRNA response exceeding 2-fold within a single donor ranged from 0% to as high as 20.4% (donor H2).

The intra-donor variability in the flumazenil-CYP3A4 mRNA response was explored further with two orthogonal methodologies. First, to better understand the magnitude of the intra-donor variability for the negative control, the mean and standard deviation of the flumazenil responses, within each donor, were used to calculate a limit of blank and limit of detection (Table 2). The limit of blank is the fold induction value beneath which there is a 95% probability that the response is a true negative. Conversely, the limit of detection is the fold induction value above which there is a 95% probability that the response is a true positive. Across the 10 hepatocyte donors examined, the calculated limit of blank or true negative value was <2-fold in 9 of 10 donors (mean of 10 donors was 1.86-fold). Therefore, a CYP3A4 mRNA fold induction value ≤1.86-fold represents a true negative response, with 95% probability based on the data sets examined. The fold induction value indicative of a true positive response above background variation, with 95% confidence or limit of detection, was calculated for the flumazenil-CYP3A4 mRNA data sets based on a means and standard deviation approach. This analysis resulted in a limit of detection ranging from 1.61- to 3.41-fold (i.e. >2-fold in five of 10 donors). Similarly, the calculated threshold for a true positive response above background variation for CYP3A4 mRNA, across all data sets, was 2.52-fold. Therefore, a CYP3A4 mRNA fold induction value ≥2.52-fold represents a true positive response with 95% probability, based on the data sets examined.

The observation of negative control values for flumazenil-CYP3A4 mRNA >2-fold was confirmed with data from a second company. Briefly, CYP3A4 mRNA data was obtained from 23 experiments conducted across nine hepatocyte donors, following treatment with a single

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

concentration of flumazenil (30 μM). In these experiments, the observed mean fold induction value for flumazenil-CYP3A4 mRNA was 1.30 (min 0.88-fold, max 3.37), with calculated limit of blank and limit of detection values of 2.12- and 2.95-fold, respectively.

The probability of a flumazenil-CYP3A4 mRNA response exceeding a 2-fold threshold in a single concentration negative control treatment group in three randomized human hepatocyte donors was assessed with Monte Carlo simulations. The simulations incorporated variability parameters (i.e. means and standard deviations) derived from data reported across the 10 donors (314 experiments). When simulated with 100,000 iterations of individual experiments, containing three donors each, the probability of observing a flumazenil-CYP3A4 mRNA response <2-fold, in all three donors, was 91.9%. Conversely, there was a probability of 8.1% that flumazenil would produce a CYP3A4 mRNA response of ≥2-fold in one or more donors. Therefore, flumazenil is likely to cause a false positive response in approximately 8% of cases if a 2-fold increase in CYP3A4 mRNA defines the threshold between a positive and negative CYP3A4 mRNA *in vitro* response.

For CYP3A activity, less intra-donor variability in the flumazenil response was observed compared to CYP3A4 mRNA. Across the 10 donors examined (n = 111 experiments), the mean fold induction values for flumazenil-CYP3A ranged from 0.95- to 1.09-fold (mean 1.03). The calculated overall mean limit of blank and limit of detection values were 1.20-fold (range 1.07- to 1.30-fold) and 1.37-fold (range 1.14- to 1.50-fold), respectively. There were no observations of flumazenil-CYP3A activities >2-fold and, therefore, the projected frequency of exceeding 2-fold was not determined.

Establishing thresholds of positive *in vitro* induction response to ensure adequate dynamic range.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Rifampicin induction in 15 hepatocyte donors, repeated on multiple occasions, are shown in Figures 1C and 1D, for CYP3A4 mRNA and CYP3A activity, respectively. Summarized statistical analyses are presented in Table 3. In total, data were collected from 581 individual experiments for rifampicin-CYP3A4 mRNA (range 13-70 experiments/donor) and from 377 individual experiments for rifampicin-CYP3A activity (range 13-70 experiments/donor). Subsequent data analyses, as with flumazenil, focused on the variability observed in the rifampicin-CYP3A4 mRNA data sets. In all cases, the rifampicin-CYP3A4 mRNA response was reported as fold induction compared to the vehicle control, DMSO. Rifampicin-CYP3A4 mRNA data sets demonstrated a non-normal distribution and were graphed on a log-based y-axis in Figure 1 (C and D) and calculation of probabilities assumed a lognormal distribution (Table 3). The median rifampicin-CYP3A4 mRNA fold induction values ranged from 7.1- to 75-fold across the 15 donors. There was notable intra-donor variability in response to rifampicin with dynamic response ranges (minimum/maximum fold induction response) of 3.4-fold to 41.5-fold and %CV values ranging from 33.6 to 93.1%. The %CV values (or RSDs), as an indicator of variability, were not dependent on the magnitude of the rifampicin-CYP3A4 mRNA response. However, the standard deviations increased in proportion with the mean response. In this regard, a higher fold change would be expected to be more variable (i.e. larger standard deviation).

Based on the observed intra-donor variability in the rifampicin-CYP3A4 mRNA response, the likelihood of exceeding a predefined positive control threshold (i.e. 6-, 10-, or 20-fold), for each hepatocyte donor, was evaluated (Table 3). The 6-fold positive control threshold was derived from EMA and FDA guidances, whereas the 10- and 20-fold thresholds are based on empirical cutoff values used by some consortium member companies. The 6-fold positive control threshold assumes that 1) the minimum positive *in vitro* induction signal is 2-fold (100% increase), 2) the minimum *in vitro* signal (2-fold) represents no more than 20% of the positive control response, and 3) a 6-fold response equates to 500% increase when the vehicle control

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

is set equal to 1-fold. When the desired rifampicin-CYP3A4 mRNA positive control response was set to 6-fold, the probability of exceeding this threshold ranged from 70 to 100% across the 15 donors examined. As the desired positive control threshold increases, the probability of achieving the response decreases. The probability of achieving rifampicin-CYP3A4 mRNA responses of greater than 10- or 20-fold across all donors ranged from 35 - 100% or 4 - 94%, respectively.

The probability of a rifampicin-CYP3A4 mRNA response above a 6-, 10- or 20-fold threshold was further examined with Monte Carlo simulations that incorporated variability parameters reported across the 15 donors (581 experiments). When simulated with 100,000 iterations, the probability of observing a rifampicin-CYP3A4 mRNA response >6-fold in all three donors was 78.4%, such that rifampicin would produce a response above the desired threshold in all three donors in nearly four of five experiments which equates to a 21.6% fail rate. The probabilities of obtaining a rifampicin-CYP3A4 mRNA response >10- or 20-fold in all three donors were 40.9 and 4.94%, respectively.

As generally observed, the amplitude of the fold induction response for rifampicin induced CYP3A activity was lower than the rifampicin-CYP3A4 mRNA response (Fahmi et al., 2010). Also, there was less intra-donor variability observed for rifampicin induced CYP3A activity. Across all 15 hepatocyte donors examined, the median fold induction values for rifampicin-CYP3A activity ranged from 3.6- to 18.1-fold (mRNA 7.1- to 75-fold). The %CV values ranged from 18.2 to 61.7%, which were, on average, less than corresponding %CV values for CYP3A4 mRNA. Monte Carlo simulations were not performed for rifampicin-CYP3A activity.

Basal CYP expression and impact on fold induction.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

The basis for the observed intra-donor variability in the rifampicin-CYP3A4 mRNA response, across repeat experiments, was further explored in hepatocyte donor H2 by analysis of RT-PCR data. Raw data (Ct values) were collected for CYP3A4 and a housekeeping gene (GAPDH) from multiple treatment groups, including the vehicle (DMSO), negative (flumazenil) and positive (rifampicin) controls (Figure 2). Figure 2A shows housekeeping gene Ct values for all treatment groups plotted in chronological order of experimentation (>50 experiments). Amongst all three treatment groups GAPDH Ct values tracked similarly and there was a consistent inter-experimental variation regardless of time (experiments conducted over ~1.5 years). Figure 2B shows raw CYP3A4 Ct values for vehicle (DMSO) and negative (flumazenil) controls. Data were rank-ordered by increasing CYP3A4 Ct values from the DMSO-treated samples. Since Ct values are inversely proportional to transcript levels, the experiments with the highest basal CYP3A4 transcript levels (lowest Ct values) are on the left side of the graph. Flumazenil CYP3A4 Ct values tracked closely with the DMSO data. CYP3A4 Ct values were normalized to GAPDH Ct values and the resultant delta Ct (ΔCt) values are plotted in Figure 2C. CYP3A4 ΔCt values for DMSO and flumazenil were generally similar. Across the experiments, the range of ΔCt values for DMSO-CYP3A4 was approximately 7, which equates to a 128-fold difference in basal CYP3A4 transcript levels (calculated by 2^7). In all cases, the rifampicin-CYP3A4 ΔCt values were lower than the corresponding vehicle control values, denoting higher levels of CYP3A4 transcript, as expected.

In Figure 2D, the resultant fold induction values (ΔΔCt) for rifampicin-CYP3A4 mRNA are ranked based on basal CYP3A4 mRNA expression (highest basal expression on the left). Figure 2D also shows that the magnitude of the rifampicin-CYP3A4 mRNA response inversely correlates with basal CYP3A4 mRNA levels. This observation suggests that hepatocytes with low basal CYP3A4 mRNA levels may demonstrate high CYP3A4 mRNA fold induction responses to rifampicin. Similar findings for CYP3A4 mRNA were observed in two additional

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

donors (H4 and H12; Supplementary Figures 1 and 2, respectively). This effect was less pronounced for rifampicin-CYP3A activity response but was based on fewer experiments from donors H2, H4 and H12 (Supplementary Figure 3).

The potential for assay noise to systematically affect the magnitude of the rifampicin-CYP3A4 mRNA response was evaluated by comparison to the corresponding intra-assay flumazenil response. This assessment was conducted across multiple repeated experiments within the same hepatocyte donor. As the fold induction for rifampicin increased in donor H2, there was no corresponding change in the negative control (flumazenil) response confirming that the variability was not a function of assay noise (Figure 2B and 2D). Similar results were observed in other donors (data not shown).

The number of experiments that might be necessary to capture the range of variability in the rifampicin-CYP3A4 mRNA response described above was evaluated, with data visualized based on chronological order of experimentation (Figures 2E and 2F). Figure 2E shows CYP3A4 ΔCt values for DMSO and rifampicin plotted by chronological experiment order and Figure 2F illustrates the resultant rifampicin-CYP3A4 mRNA fold induction values. There was no clear trend in the data with respect to time in either ΔCt or fold induction values. Consequently, the number of repeat experiments required to capture variability in the rifampicin-CYP3A4 mRNA induction response in a single hepatocyte donor is considerable (e.g. ≥5 repeat experiments) and may vary between donors.

Normalizing *in vitro* induction data to a positive control.

Multiple datasets, with maximum fold-induction for rifampicin and test compound, were combined to explore the utility of normalizing data as percent of positive control. Figure 3A shows percent positive control (rifampicin) data for CYP3A4 mRNA induction for 30 compounds

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

in three donors. The untransformed fold-induction data is shown in Supplementary Figure 4. Note that in some donors the rifampicin response was on the low side (~6-fold). Since data for test compound indicated positive *in vitro* CYP3A4 mRNA induction (>2-fold) this dataset holds value and was included. There were marked differences observed for compounds when looking at percent rifampicin control response across donors. For example, the carbamazepine response was 52, 28 and 218%, phenobarbital was 96, 36 and 106% and phenytoin was 40, 23 and 44% (where rifampicin maximum induction was 7-, 16- and 7-fold for 1st, 2nd and 3rd donor, respectively, within the same laboratory). A similar trend was observed in a dataset generated across laboratories using different donors. Here the felbamate response was 20, 34 and 33%, oxcarbazepine was 28, 105 and 88% of rifampicin response (where rifampicin maximum induction was 19-, 13- and 19-fold for 1st, 2nd and 3rd donor, respectively). Similarly, in a third dataset, where each compound was tested in a single laboratory using multiple donors, Cmpd1 was 64, 121 and 136% of rifampicin response which was 25-, 24- and 17-fold, respectively, Cmpd4 was 15, 18 and 44% of rifampicin response which was 41-, 133- and 96-fold, respectively, and Cmpd7 was 21, 27 and 20% of rifampicin response which was 17-, 12- and 11-fold, respectively.

