

Review Article

Estimating the predictive validity of diabetic animal models in rosiglitazone studies

O. E. Varga^{1,2}, N. Zsíros³ and I. A. S. Olsson¹

¹Laboratory Animal Science group, IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal;

²Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary; ³1st Department of Internal Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

Received 18 November 2014; revised 10 January 2015; accepted 2 February 2015

Address for correspondence: Dr OE Varga, Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Kassai út 26, Debrecen 4028, Hungary.
E-mail: varga.orsolya@sph.unideb.hu

Summary

For therapeutic studies, predictive validity of animal models – arguably the most important feature of animal models in terms of human relevance – can be calculated retrospectively by obtaining data on treatment efficacy from human and animal trials. Using rosiglitazone as a case study, we aim to determine the predictive validity of animal models of diabetes, by analysing which models perform most similarly to humans during rosiglitazone treatment in terms of changes in standard diabetes diagnosis parameters (glycosylated haemoglobin [HbA1c] and fasting glucose levels). A further objective of this paper was to explore the impact of four covariates on the predictive capacity: (i) diabetes induction method; (ii) drug administration route; (iii) sex of animals and (iv) diet during the experiments. Despite the variable consistency of animal species-based models with the human reference for glucose and HbA1c treatment effects, our results show that glucose and HbA1c treatment effects in rats agreed better with the expected values based on human data than in other species. Induction method was also found to be a substantial factor affecting animal model performance. The study concluded that regular reassessment of animal models can help to identify human relevance of each model and adapt research design for actual research goals.

Keywords: Animal models, human relevance, predictive validity, type 2 diabetes.

obesity reviews (2015)

Introduction

Although animal research is considered to be a central element of contemporary biomedical science and arguably has contributed greatly to the understanding of disease mechanisms and development of treatments, the predictive validity of different animal models is generally assumed and rarely measured. Having information on this feature of animals is crucial especially for therapeutic studies. The predictive validity of animal models, which means ‘to what extent research data from animals can predict human response to particular drugs’ can be calculated retrospectively, after obtaining data on treatment efficacy from humans and animals.

Even though only human–animal comparative studies can produce evidence on human relevance of animal models and justify their use from a scientific point of view, relatively few studies have addressed their methodology. Several animal studies speculate on the translatability of results from animal studies by assuming on the validity of animal models. However, unique characteristics of animal models deserve special consideration on whether results can be translated to humans or not (1). As Garth Whiteside has noted, comparison of large datasets from preclinical efficacy studies and human trials is a complex matter, and the predictive capacity of these models cannot be described as ‘worked’, ‘didn’t work’, or ‘failed’; deeper analysis is needed (2).

To evaluate the predictive validity of animal models, one may opt for (i) assessing the predictive validity of a single or very few models by comparing treatment effects of series of interventions (e.g. administration of anti-diabetic drugs) in that/those animal model(s) and humans or (ii) assessing the predictive validity of several animal models by comparing the treatment effect of a particular intervention in those animals and humans. In either case, this comparison should be based on quantitative data that allow correlations between humans and different animal species to be calculated. This requires access to data where the same quantitative information (outcomes) is available for treatment effects both in animal and human subjects. Although research outcomes of an identical nature are regularly reported in human and animal studies, any study selection effort entails the setting of minimum criteria for study design quality. Furthermore, data availability is also an issue.

The quality of the data has a huge impact on the conclusions that can be drawn, but raw data are rarely presented. The ideal human data for the calculation would be complete data from clinical trials. Because of the recent tight regulations on clinical trials (e.g. requirement of authorization (3) and prior registration (4)), transparency on the conduct and results of clinical trials has been improved. However, there are still problems with study design and publication from these trials. Regarding design, a widely discussed issue, among others, is the obscure management of missing data (5). Additionally, we know that not all studies get published in their entirety after the clinical trials (6); what is more, what does get published may be different from in-house interpretation and more likely to make a drug look favourable (7).

Also for animal data, it is increasingly evident that shortcomings in research design and publication bias resulting from selective publishing of desirable results are the cause of overestimated treatment effects (8,9). The retrospective evaluation of the predictive validity of animal models is further complicated by statistical weaknesses. Animal studies regularly report data from small samples of animals, and usually such studies are not repeated by independent third-party laboratories.

An additional challenge is that of comparing different species (10). The efficient drug dose widely varies across species, mostly because of the pharmacokinetics of a particular drug being different from species to species (11). It is commonly observed that small animals need to be administered larger doses (per kilogram body weight) as compared with big animals or humans to achieve similar pharmacological effects. For instance, about fivefold higher doses of prednisolone and caffeine have been reported for rats as compared with humans (12). Providing cross-species comparisons of activity and toxicity of various drugs, two important methods are used for dosage conversions. One is

based on per body surface area calculation (mg/m^2), which is the method required by the Food and Drug Administration (FDA); the alternative method considers the daily expenditure of energy expressed per metabolically active mass (12).

In the present