Finally, the utility of normalization to a positive control response to address intra-donor variability, was explored. Figure 3B shows rosiglitazone and pioglitazone induction as percent positive control response (rifampicin CYP3A4 mRNA) in three different donors that were repeated on five separate occasions within the same laboratory. Within a single donor over time, similar to the previous dataset, a lack of normalization was observed, with the percent positive control values spanning a wide range for each compound. For example, rosiglitazone was 88, 31 and 61% and pioglitazone was 47, 41 and 77% of rifampicin response for each donor in the second experimental repeat. In the third experimental repeat, rosiglitazone was 76, 13 and 129% and pioglitazone was 31, 100 and 85% of rifampicin response for each donor.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

***In vitro* induction parameters and reproducibility across donors and labs**

Following analysis of a large dataset of single concentration data from two labs, the IWG extended analysis to concentration-response induction data generated in multiple laboratories, under different conditions in multiple human donors. Rifampicin CYP3A4 mRNA EC_{50} and E_{max} values were collated from five literature sources (at least n=3 unique donors for inclusion) and from multiple IQ member companies (Supplementary Table 3). Variability was observed for both the EC_{50} and E_{max} parameters calculated within the six data sets (as given by %CV) (EC_{50}: 51.6 to 144% CV; E_{max}: 28.6 to 104% CV). Overall the mean and median values across the data sets were within 2-fold of each other with the exception of the EC_{50} for the IWG data, which was within 2.5-fold.

An additional rifampicin dataset was collected to further examine this variability. The reproducibility within a donor, under the same experimental conditions, in the same laboratory, was examined. Rifampicin CYP3A4 EC_{50} and E_{max} data were collated from three different companies (nine donors) where at least three experiments were available for each donor (Figure 4 Supplementary Table 4). Variability, within each donor (expressed as %CV) ranged from 28.6 to 77.3% for EC_{50} and 22.9 to 125% for E_{max} values. The mean and median values of the data set were within 2-fold of one another. The spread in minimum to maximum values observed within each donor ranged from 1.1- to 9.5-fold for EC_{50} and 1.49- to 9.19-fold for E_{max}. The variability observed in CYP3A4 EC_{50} and E_{max} parameters was not unique to rifampicin (Figure 5; Supplementary Table 5). Similar %CV values were noted for eight other CYP3A4 inducers (troglitazone, pioglitazone, ritonavir, nifedipine, phenobarbital, carbamazepine, rosiglitazone and phenytoin) and ranged from 72 to 133% for EC_{50} and 59 to 119% for E_{max} values.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Dataset for DDI IVIVE.

In vitro CYP3A4 and clinical data were collected for 51 compounds covering both clinical and *in vitro* induction response from inhibition, no-effect and induction (Figures 5 and 6; Supplementary Table 6).

***In vitro* dataset.**

For most inducers a minimum of three donors were available for generating median induction parameters. In the case of saquinavir, teriflunomide, Cmpd 3, Cmpd 8, Cmpd 9 and Cmpd 15, data were only available, or induction parameters could only be defined, from two donors. In the case of Cmpd 5 and Cmpd 14, induction parameters could only be determined from one of the three donors investigated. For Cmpd 5, only one donor resulted in measurable increases in CYP3A4 mRNA (>2-fold), two donors were negative. For Cmpd 14, while three donors were evaluated, only one donor included enough concentrations to characterize the concentration response profile. In both of these cases, the clinical observation was inhibition. The weak *in vitro* inducers, defined as those eliciting a <3-fold CYP3A4 mRNA induction in at least one of the donors, were aprepitant, omeprazole, pioglitazone, pleconaril and terbinafine. In some cases, moderate to strong clinical inducers, including carbamazepine, Cmpd 7 and phenytoin, had at least one donor with an E_{max} value <4-fold. In general, the *in vitro* variability for all of the inducers was consistent with that observed for rifampicin (Figure 5). There were some trends discernible for EC_{50}, where moderate and strong clinical inducers generally exhibited much lower EC_{50} values compared with compounds that had weak or no clinical induction (Figure 5A). However, an exception was noted for Cmpd 3, which showed moderate clinical induction due to its relatively high unbound circulating concentration (5.6 μM). As one might expect, there was no trend in EC_{50} values with clinical DDI magnitude for the compounds that exhibited both *in vitro* induction and inhibition (Figure 5B). In general, the E_{max} values for rifampicin, while variable, were higher than those observed from weak or non-clinical inducers such as

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

perampanel or lersivirine. E_{max} values for compounds with *in vitro* induction only (Figure 5C) generally trended down with increasing EC_{50} value. There were no discernible trends in E_{max} for compounds that exhibited both *in vitro* induction and inhibition (Figure 5D). In the case of rifapentine, nifedipine and rosiglitazone, the E_{max} values were comparable to those determined for rifampicin, although these drugs result in no clinical induction (Figure 5D).

Clinical dataset

The IWG collected data for 35 literature compounds and 16 proprietary compounds from the IQ member companies. When considering the median clinical AUCR and DDI category relative to the FDA guidance (FDA, 2012), there were eight compounds with clinical inhibition, 16 with no effect (AUCR; 0.8-1.25), 16 with weak induction (AUCR; 0.5-0.8), nine with moderate induction (AUCR; 0.2-0.5), and two with strong induction (AUCR; <0.2). When considering the worst-case (or greatest induction) clinical AUCR, there were six compounds with clinical inhibition, nine with no effect, 15 with weak induction, 16 with moderate induction, and five with strong induction (Supplementary Table 7). Of these compounds, 31 of 51 (61%) exhibited mixed DDI mechanisms towards CYP3A (i.e. *in vitro* induction plus inhibition and/or inactivation).

Data from 1,048 clinical trials were collected for *in vitro* CYP3A inducers (Supplementary Table 9). These trials included all substrates with some role of CYP3A in the overall metabolism, as determined by literature searches for *in vitro* or *in vivo* metabolism data. When the clinical data were refined to include only rifampicin doses 600 mg or greater, and dosing regimens 5 days or longer, there were 835 datasets remaining. This translated to a total of 181 clinical DDI datasets, when considering only the sensitive CYP3A substrates, and 74 studies that used the recommended index substrates, midazolam or triazolam, (71 and three, respectively) (Figure 6). All the proprietary clinical datasets included midazolam as the probe substrate to assess induction of CYP3A. In general, the AUCR range was similar whether all data was considered

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

or only the sensitive CYP3A victim drugs, with the exception of some potent mixed mechanism DDI compounds (e.g. ritonavir). The prevalence of induction (i.e. AUCR <0.8) was determined to be 56% using median AUCR values and 72% using the worst-case AUCR values. Despite this refinement of the data, a reasonable degree of variability remained in the clinical induction response as can be visualized in the rifampicin and ritonavir data (Figure 6).

Translating *in vitro* induction data to clinically relevant risk of induction DDI

The large datasets collected (Figures 5 and 6) enabled evaluation of various simplistic models for predicting clinical induction risk. The potential for each method to provide meaningful risk assessment was considered based on the number of false negative or false positive compounds (Table 4).

High false positive rates (>35.7%) were observed when comparing the output from the recommended models and the median observed clinical AUCR, with the exception of the mechanistic static model that considered both induction and inhibition (16.7% false positive rate using the median *in vitro* donor induction data). The quantitative prediction accuracy, using the induction/inhibition mechanistic static model, (17% within bioequivalence and 43% within 2-fold), was not as high as that of other methods such as the R3 using unbound $C_{av,ss}$ (31% within bioequivalence and 94% within 2-fold when using the median *in vitro* donor data) and the percentage of false negatives was higher with inhibition/induction mechanistic static model than other methods (27 to 36%). Compiling all *in vitro* data into the RIS or slope correlation curves enabled quantitative prediction and a minimal number of false negatives (Table 4). A noted limitation of this approach is that no test sets were available to evaluate true predictive performance, as predictions were made for compounds that were used to build the correlation model. A similar observation was made using a d-value of 0.3, based on the large multi-donor *in vitro* dataset collected here. When an R3 cut-off of 0.8 is used rather than 0.9, with total

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

$C_{max,ss}$ as the input and a d-value of 1, the percentage of true negatives was significantly improved from 3% to 17% with only a small effect on the false negatives (increased from one to two). Using the recommended equation in the draft FDA 2017 DDI guidance (Table 1C), which incorporates a 10-fold multiplier to the $C_{max,ss,u,}$ resulted in two more false negatives (pleconaril and Cmpd15 in addition to dexamethasone) than the 2012 guidance (dexamethasone). Applying a multiplier of 50-fold rather than 10-fold reduced the number of false negatives from three to zero. Using the gut concentration as the input for the R3 and F2 models also reduced false negatives. Limiting the input for unbound plasma protein binding to 1% resulted in fewer false negatives. However, those false negatives that remained (dexamethasone and oxcarbazepine) had only moderate plasma protein binding and the inclusion of compounds with unbound plasma concentrations <1%, including Cmpd 13, efavirenz, rosiglitazone and teriflunomide resulted in appropriate binning when the reported unbound plasma protein value was used. Of all of the methodologies investigated, using the average unbound $C_{max,ss}$ resulted in the fewest number of false positives but increased the number of false negatives (from one to six when using the median induction parameters). The average unbound $C_{max,ss}$ also resulted in the highest number of predictions within 2-fold or bioequivalence, 94% and 31%, respectively. Using the unbound $C_{max,ss}$ for hepatic and the portal concentration for the gut component resulted in two false negatives (dexamethasone and pleconaril) and improved the percentage of false positives over many of the other IVIVE methods.

When *in vitro* induction parameters cannot be defined, either due to solubility or cytotoxicity limitations, the F2 or slope values can often be estimated. The slope tended to over-predict the magnitude of induction compared with the EC_{50} and E_{max} values, while the F2 value resulted in four false negatives (dexamethasone, pleconaril, Cmpd 2, Cmpd 15) compared to one false negative using the R3 equation with total $C_{max,}$ d = 1 and a cut-off of 0.8. When the $F2/C_{max,ss,u}$ multiplier was reduced from 50-fold to 30-fold, there was no impact on the false negative rate.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

However, the false positive rate decreased, from 83 to 70% using median data and 87 to 78% using worst-case data. In order to evaluate the ability of the F2 value to predict induction at the gut level, the F2 equation was solved for the dose level of perpetrator using molecular weight and the equation in the EMA guideline (0.1 x Dose/250 mL). When applying a cutoff value of 0.25 for dose level=F2/therapeutic dose level, the only false negative observed was dexamethasone.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

DISCUSSION

The IWG compiled extensive *in vitro* and clinical induction datasets focusing on interpretation of *in vitro* induction data for CYP3A4 mRNA and its clinical relevance. Strikingly, there was a large degree of variability in both clinical and *in vitro* induction responses (Figures 5 and 6). Variability occurred, irrespective of experimental conditions, laboratory and test compound, and was not solely accounted for by differences in donor response as previously suspected. Importantly, despite being variable, *in vitro* induction data has utility in clinical DDI risk assessment and decision-making. Six recommendations are derived from this analysis.

CYP induction should continue to be evaluated in three separate human donors *in vitro.* *In vitro* CYP3A4 mRNA data for rifampicin included diverse sets of multiple repeats within a donor, from the same laboratory/experimental condition (Figure 4). The intra-donor variability was similar to that observed between donors and across-laboratories (Figure 5 and Supplementary Table 3). Beyond rifampicin, variability exists across the compound dataset (Figure 5). Of note, CYP3A4 activity appeared to be less variable for the single concentration rifampicin dataset (Table 3). However, there was insufficient EC_{50} and E_{max} data for further evaluation. Given the observed variability, *in vitro* CYP induction should continue to be evaluated in three separate human donors, thus supporting existing recommendations from regulators.

Why might this variability exist? It is possible that small differences in cell culture; temperature, plate agitation, pipetting speed and media change times, between each experiment, could drive variability since all may impact efficient attachment, cell morphology and basal CYP expression (Hamilton et al., 2001; Hewitt et al., 2007). Intra-donor variability in basal CYP expression appeared to determine variability in fold induction response for rifampicin (Figure 2). Additionally, differences in intracellular drug concentration, in response to changes in enzyme or

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

transporter expression, could contribute to a range of induction responses inter- and intra-donor (Chu et al., 2013; Sun et al., 2017).

In light of empirically divergent responses in rifampicin control and most test inducers, normalization of data to percent positive control appears to be of limited benefit. To account for donor variability in induction response, normalization to percent positive control was previously suggested (Bjornsson et al., 2003). The assumption was that although the absolute fold induction value might be different between donors, the relative magnitude of response for different compounds would be preserved within a donor. This is commonly used for reporter gene assays (Persson et al., 2006; Sinz et al., 2006a), but reports in human hepatocytes are from smaller studies (Kamiguchi et al., 2010). The range of percent control response for each compound shown in Figure 3A suggests that this does not normalize the induction response across donors or laboratories; nor does it normalize response within a donor over time (Figure 3B). Thus normalization to % rifampicin response provides limited benefit in aiding data interpretation. Why is this normalization not successful? There is no mechanistic evaluation that explains the compelling data observed here. Do different metabolic pathways predominate in different donors for a compound (Richert et al., 2006; Heslop et al., 2017)? In this case, changes in metabolism between donors, resulting in different effective drug concentrations, might explain why test and control compound response do not track. Several genetic variants of PXR and CAR exist and could contribute to inter-individual variation in induction response (Lamba et al., 2005). Further, if test and control compound differ in regulation of PXR and CAR, and donors differ in PXR and CAR expression, then test and control response may not track (Faucette et al., 2006). Subtle differences in intracellular concentration between donors and compounds could also be confounding. This could occur by multiple factors, such as, small changes in amount of drug dosed *in vitro,* differences in seeding density and cell attachment,

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

and thus changes in unbound fraction in the incubation, and different drivers of cellular uptake such as transporter expression and albumin concentration (Miyauchi et al., 2018).

Two-fold induction, with concentration dependence, is an acceptable threshold for positive identification of *in vitro* CYP3A4 mRNA induction. Regulatory agencies recommend >2-fold change, relative to vehicle control, to identify a positive *in vitro* inducer. This recommendation has evoked considerable discussion as being too stringent a threshold for clinical relevance (Fahmi et al., 2010), especially for changes in CYP3A4 mRNA. A large flumazenil CYP3A4 mRNA dataset helped interrogate the appropriateness of a 2-fold cut off. Flumazenil is not an inducer of CYP3A4 mRNA or activity *in vitro* (Fahmi et al., 2010), nor is it a CYP3A inducer clinically (Ma et al., 2009; Fahmi et al., 2010). A limit of blank and limit of detection analysis in 10 donors, was used to understand thresholds, based on assay signal-to-noise. This defined a true negative as ≤1.86-fold and true positive as ≥2.52-fold (Table 2). This was consistent with a smaller dataset (true negative ≤2.12-fold, true positive ≥2.95-fold). This statistical analysis supports 2-fold as the threshold to define positive induction of CYP3A4 mRNA. A compound can confidently be assigned as having no CYP3A4 induction if three donors all result in <2-fold induction, at clinically relevant concentrations. Note, identifying a compound as positive *in vitro* does not necessarily mean a clinical study is warranted, only that further evaluation of risk is required using mathematical DDI prediction models.

To reduce the risk of false positives, in the absence of a concentration dependent response, induction ≥2-fold should be observed in more than one donor to classify a compound as an *in vitro* inducer. Monte Carlo simulations of flumazenil (100,000 iterations; Table 2), indicates that the probability of observing a false positive of >2-fold response in one of three donors is ~8%. Thus, a single donor with a weak CYP3A4 mRNA induction >2-fold is not sufficient to define a true positive. Two or more data points above 2-fold, and concentration

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

dependence, is recommended to confidently define a positive. For weak induction, the IWG acknowledge that defining concentration dependence can be somewhat subjective. The following should be considered; evidence of concentration response (visual inspection), statistical significance (correlation and linear regression), relevance (i.e. above 2-fold). If only one donor exhibits induction, adding a fourth could be considered to probe for a false positive. If the fourth donor is negative, this strongly suggests a false positive and may obviate the need for follow up. To further avoid false positives, in the absence of a concentration dependence (providing cytotoxicity or solubility is not limiting), if only a single point is >2-fold CYP3A4 mRNA in more than one donor, additional investigation is warranted to contextualize *in vitro* observations to clinical relevance.

If qualifying a compound as negative for CYP3A4 mRNA induction, the magnitude of maximal rifampicin response in that donor should be ≥10-fold. Applying a minimum rifampicin response threshold ensures that weak inducers are not overlooked. A ≥10-fold threshold provides sufficient dynamic range, giving confidence in a negative determination (<2-fold) and affords a window to determine weak, but clinically relevant inducers (e.g., pleconaril, felbamate and Cmpd 7 which had median E_{max} values of 3-, 4.7- and 3.4-fold, respectively, Supplementary Table 6). This threshold is only critical when a test compound has data <2-fold and the result is being interpreted as negative for *in vitro* induction. For compounds with clear concentration response and EC_{50} and E_{max} values readily determined, those data are of value, independent of the rifampicin response in that experiment. The selection of 10-fold is somewhat arbitrary but pragmatic and data driven. Using single concentration rifampicin (Table 3), setting the threshold at >20-fold would result in too many donors not passing. Indeed, Monte Carlo simulations indicate a ~5% frequency of all three donors tested reaching >20-fold. Conversely, while setting the threshold at 6-fold would result in most datasets falling into range, there would not be sufficient window to detect weak inducers, between the true and false positive frequency,

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

since there is an 8% probability of false positives >2-fold. At the proposed >10-fold threshold, there is >40% probability of all three donors tested falling into range (Table 2). The potential for a high *in vitro* assay failure rate is naturally concerning. However, the additional *in vitro* investment becomes more palatable in contrast to potentially unnecessary clinical DDI trials due to insufficient confidence in defining negative *in vitro* induction. Finally, it should be noted that the 15 hepatocyte donors examined here were initially characterized to produce, at minimum, a >6-fold rifampicin-CYP3A4 mRNA response. Thus the calculated probabilities could be biased based on this initial acceptance criterion. An alternative approach might be a weak inducer control to demonstrate confidence that clinically relevant inducers in the 3- to 4-fold range could be identified. However, there is insufficient data available to evaluate the utility of this approach.

Inclusion of a negative control adds no value beyond that of the vehicle control. Vehicle and negative control data are superimposable (Figure 2C). The flumazenil data was useful for determining false positive frequency. An appropriate vehicle control should be included.

Interestingly, the rate of false positive prediction of induction DDI was generally lower when using the *in vitro* donor median vs. the worst-case parameters. Previously, various static and dynamic modeling methods were used to predict clinical CYP3A induction in 28 clinical trials for 13 compounds (Einolf et al., 2014). Expanding this, we evaluated data from over 1000 clinical trials for 51 compounds. However, dynamic modeling was out of scope. All prediction methods, which were variations of the five different approaches detailed by regulatory agencies (F2, RIS, R3, slope correlation and static modeling), had some incidence of false positive prediction (Figure 7) compared to the median AUCR (Table 4). However, the rate of false positive prediction was lower (19 out of 23 methods) when using *in vitro* donor median vs. worst-case parameters. Conversely, the rate of false negative predictions was higher (10 out of 23 methods) using *in vitro* median versus worst-case (Table 4 and Supplementary Table 8),

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

particularly with unbound concentrations. Using slope or F2 as *in vitro* induction input parameters served as a reasonable surrogate for EC_{50} and E_{max} when binning the clinical induction risk. Additionally, applying unbound concentrations generally lowered the false positive rate but increased false negatives. Quantitative accuracy, as assessed by % predictions within 2-fold, was better when unbound concentrations were used. Thus, the situational preference for avoiding false negatives or positives could drive selection of the prediction approach. Historically, regulatory agencies advocated total plasma concentration as a conservative estimate to avoid false negative results in I/K_i calculations (Zhang et al., 2009). Importantly, median donor data improves quantitative estimation of risk by increasing the number of predictions within bioequivalence and 2-fold of observed clinical data (Table 4). Thus, median data of three *in vitro* donors, rather than the worst-case donor, should be considered for induction DDI risk assessment.

The above analysis and recommendations only pertain to CYP3A4 mRNA. It is possible that some findings are CYP isoform specific and additional work is necessary to evaluate CYP1A2 and CYP2B6. Since recent regulatory recommendations have focused on changes in mRNA, there was limited enzyme activity data to mine. A prevalence of CYP3A4 time-dependent inhibitors limits the value of CYP activity as an endpoint. However, in the absence of TDI, it retains significant value. Using activity, would the apparent decrease in variability of single concentration positive control (Figure 1) extend to EC_{50} and E_{max} data? Additionally, attempting to avoid false positives, if CYP3A4 mRNA response is >2-fold but activity is increased <2-fold (without TDI), would there be more confidence in a negative induction definition? Another shortfall of the current analysis is the use of nominal *in vitro* concentrations, since actual concentration data were not available. Not accounting for incubational binding or compound loss by metabolism may result in over estimation of EC_{50}, subsequently impacting IVIVE (Sun et al., 2017). Finally, whilst PBPK modeling was out of scope, dynamic simulation of inducer

concentration could yield further improvements to IVIVE (Einolf et al., 2014; Almond et al., 2016; Ke et al., 2016). Given the incidence of complex DDI involving multiple mechanisms, predicting DDI should address both inhibition and induction (Kirby et al., 2011; Fukushima et al., 2013).

The IWG has presented a data driven evaluation of *in vitro* human CYP induction with several recommendations (Figure 8). The analysis supports the regulators' recommendations to use three human donors *in vitro* to assess induction and application of a 2-fold CYP3A4 mRNA threshold, coupled with concentration dependency, to determine a positive *in vitro* induction signal. The IWG propose several actions around use of controls to aid data interpretation, and showed that while both *in vitro* and *in vivo* induction data are somewhat variable, simple static models of clinical risk using *in vitro* data can be used for decision making.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Acknowledgments

The authors would like to thank the IQ IWG for valuable discussion of data and manuscript review, the IQ DMLG for manuscript review, the IQ member companies for data, Drs Scott Obach (Pfizer), Christopher Gibson (Merck), Michael Sinz (BMS) and Edward LeCluyse (LifeNet Health) for valuable feedback on draft manuscripts. Sophie Mukadam (Genentech) for assistance in literature data-mining. Philip Yates (Pfizer) for valuable statistical discussions. Marina Slavsky (Sanofi), Amanda Moore (Vertex) and Thuy Ho, Rob Clark and Sarah Trisdale (Corning) for *in vitro* human hepatocyte induction studies for clinically weak inducers.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Authorship Contributions

Participated in research design: Kenny, Ramsden, Buckley, Dallas, Fung, Mohutsky, Einolf, Chen, Dekeyser, Fitzgerald, Goosen, Siu, Walsky, Zhang, Tweedie and Hariparsad.

Conducted experiments: Ramsden, Fitzgerald, Zhang and Hariparsad.

Contributed new reagents or analytical tools: NA

Performed data analysis: Kenny, Ramsden, Buckley, Dallas, Fung, Mohutsky and Hariparsad.

Wrote or contributed to the writing of the manuscript: Kenny, Ramsden, Buckley, Dallas, Fung, Mohutsky, Einolf, Tweedie and Hariparsad.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

REFERENCES

Almond LM, Mukadam S, Gardner I, Okialda K, Wong S, Hatley O, Tay S, Rowland-Yeo K, Jamei M, Rostami-Hodjegan A, and Kenny JR (2016) Prediction of Drug-Drug Interactions Arising from CYP3A induction Using a Physiologically Based Dynamic Model. *Drug Metab Dispos* **44:**821-832.

Armbruster DA, Pry T (2008) Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev.* **29** Suppl 1:S49-52.

Bjornsson TD, Callaghan JT, Einolf HJ, Fischer V, Gan L, Grimm S, Kao J, King SP, Miwa G, Ni L, Kumar G, McLeod J, Obach RS, Roberts S, Roe A, Shah A, Snikeris F, Sullivan JT, Tweedie D, Vega JM, Walsh J, and Wrighton SA (2003) The conduct of in vitro and in vivo drug-drug interaction studies: A Pharmaceutical Research and Manufacturers of America (PhRMA) perspective. *Drug Metab Dispos* **31:**815-832.

Chu X, Korzekwa K, Elsby R, Fenner K, Galetin A, Lai Y, Matsson P, Moss A, Nagar S, Rosania GR, Bai JP, Polli JW, Sugiyama Y, Brouwer KL, and International Transporter Consortium (2013) Intracellular drug concentrations and transporters: measurement, modeling, and implications for the liver. *Clin Pharmacol Ther* **94:**126-141.

Einolf HJ (2007) Comparison of different approaches to predict metabolic drug-drug interactions. *Xenobiotica* **37:**1257-1294.

Einolf HJ, Chen L, Fahmi OA, Gibson CR, Obach RS, Shebley M, Silva J, Sinz MW, Unadkat JD, Zhang L, and Zhao P (2014) Evaluation of various static and dynamic modeling methods to predict clinical CYP3A induction using in vitro CYP3A4 mRNA induction data. *Clin Pharmacol Ther* **95:**179-188.

EMA (2010) Committee for Medicinial Products for Human Use (CHMP): Guideline on the Investigation of Bioequivalence. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf

EMA (2012) Guideline on the Investigation of Drug Interactions http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/07/WC500129606.pdf,

Fahmi OA, Kish M, Boldt S, and Obach RS (2010) Cytochrome P450 3A4 mRNA is a more reliable marker than CYP3A4 activity for detecting pregnane X receptor-activated induction of drug-metabolizing enzymes. *Drug Metab Dispos* **38:**1605-1611.

Fahmi OA and Ripp SL (2010) Evaluation of models for predicting drug-drug interactions due to induction. *Expert Opin Drug Metab Toxicol* **6:**1399-1416.

Fahmi OA, Shebley M, Palamanda J, Sinz MW, Ramsden D, Einolf HJ, Chen L, and Wang H (2016) Evaluation of CYP2B6 Induction and Prediction of Clinical Drug-Drug Interactions: Considerations from the IQ Consortium Induction Working Group-An Industry Perspective. *Drug Metab Dispos* **44:**1720-1730.

Faucette SR, Sueyoshi T, Smith CM, Negishi M, Lecluyse EL, and Wang H (2006) Differential regulation of hepatic CYP2B6 and CYP3A4 genes by constitutive androstane receptor but not pregnane X receptor. *J Pharmacol Exp Ther* **317:**1200-1209.

FDA (2001) Guidance for Industry: Statistical Approaches to Establishing Bioequivalence https://www.fda.gov/downloads/drugs/guidances/ucm070244.pdf

FDA (2012) Guidance for Industry: Drug Interaction Studies-Study Design, Data Analysis, Implications for Dosing and Labeling Recommendations (Draft).

FDA (2017) Guidance for Industry: Drug Interaction Studies-Study Design, Data Analysis, Implications for Dosing and Labeling Recommendations (Draft). https://www.fda.gov/downloads/drugs/guidances/ucm292362.pdf

Fukushima K, Kobuchi S, Mizuhara K, Aoyama H, Takada K, and Sugioka N (2013) Time-dependent interaction of ritonavir in chronic use: the power balance between inhibition and induction of P-glycoprotein and cytochrome P450 3A. *J Pharm Sci* **102:**2044-2055.

Gabriel L, Tod M, and Goutelle S (2016) Quantitative Prediction of Drug Interactions Caused by CYP1A2 Inhibitors and Inducers. *Clin Pharmacokinet* **55:**977-990.

Hamilton GA, Jolley SL, Gilbert D, Coon DJ, Barros S, and LeCluyse EL (2001) Regulation of cell morphology and cytochrome P450 expression in human hepatocytes by extracellular matrix and cell-cell interactions. *Cell Tissue Res* **306:**85-99.

Hariparsad N, Ramsden D, Palamanda J, Dekeyser JG, Fahmi OA, Kenny JR, Einolf H, Mohutsky M, Pardon M, Siu YA, Chen L, Sinz M, Jones B, Walsky R, Dallas S, Balani SK, Zhang G, Buckley D, and Tweedie D (2017) Considerations from the IQ Induction Working Group in Response to Drug-Drug Interaction Guidance from Regulatory Agencies: Focus on Downregulation, CYP2C Induction, and CYP2B6 Positive Control. *Drug Metab Dispos* **45:**1049-1059.

Heslop JA, Rowe C, Walsh J, Sison-Young R, Jenkins R, Kamalian L, Kia R, Hay D, Jones RP, Malik HZ, Fenwick S, Chadwick AE, Mills J, Kitteringham NR, Goldring CE, and Kevin Park B (2017) Mechanistic evaluation of primary human hepatocyte culture using global proteomic analysis reveals a selective dedifferentiation profile. *Arch Toxicol* **91:**439-452.

Hewitt NJ, Lechon MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GM, and Hengstler JG (2007) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metab Rev* **39:**159-234.

Kamiguchi N, Aoyama E, Okuda T, and Moriwaki T (2010) A 96-well plate assay for CYP4503A induction using cryopreserved human hepatocytes. *Drug Metab Dispos* **38:**1912-1916.

Ke A, Barter Z, Rowland-Yeo K, and Almond L (2016) Towards a Best Practice Approach in PBPK Modeling: Case Example of Developing a Unified Efavirenz Model Accounting for Induction of CYPs 3A4 and 2B6. *CPT Pharmacometrics Syst Pharmacol* **5:**367-376.

Kirby BJ, Collier AC, Kharasch ED, Whittington D, Thummel KE, and Unadkat JD (2011) Complex drug interactions of HIV protease inhibitors 1: inactivation, induction, and inhibition of cytochrome P450 3A by ritonavir or nelfinavir. *Drug Metab Dispos* **39:**1070-1078.

Kozawa M, Honma M, and Suzuki H (2009) Quantitative prediction of in vivo profiles of CYP3A4 induction in humans from in vitro results with a reporter gene assay. *Drug Metab Dispos* **37:**1234-1241.

Lamba J, Lamba V, and Schuetz E (2005) Genetic variants of PXR (NR1I2) and CAR (NR1I3) and their implications in drug metabolism and pharmacogenetics. *Curr Drug Metab* **6:**369-383.

Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25:**402-408.

Ma JD, Lawendy NM, Fullerton T, Snyder PJ, Nafziger AN, and Bertino JS, Jr. (2009) Effect of intravenous flumazenil on oral midazolam pharmacokinetics and pharmacodynamics for use as a cytochrome P450 3A probe. *Int J Clin Pharmacol Ther* **47:**111-119.

Meddings JB, Scott RB, and Fick GH (1989) Analysis and comparison of sigmoidal curves: application to dose-response data. *Am J Physiol* **257:**G982-989.

Miyauchi S, Masuda M, Kim SJ, Tanaka Y, Lee KR, Iwakado S, Nemoto M, Sasaki S, Shimono K, Tanaka Y, and Sugiyama Y (2018) The Phenomenon of Albumin-Mediated Hepatic Uptake of Organic Anion Transport Polypeptide Substrates: Prediction of the In Vivo Uptake Clearance from the In Vitro Uptake by Isolated Hepatocytes Using a Facilitated-Dissociation Model. *Drug Metab Dispos* **46:**259-267.

Persson KP, Ekehed S, Otter C, Lutz ES, McPheat J, Masimirembwa CM, and Andersson TB (2006) Evaluation of human liver slices and reporter gene assays as systems for predicting the cytochrome p450 induction potential of drugs in vivo in humans. *Pharm Res* **23:**56-69.

PMDA (2014) Drug Interaction Guideline for Drug Development and Labeling Recommendations (MHLW, 2014), updated 2017, English translation not yet available from PMDA.

Ramsden D, Zhou J, and Tweedie DJ (2015) Determination of a Degradation Constant for CYP3A4 by Direct Suppression of mRNA in a Novel Human Hepatocyte Model, HepatoPac. *Drug Metab Dispos* **43:**1307-1315.

Richert L, Liguori MJ, Abadie C, Heyd B, Mantion G, Halkic N, and Waring JF (2006) Gene expression in human hepatocytes in suspension after isolation is similar to the liver of origin, is not affected by hepatocyte cold storage and cryopreservation, but is strongly changed after hepatocyte plating. *Drug Metab Dispos* **34:**870-879.

Ruane PJ, Luber AD, Wire MB, Lou Y, Shelton MJ, Lancaster CT, Pappa KA, and Team COLS (2007) Plasma amprenavir pharmacokinetics and tolerability following administration of 1,400 milligrams of fosamprenavir once daily in combination with either 100 or 200 milligrams of ritonavir in healthy volunteers. *Antimicrob Agents Chemother* **51:**560-565.

Sane RS, Ramsden D, Sabo JP, Cooper C, Rowland L, Ting N, Whitcher-Johnstone A, and Tweedie DJ (2015) Contribution of major metabolites towards complex drug-drug interactions of deleobuvir: in vitro predictions and in vivo outcomes. *Drug Metab Dispos* **44:**466-475

Sinz M, Kim S, Zhu Z, Chen T, Anthony M, Dickinson K, and Rodrigues AD (2006a) Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Curr Drug Metab* **7:**375-388.

Sinz M, Kim S, Zhu Z, Chen T, Anthony M, Dickinson K, and Rodrigues AD (2006b) Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Curr Drug Metab* **7:**375-388.

Sun Y, Chothe PP, Sager JE, Tsao H, Moore A, Laitinen L, and Hariparsad N (2017) Quantitative Prediction of CYP3A4 Induction: Impact of Measured, Free, and Intracellular Perpetrator Concentrations from Human Hepatocyte Induction Studies on Drug-Drug Interaction Predictions. *Drug Metab Dispos* **45:**692-705.

Vieira ML, Kirby B, Ragueneau-Majlessi I, Galetin A, Chien JY, Einolf HJ, Fahmi OA, Fischer V, Fretland A, Grime K, Hall SD, Higgs R, Plowchalk D, Riley R, Seibert E, Skordos K, Snoeys J, Venkatakrishnan K, Waterhouse T, Obach RS, Berglund EG, Zhang L, Zhao P, Reynolds KS, and Huang SM (2014) Evaluation of various static in vitro-in vivo extrapolation models for risk assessment of the CYP3A inhibition potential of an investigational drug. *Clin Pharmacol Ther* **95:**189-198.

Zhang JG, Ho T, Callendrello AL, Clark RJ, Santone EA, Kinsman S, Xiao D, Fox LG, Einolf HJ, and Stresser DM (2014) Evaluation of calibration curve-based approaches to predict clinical inducers and noninducers of CYP3A4 with plated human hepatocytes. *Drug Metab Dispos* **42:**1379-1391.

Zhang JG, Ho T, Callendrello AL, Crespi CL, and Stresser DM (2010) A multi-endpoint evaluation of cytochrome P450 1A2, 2B6 and 3A4 induction response in human hepatocyte cultures after treatment with beta-naphthoflavone, phenobarbital and rifampicin. *Drug Metab Lett* **4:**185-194.

Zhang L, Zhang YD, Zhao P, and Huang SM (2009) Predicting drug-drug interactions: an FDA perspective. *The AAPS journal* **11:**300-306.

FOOTNOTES

Jane R Kenny, Diane Ramsden and David B. Buckley contributed equally to this work.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

FIGURE LEGENDS

Figure 1: Reproducibility of fold induction responses for CYP3A4 mRNA and CYP3A enzyme activity from repeat experiments with a single concentration of a negative (flumazenil; 25 μM) or positive (rifampicin; 10 or 20 μM) control. Values reported within each hepatocyte donor (H#) were generated in a single laboratory and represent fold induction responses collected across experiments conducted under the same experimental conditions. Data were collected from two different laboratories (donors H1-H10 and H11-H15, respectively). Individual flumazenil (negative control) data for (A) CYP3A4 mRNA (closed circles) and (B) CYP3A enzyme activity (open circles) were normally distributed and graphed on an arithmetic y-axis. Dotted lines represent a 2-fold change from vehicle control (0.5- and 2-fold). Individual rifampicin (positive control) data for (C) CYP3A4 mRNA (closed circles) and (D) CYP3A enzyme activity (open circles) were not normally distributed and graphed on a log-based y-axis. Dotted line represents 6-fold induction. Solid black lines represent mean fold induction values within each donor. Negative control responses were typically within 2-fold of the vehicle control response (0.5 – 2.0-fold) whereas rifampicin responses demonstrated marked intra-donor variability. The magnitude of the fold induction response was typically greater for mRNA compared to enzyme activity, whereas less intra-donor variation was observed for enzyme activity.

Figure 2: Reproducibility of CYP3A4 gene expression data from >60 individual repeated experiments conducted in a single laboratory with hepatocyte donor H2 as measured by real time PCR (TaqMan® RT-PCR). Cycle threshold (Ct) values for the housekeeping gene (GAPDH) and CYP3A4 were collected from each experiment for vehicle control (DMSO; 0.1%), negative control (flumazenil; 25 μM) and positive control (rifampicin; 20 μM) treatment groups. (A) Ct values for the housekeeping gene (GAPDH) over experimental repeat in chronological order. (B) CYP3A4 Ct values for vehicle and negative controls, rank-ordered by increasing vehicle CYP3A4 Ct values (Ct values are inversely proportional to transcript levels). (C)

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

CYP3A4 ΔCt values (ΔCt values; normalized to GAPDH) for vehicle, negative and positive controls rank-ordered by increasing vehicle (DMSO) CYP3A4 ΔCt values. (D) CYP3A4 mRNA fold-induction values (or ΔΔCt) for negative and positive controls rank-ordered by increasing vehicle CYP3A4 ΔCt values. (E) CYP3A4 ΔCt values (ΔCt values; normalized to GAPDH) for vehicle, negative and positive controls over experimental repeat in chronological order. (F) CYP3A4 mRNA fold-induction values (or ΔΔCt) for negative and positive controls over experimental repeat in chronological order. Vehicle control and flumazenil responses tracked across individual experiments and there was no correlation between flumazenil and rifampicin responses. In general, low basal expression of CYP3A4 mRNA resulted in higher rifampicin fold induction values. There was no trend in the response over time.

Figure 3: Impact of normalizing CYP3A4 mRNA fold-induction of test compound to positive control fold-induction (rifampicin). Emax for rifampicin and compound were used rather than response at a single maximum concentration. (A) shows data combined from three different sources using different donors and experimental conditions; IWG generated data for mild clinical inducers across three different laboratories and three different donors (felbamate, flucloxacillin, lersivirine, oxcarbazepine and rufinamide), literature data from Zhang et al 2014 using the same three donors in a single laboratory under the same experimental conditions, and IWG gathered proprietary compound data across different laboratories and experimental conditions in three different donors (untransformed fold-induction data Supplementary Figure 4). (B) shows rosiglitazone and pioglitazone in three donors that were repeated on five different occasions by the same laboratory under the same experimental conditions.

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 4: Rifampicin CYP3A4 mRNA for (A) EC_{50} and (B) E_{max} upon repeat experiments in 9 human hepatocyte donors. Data gathered from three different laboratories by the IQ IWG. Lines represent the median values.

Figure 5: *In vitro* human hepatocyte CYP3A4 mRNA induction data for 50 compounds for which clinical induction DDI data are available. (A) and (B) show EC_{50} while (C) and (D) show E_{max}. Compounds are arranged in order of ascending median *in vitro* induction potency (EC_{50}). Each point represents a distinct human hepatocyte donor. Data are from at least three different laboratories (sourced via IQ member company survey, from the literature, or generated by IQ induction group member companies specifically for this analysis). Compounds are grouped as exhibiting either *in vitro* induction only (A and C) or a combination of *in vitro* induction and inhibition (B and D). Color-coding is by median clinical AUC ratio (AUCR), where red represents strong DDI (induction AUCR <0.2 or inhibition AUCR >5), orange represents moderate DDI (induction AUCR 0.200-0.499 or inhibition AUCR 2.001–5.000), yellow represents mild DDI (induction AUCR 0.500–0.799 or inhibition AUCR 1.250–2.000), and green represents no DDI effect (AUCR within bioequivalence 0.800–1.249). Marker shapes distinguish median clinical induction effects as defined above: circles for induction, stars for bioequivalence, and squares for inhibition.

Figure 6: Clinical CYP3A DDI data for 51 compounds in order of ascending median victim drug AUC ratio (AUCR). Compounds are grouped as exhibiting either *in vitro* induction only (A) or a combination of *in vitro* induction and inhibition (B). Color-coding is by clinical AUCR, where red represents strong DDI (induction AUCR <0.2 or inhibition AUCR >5), orange represents moderate DDI (induction AUCR 0.200-0.499 or inhibition AUCR 2.001-5.000), yellow represents mild DDI (induction AUCR 0.500 – 0.799 or inhibition AUCR 1.250 – 2.000), and green represents no DDI effect (AUCR within bioequivalence 0.800-1.249). Triangles represent

DMD Fast Forward. Published on June 29, 2018 as DOI: 10.1124/dmd.118.081927
This article has not been copyedited and formatted. The final version may differ from this version.

midazolam or triazolam used as the clinical probe victim drug, while circles, stars, and squares represent induction, bioequivalence, or inhibition, respectively, for other clinical probe victim drugs.

Figure 7: Incidence of false positive (%FP) and false negative (%FN) predictions of DDI for 51 compounds compared with observed median CYP3A4 clinical DDI data using different IVIVE approaches (equations in Table 1) for (A) median induction *in vitro* parameters and (B) worst-case induction *in vitro* parameters

Figure 8: Summary of recommendations from the IQ IWG on CYP3A4 mRNA *in vitro* response thresholds, variability, and clinical relevance

Table 1. Equations

Equation designation	Parameter	Equation
A	R_3[a] value (2012 FDA and PMDA) [I][b] = $C_{max,ss}$	$R_3 = \frac{1}{(1 + d \times \frac{E_{max} \times [I]}{EC_{50} + [I]})}$
B	F2[c]	$F2 = \frac{2}{(Top - 1)^{\frac{1}{1} * EC_{50}}}$
C	R_3[a] value (2017 FDA and PMDA), [I][b] = $C_{max,ss,ub}$	$R_3 = \frac{1}{(1 + d \times \frac{E_{max} \times 10 \times [I]}{EC_{50} + 10 \times [I]})}$
D	Static mechanistic model	$\mathrm{AUCi/AUC} = \left(\frac{1}{[A_h \times B_h \times C_h] \times f_m + (1 - f_m)}\right) X \left(\frac{1}{[A_g \times B_g \times C_g] \times (1 - f_g) + f_g}\right)$
Da	A	$A_h = \frac{1}{1 + \frac{[I]_h}{K_i}} \quad A_g = \frac{1}{1 + \frac{[I]_g}{K_i}}$
Db	B	$B_h = \frac{k_{deg,h}}{k_{deg,h} + \frac{[I]_h \times k_{inact}}{[I]_h + K_I}} \quad B_g = \frac{k_{deg,g}}{k_{deg,g} + \frac{[I]_g \times k_{inact}}{[I]_g + K_I}}$
Dc	C (Induction only)	$C_h = 1 + \frac{d \times E_{max} \times [I]_h}{[I]_h + EC_{50}} \quad C_g = 1 + \frac{d \times E_{max} \times [I]_g}{[I]_g + EC_{50}}$
E	RIS (relative induction score) [I][b] = $C_{max,ss,ub}$	$\frac{E_{max} \times [I]}{EC_{50} + [I]}$
F	3-Parameter equation	$Y = Bottom + \frac{E_{max} - Bottom}{1 + 10^{Log\ EC_{50} - X}}$
G	Percent of prototypical inducer response	$\%PI = 100 \times \frac{Compound\ Signal - Blank\ Signal}{Total\ Signal - Blank\ Signal}$
H	4-Parameter equation	$Y = Bottom + \frac{E_{max} - Bottom}{1 + 10^{(Log\ EC_{50} - X) \times HS}}$
I	R_3 value using slope, [I][b] = $C_{max,ss}$	$R_3 = \frac{1}{1 + slope \times [I]}$
J	True positive (TP)	$\frac{TP}{TP + FN}$

K	True negative (TN)	$\frac{TN}{TN + FP}$
L	False negative (FN)	$\frac{FN}{TP + FN}$
M	False positive (FP)	$\frac{\boldsymbol{FP}}{\boldsymbol{TN + FP}}$

[a]R_3 = As described in the FDA and PMDA guidance, for in vitro induction characterization, the R value represents the ratio of the intrinsic clearance for an index substrate in the absence and presence of an inducer. Under the assumption that the intrinsic clearance is proportional to the total clearance the R value represents the AUC ratio in the presence and absence of the inducer.
[b]I = the concentration of the inducer used in the equation
[c]F2= the in vitro concentration where a 2-fold increase in mRNA is observed

Table 2. CYP3A4 mRNA levels (n = 314 experiments) and CYP3A enzyme activity (n = 111 experiments) in 10 hepatocyte donors following treatment with a single concentration of flumazenil (25 μM).

Flumazenil (25 μM; Negative Control)			H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Mean (All donors)
CYP3A4 mRNA	Fold Induction Response	*n (#)*	29	57	25	54	23	27	23	25	27	24	31.4
		Min	0.56	0.38	0.79	0.61	0.75	0.69	0.54	0.82	0.62	0.98	0.67
		Max	1.37	2.99	1.83	2.30	1.63	2.02	2.37	1.47	1.49	2.05	1.95
		Mean[a]	1.01	1.53	1.14	1.20	1.06	1.26	1.16	1.04	1.03	1.19	1.20
		SD[a]	0.19	0.57	0.25	0.35	0.24	0.38	0.45	0.17	0.23	0.29	0.40
		Limit of Blank[b]	1.31	2.47	1.55	1.78	1.45	1.89	1.90	1.32	1.41	1.65	1.86
		Limit of Detection[c]	1.62	3.41	1.97	2.35	1.84	2.52	2.64	1.61	1.79	2.12	2.52
	Probability of Exceeding 2-fold		0.0%	20%	0.0%	1.1%	0.0%	2.6%	3.1%	0.0%	0.0%	0.2%	2.3%
	Monte Carlo Simulations[d]	Negative control threshold	Frequency of a flumazenil (negative control) response ≥ 2.0-fold										
			In all 3 donors		In ≥ 2 donors		In ≥ 1 donor		Not Observed				
		2.0-fold	0.000%		0.117%		8.13%		91.9%				
			H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**	**H10**	**Mean (All Donors)**
CYP3A Activity	Fold Induction Response	*n (#)*	*6*	*8*	*15*	*24*	*13*	*8*	*4*	*10*	*8*	*15*	11.1
		Min	0.72	0.97	0.91	0.76	0.80	0.92	0.96	0.91	0.82	0.83	0.86
		Max	1.15	1.21	1.19	1.27	1.27	1.27	1.06	1.14	1.15	1.17	1.19
		Mean	0.95	1.07	1.05	1.03	1.04	1.09	1.00	1.03	0.99	1.01	1.03
		SD	0.16	0.09	0.09	0.11	0.14	0.13	0.04	0.08	0.11	0.10	0.11
		Limit of Blank[b]	1.22	1.22	1.20	1.21	1.26	1.30	1.07	1.17	1.17	1.17	1.20
		Limit of Detection[c]	1.48	1.38	1.34	1.40	1.49	1.50	1.14	1.31	1.36	1.33	1.37
	Probability of Exceeding 2-fold		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Flumazenil data available for H1-H10 only

n number of experiments per donor

SD Standard Deviation

[a] *Data normally distributed*

[b] *Limit of Blank (LOB) = Mean + (1.645 x SD negative control) --- 95% probability that a blank or negative control response falls below this value*

[c] *Limit of Detection = LOB + (1.645 x SD negative control) ---5% Type I and Type II error (false positive or negative)*

[d] *Monte Carlo Simulations conducted with 100,000 theoretical experiments each containing three random hepatocyte donors, rounded to 3 significant figures*

Table 3. CYP3A4 mRNA levels (*n* = 581 experiments) and CYP3A enzyme activity (*n* = 377 experiments) in 15 hepatocyte donors following treatment with a single concentration of rifampicin (10 or 20 μM).

Rifampicin (Positive Control)			H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15
CYP3A4 mRNA		*n (#)*	*43*	*65*	*24*	*64*	*35*	*31*	*31*	*41*	*36*	*24*	*46*	*70*	*13*	*43*	*15*
	Fold Induction Response	Min	3.7	4.2	7.4	3.3	3.6	4.5	5.3	3.2	5.0	14.2	9.4	6.4	20.3	8.5	6.6
		Median	11.9	13.4	17.3	29.9	7.1	10.5	8.5	7.6	18.7	75.0	31.6	13.0	31.8	19.6	12.0
		Max	42.0	71.9	92.1	137	40.9	47.7	22.6	52.8	58.9	134	89.0	26.8	68.9	35.2	42.0
		%CV	53.5	60.9	93.1	79	76.7	70.4	46.3	84	65.3	54.3	52.9	34	39.2	33.6	57.6
		Max/Min	11.4	17.1	12.4	41.5	11.4	10.6	4.3	16.5	11.8	9.4	9.5	4.2	3.4	4.1	6.4
	Probability of exceeding X-fold response[a]	> 6-fold	92%	96%	92%	98%	70%	87%	87%	70%	97%	100%	100%	89%	100%	100%	97%
		> 10-fold	66%	77%	78%	91%	35%	61%	47%	34%	87%	100%	98%	80%	100%	95%	73%
		> 20-fold	17%	26%	47%	67%	5%	19%	4%	5%	50%	94%	80%	40%	94%	41%	16%
	Monte Carlo Simulations[b]	Positive control threshold	Frequency of a rifampicin response ≥ X-fold														
			In all 3 donors		In ≥ 2 donors		In ≥ 1 donor		Not Observed								
		6-fold	78.4%		98.5%		99.9%		0.04%								
		10-fold	40.9%		84.5%		98.6%		1.39%								
		20-fold	4.94%		32.2%		77.9%		22.1%								
			H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**	**H10**	**H11**	**H12**	**H13**	**H14**	**H15**
CYP3A Activity		*n (#)*	*15*	*37*	*15*	*27*	*13*	*13*	*19*	*21*	*17*	*13*	*46*	*70*	*13*	*43*	*15*
	Fold Induction Response	Min	4.5	1.1	2.2	3.0	2.7	3.0	2.3	4.0	9.1	3.4	5.7	3.1	6.4	5.1	2.9
		Median	7.0	5.3	6.7	14.1	4.2	6.3	3.6	6.4	18.1	9.2	13.5	6.2	9.8	10.0	4.2
		Max	11.1	19.8	9.7	29.4	7.3	11.2	5.3	12.6	34.7	19.1	27.6	12.0	15.3	15.2	5.5
		%CV	27.2	61.7	29.3	44.5	29.9	38.4	21.5	34	40.7	38.4	34.7	26	25.4	18.2	21.2
		Max/Min	2.5	18	4.4	9.8	2.7	3.7	2.3	3.2	3.8	5.6	4.8	3.9	2.4	3.0	1.9
	Probability of exceeding 6-fold response[a]		65%	32%	60%	88%	10%	60%	1%	64%	100%	84%	99%	54%	97%	94%	5%

Donors H1 – H10 and donors H11 – H15 were treated with 20 or 10 μM rifampicin, respectively.

n number of experiments per donor

SD Standard deviation

[a] *Data not normally distributed. Probabilities derived from mean and standard deviation of log-transformed data*

[b] *Monte Carlo Simulations conducted with 100,000 theoretical experiments each containing three random hepatocyte donors, rounded to 3 significant figures*

Table 4. Clinical induction risk assessment equations and incidence of false positive and false negative prediction based on different approaches to IVIVE in regulatory guidance using median observed clinical AUCR.

Regulator	Equation	Input [Inducer]	% False Negative		% False Positive		% within 2-fold		% within BE	
			Median	Worst case	Median	Worst case	Median	Worst case	Median	Worst case
EMA	F2 (50-fold $C_{max,ss,u}$)	$C_{max,ss,u}$	14.8[a]	11.1[b]	82.6	87.0	NC	NC	NC	NC
	F2 (30-fold $C_{max,ss,u}$)	$C_{max,ss,u}$	14.8[a]	14.8[a]	69.6	78.3	NC	NC	NC	NC
	F2 (0.25 gut)	0.1 * Dose/250 mL	5.3[c]	5.3[c]	80.0	93.3	NC	NC	NC	NC
	RIS	$C_{max,ss,u}$	7.4[d]	0	82.6	87.0	38	36	20	12
		$Portal_{,ss,u}$	0	0	93.3	100	42	39	24	15
FDA *current*	R3 = 0.9	$C_{max,total}$	3.7[c]	3.7[c]	95.7	95.7	26	16	12	4
		$C_{max,ss,u}$	25.9[e]	18.5[f]	60.9	73.9	72	54	28	18
		$C_{max,ss,u}$ Fu = 0.01	14.8[g]	7.4[h]	60.9	73.9	72	56	32	22
		$C_{av,total}$	11.1[i]	11.1[i]	78.6	92.9	47	28	25	6
		$C_{av,ss,u}$	33.3[j]	22.1[k]	35.7	50	94	81	31	31
		0.1 * Dose/250 mL	0	0	100	100	NC	NC	NC	NC
	R3 = 0.8, d = 1	$C_{max,total}$	7.4[i]	3.7[c]	87.0	91.3	26	16	12	4
	R3 = 0.95, d = 1	$C_{max,ss,u}$	14.8[a]	11.1[b]	73.9	87.0	72	54	28	18
	R3 = 0.9, d = 0.3	$C_{max,total}$	7.4[i]	7.4[i]	82.6	91.3	60	46	26	12
FDA and PMDA proposed 2017	R3 = 0.8, d = 1, SF = 10X	$C_{max,ss,u}$	11.1[b]	7.4[l]	78.3	87.0	36	30	18	12
	R3 = 0.8, d = 1, SF = 50X	$C_{max,ss,u}$	0	0	91.3	95.7	24	20	8	0
	R3 = 0.9, calculated from slope	$C_{max,total}$	3.7[c]	3.7[c]	95.7	95.7	20	12	10	4
	R3 = 0.9, calculated from slope	$C_{max,ss,u}$	25.9[e]	18.5[f]	60.9	73.9	68	52	28	14
	Slope correlation	NA	0	0	100	100	54	52	18	22
Mechanistic static model	Induction only	$Portal_{,ss,u}$ and Igut ($Fa*Ka*Dose/Q_{en}$)	5.3[c]	5.3[c]	100	100	12	6	3	3

	$C_{max,ss,u}$ and I gut = $Portal_{,ss,u}$	10.5[h]	10.5[h]	46.7	66.7	59	44	26	12
Induction + Inhibition	$Portal_{,ssu}$ and Igut ($Fa*Ka*Dose/Q_{en}$)	36.4[m]	36.4[m]	16.7	25.0	30	26	22	22
	$C_{max,ss,u}$ and I gut = $Portal_{,ss,u}$	36.4[m]	27.3[n]	16.7	25.0	43	48	17	13

[a]dexamethasone, pleconaril, Cmpd 2, Cmpd 15, [b]dexamethasone, pleconaril, Cmpd 15,[c]dexamethasone, [d]pleconaril, Cmpd 15, [e]dexamethasone, flucloxacillin, oxcarbazepine, pleconaril, Cmpd 2, Cmpd 11, Cmpd 15, [f] dexamethasone, pleconaril, Cmpd 2, Cmpd 11, Cmpd 15, [g]dexamethasone, flucloxacillin, oxcarbazepine, pleconaril, [h]dexamethasone, pleconaril, [i]dexamethasone, oxcarbazepine, [j]clobazam, dexamethasone, efavirenz, flucloxacillin, oxcarbazepine, pleconaril, [k]dexamethasone, efavirenz, oxcarbazepine, pleconaril

[l]dexamethasone, Cmpd 15, [m]dexamethasone, lopinavir, nevirapine, troglitazone, [n]dexamethasone, lopinavir, nevirapine

Figure 1

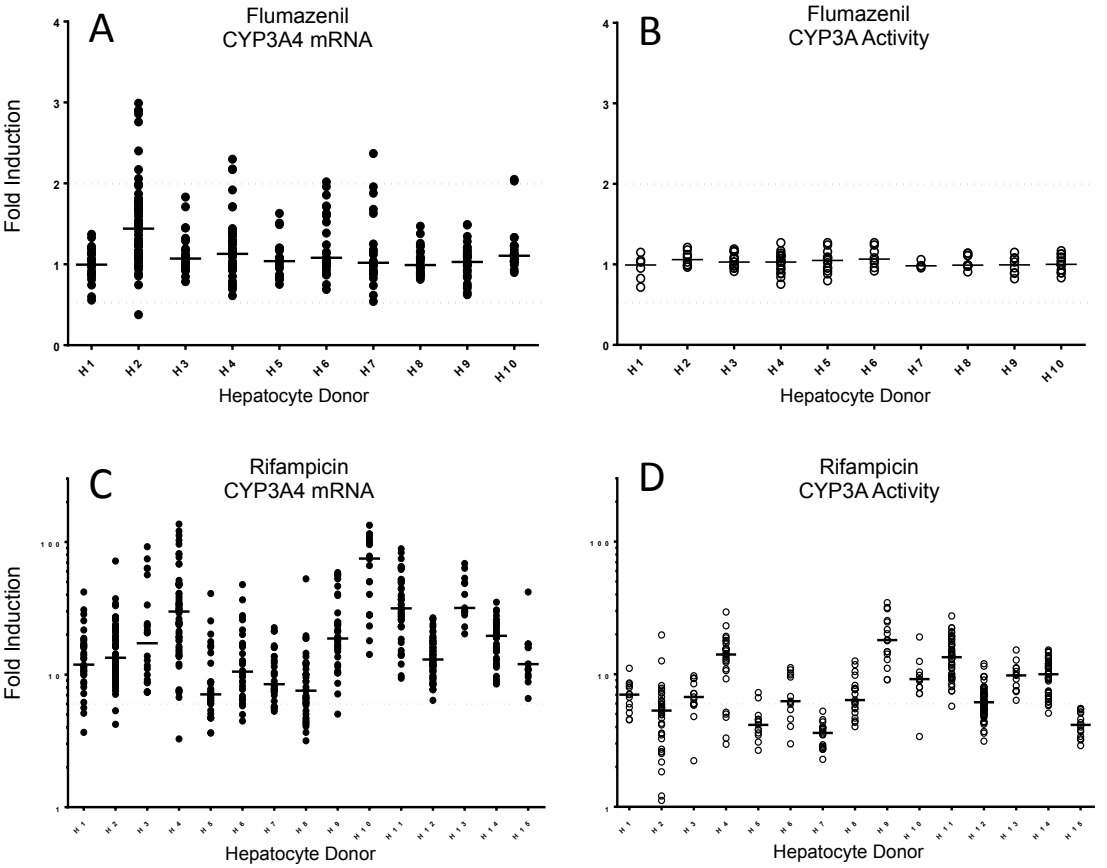

A
Flumazenil
CYP3A4 mRNA
Fold Induction
Hepatocyte Donor
B
Flumazenil
CYP3A Activity
Hepatocyte Donor
C
Rifampicin
CYP3A4 mRNA
Fold Induction
Hepatocyte Donor
D
Rifampicin
CYP3A Activity
Hepatocyte Donor

Figure 2

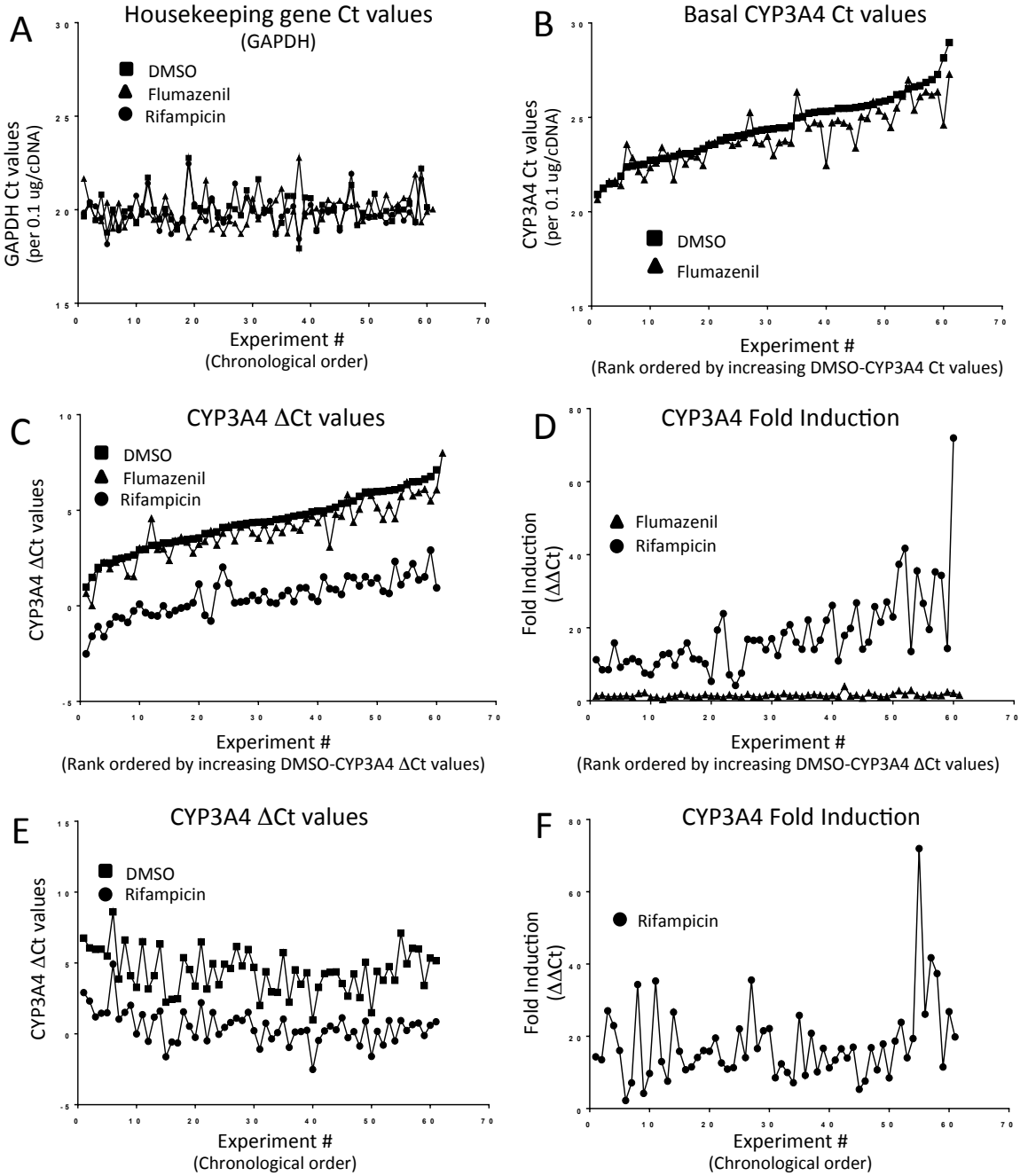
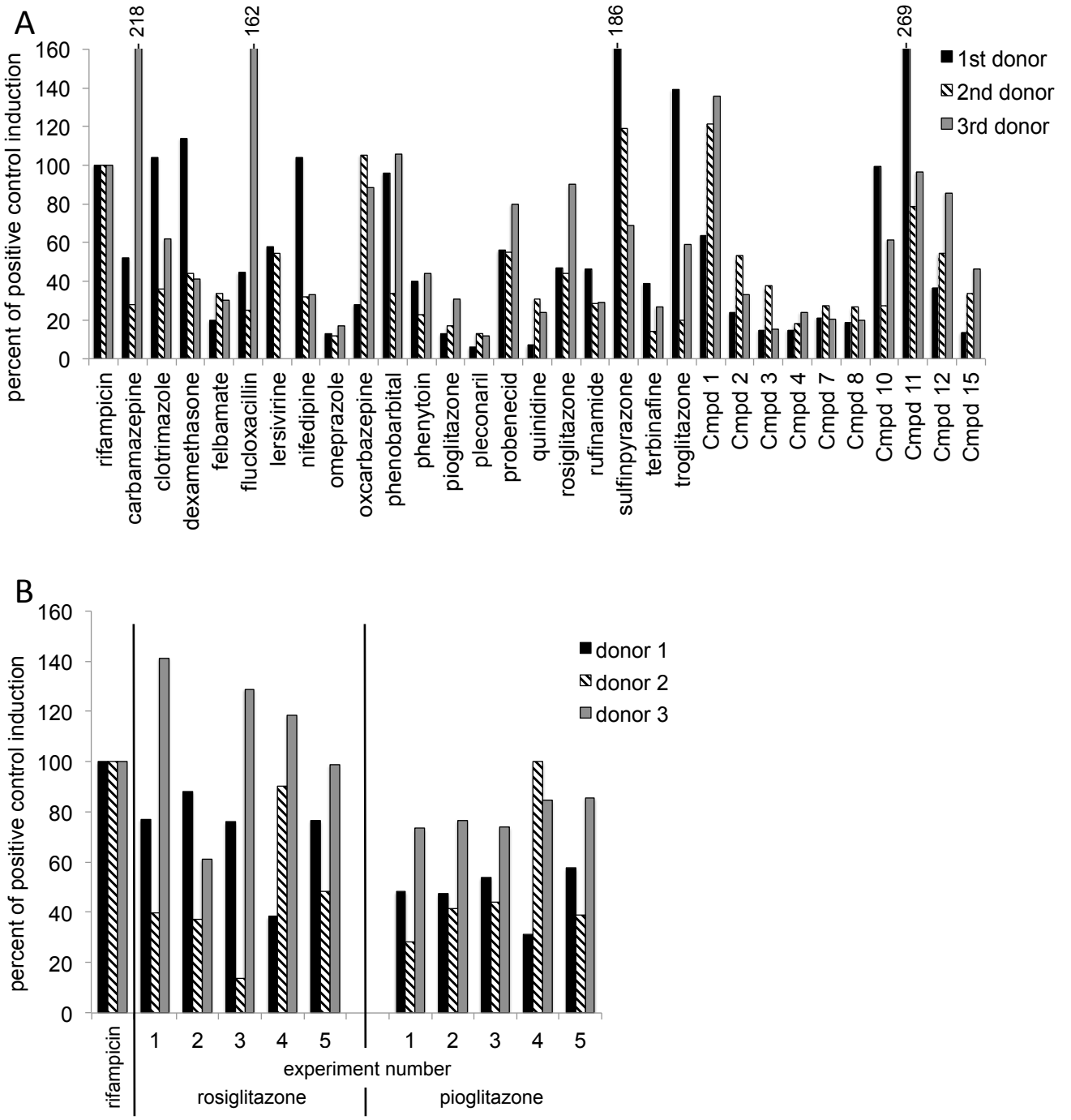

A
Housekeeping gene Ct values
(GAPDH)
DMSO
Flumazenil
Rifampicin
GAPDH Ct values
(per 0.1 ug/cDNA)
Experiment #
(Chronological order)
B
Basal CYP3A4 Ct values
CYP3A4 Ct values
(per 0.1 ug/cDNA)
DMSO
Flumazenil
Experiment #
(Rank ordered by increasing DMSO-CYP3A4 Ct values)
C
CYP3A4 ΔCt values
DMSO
Flumazenil
Rifampicin
CYP3A4 ΔCt values
Experiment #
(Rank ordered by increasing DMSO-CYP3A4 ΔCt values)
D
CYP3A4 Fold Induction
Flumazenil
Rifampicin
Fold Induction
(ΔΔCt)
Experiment #
(Rank ordered by increasing DMSO-CYP3A4 ΔCt values)
E
CYP3A4 ΔCt values
DMSO
Rifampicin
CYP3A4 ΔCt values
Experiment #
(Chronological order)
F
CYP3A4 Fold Induction
Rifampicin
Fold Induction
(ΔΔCt)
Experiment #
(Chronological order)

Figure 3


A
218
162
186
269
percent of positive control induction
160
140
120
100
80
60
40
20
0
1st donor
2nd donor
3rd donor
rifampicin
carbamazepine
clotrimazole
dexamethasone
felbamate
flucloxacillin
lersivirine
nifedipine
omeprazole
oxcarbazepine
phenobarbital
phenytoin
pioglitazone
pleconaril
probenecid
quinidine
rosiglitazone
rufinamide
sulfinpyrazone
terbinafine
troglitazone
Cmpd 1
Cmpd 2
Cmpd 3
Cmpd 4
Cmpd 7
Cmpd 8
Cmpd 10
Cmpd 11
Cmpd 12
Cmpd 15

B
percent of positive control induction
160
140
120
100
80
60
40
20
0
donor 1
donor 2
donor 3
rifampicin
1
2
3
4
5
1
2
3
4
5
experiment number
rosiglitazone
pioglitazone

Figure 4

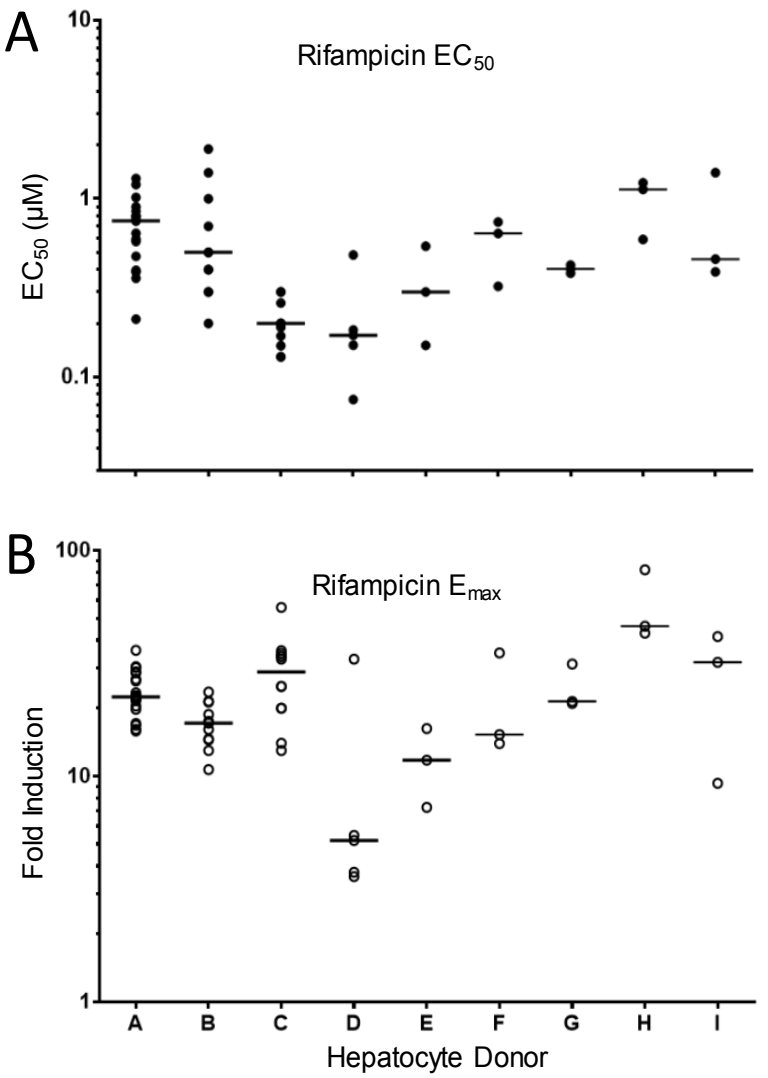

A
Rifampicin EC50
EC50 (μM)
10
1
0.1
B
Rifampicin Emax
Fold Induction
100
10
1
A
B
C
D
E
F
G
H
I
Hepatocyte Donor

Figure 5

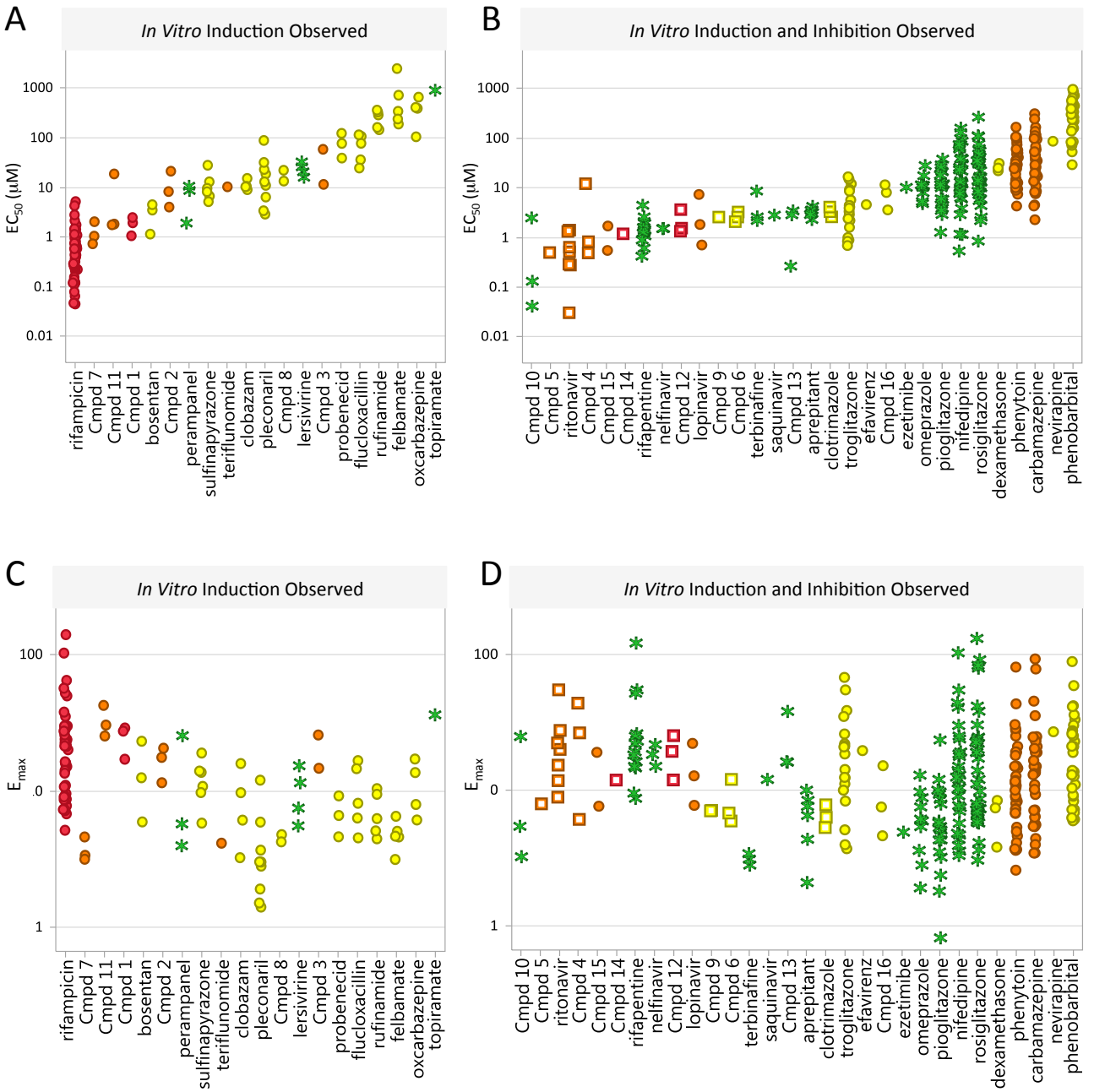

A
In Vitro Induction Observed
EC50 (µM)
1000
100
10
1
0.1
0.01
rifampicin
Cmpd 7
Cmpd 11
Cmpd 1
bosentan
Cmpd 2
perampanel
sulfinapyrazone
teriflunomide
clobazam
pleconaril
Cmpd 8
lersivirine
Cmpd 3
probenecid
flucloxacillin
rufinamide
felbamate
oxcarbazepine
topiramate
B
In Vitro Induction and Inhibition Observed
EC50 (µM)
1000
100
10
1
0.1
0.01
Cmpd 10
Cmpd 5
ritonavir
Cmpd 4
Cmpd 15
Cmpd 14
rifapentine
nelfinavir
Cmpd 12
lopinavir
Cmpd 9
Cmpd 6
terbinafine
saquinavir
Cmpd 13
aprepitant
clotrimazole
troglitazone
efavirenz
Cmpd 16
ezetimibe
omeprazole
pioglitazone
nifedipine
rosiglitazone
dexamethasone
phenytoin
carbamazepine
nevirapine
phenobarbital
C
In Vitro Induction Observed
Emax
100
0
1
rifampicin
Cmpd 7
Cmpd 11
Cmpd 1
bosentan
Cmpd 2
perampanel
sulfinapyrazone
teriflunomide
clobazam
pleconaril
Cmpd 8
lersivirine
Cmpd 3
probenecid
flucloxacillin
rufinamide
felbamate
oxcarbazepine
topiramate
D
In Vitro Induction and Inhibition Observed
Emax
100
0
1
Cmpd 10
Cmpd 5
ritonavir
Cmpd 4
Cmpd 15
Cmpd 14
rifapentine
nelfinavir
Cmpd 12
lopinavir
Cmpd 9
Cmpd 6
terbinafine
saquinavir
Cmpd 13
aprepitant
clotrimazole
troglitazone
efavirenz
Cmpd 16
ezetimibe
omeprazole
pioglitazone
nifedipine
rosiglitazone
dexamethasone
phenytoin
carbamazepine
nevirapine
phenobarbital

Figure 6

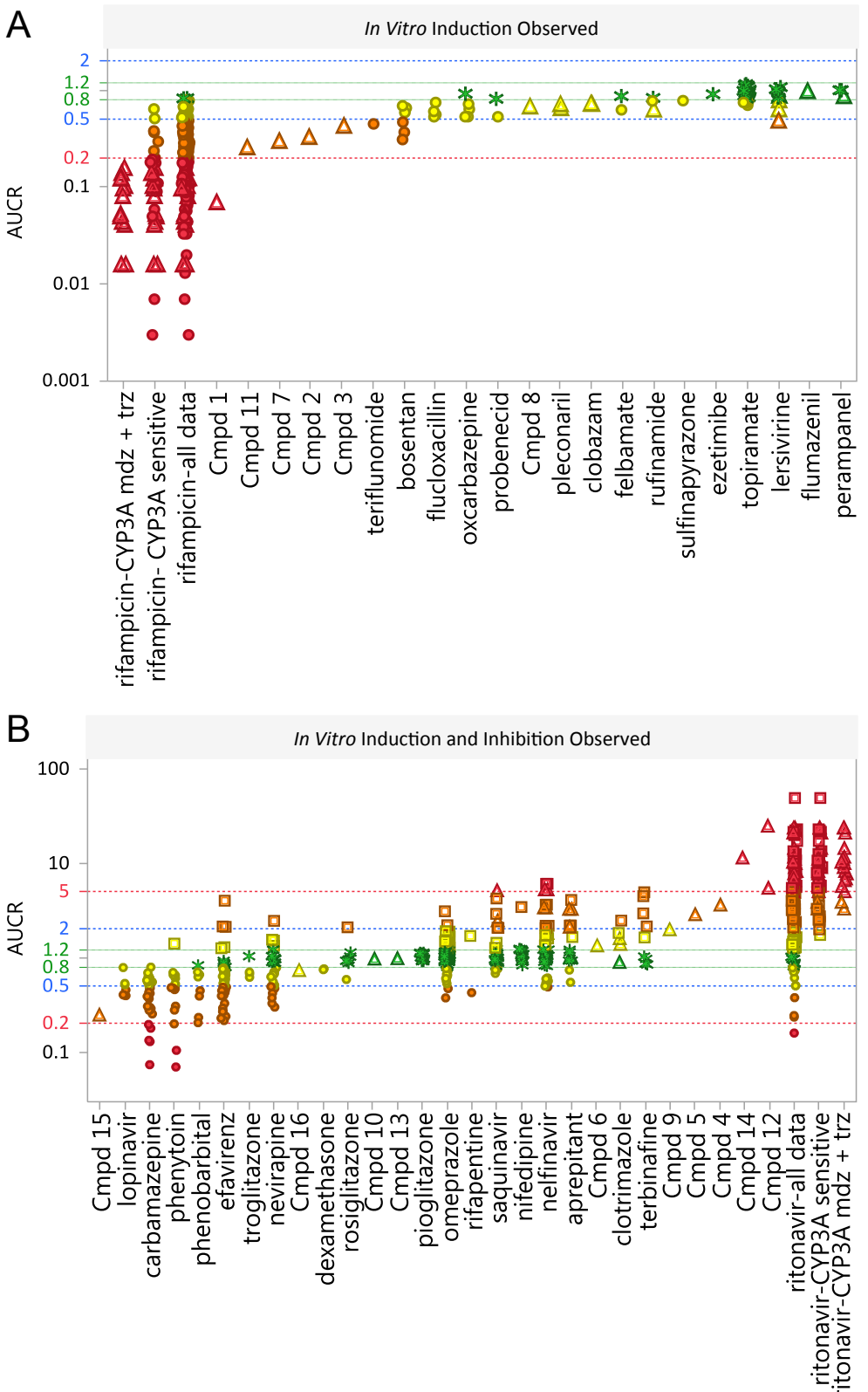
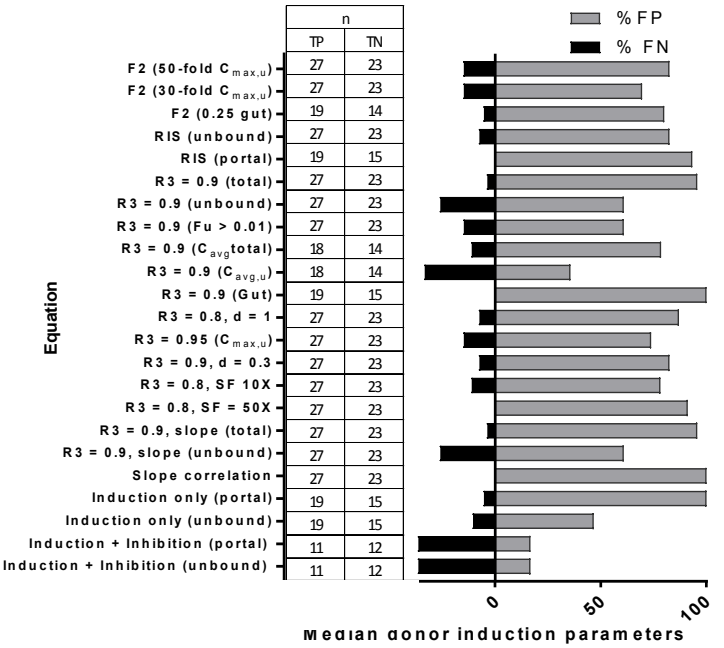

A
In Vitro Induction Observed
AUCR
2
1.2
0.8
0.5
0.2
0.1
0.01
0.001
rifampicin-CYP3A mdz + trz
rifampicin- CYP3A sensitive
rifampicin-all data
Cmpd 1
Cmpd 11
Cmpd 7
Cmpd 2
Cmpd 3
teriflunomide
bosentan
flucloxacillin
oxcarbazepine
probenecid
Cmpd 8
pleconaril
clobazam
felbamate
rufinamide
sulfinapyrazone
ezetimibe
topiramate
lersivirine
flumazenil
perampanel
B
In Vitro Induction and Inhibition Observed
AUCR
100
10
5
2
1.2
0.8
0.5
0.2
0.1
Cmpd 15
lopinavir
carbamazepine
phenytoin
phenobarbital
efavirenz
troglitazone
nevirapine
Cmpd 16
dexamethasone
rosiglitazone
Cmpd 10
Cmpd 13
pioglitazone
omeprazole
rifapentine
saquinavir
nifedipine
nelfinavir
aprepitant
Cmpd 6
clotrimazole
terbinafine
Cmpd 9
Cmpd 5
Cmpd 4
Cmpd 14
Cmpd 12
ritonavir-all data
ritonavir-CYP3A sensitive
ritonavir-CYP3A mdz + trz

Figure 7

A


n
TP
TN
% FP
% FN
F2 (50-fold C max,u) 27 23
F2 (30-fold C max,u) 27 23
F2 (0.25 gut) 19 14
RIS (unbound) 27 23
RIS (portal) 19 15
R3 = 0.9 (total) 27 23
R3 = 0.9 (unbound) 27 23
R3 = 0.9 (Fu > 0.01) 27 23
R3 = 0.9 (C avg total) 18 14
R3 = 0.9 (C avg,u) 18 14
R3 = 0.9 (Gut) 19 15
R3 = 0.8, d = 1 27 23
R3 = 0.95 (C max,u) 27 23
R3 = 0.9, d = 0.3 27 23
R3 = 0.8, SF 10X 27 23
R3 = 0.8, SF = 50X 27 23
R3 = 0.9, slope (total) 27 23
R3 = 0.9, slope (unbound) 27 23
Slope correlation 27 23
Induction only (portal) 19 15
Induction only (unbound) 19 15
Induction + Inhibition (portal) 11 12
Induction + Inhibition (unbound) 11 12
Equation
0
50
100
Median donor induction parameters

B

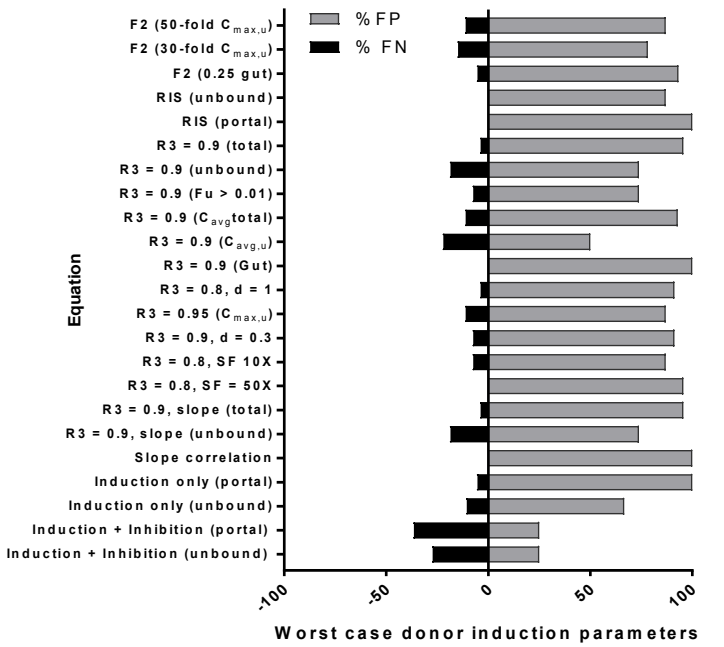

% FP
% FN
F2 (50-fold C max,u)
F2 (30-fold C max,u)
F2 (0.25 gut)
RIS (unbound)
RIS (portal)
R3 = 0.9 (total)
R3 = 0.9 (unbound)
R3 = 0.9 (Fu > 0.01)
R3 = 0.9 (C avg total)
R3 = 0.9 (C avg,u)
R3 = 0.9 (Gut)
R3 = 0.8, d = 1
R3 = 0.95 (C max,u)
R3 = 0.9, d = 0.3
R3 = 0.8, SF 10X
R3 = 0.8, SF = 50X
R3 = 0.9, slope (total)
R3 = 0.9, slope (unbound)
Slope correlation
Induction only (portal)
Induction only (unbound)
Induction + Inhibition (portal)
Induction + Inhibition (unbound)
Equation
-100
-50
0
50
100
Worst case donor induction parameters

Figure 8

- CYP induction should continue to be evaluated in three separate human donors in vitro.
- In light of empirically divergent responses in rifampicin control and most test inducers, normalization of data to percent positive control appears to be of limited benefit.
- Two-fold induction, with concentration dependence, is an acceptable threshold for positive identification of in vitro CYP3A4 mRNA induction.
- To reduce the risk of false positives, in the absence of a concentration dependent response, induction ≥ 2-fold should be observed in more than one donor to classify a compound as an in vitro inducer.
- If qualifying a compound as negative for CYP3A4 mRNA induction, the magnitude of maximal rifampicin response in that donor should be ≥10-fold.
- Inclusion of a negative control adds no value beyond that of the vehicle control.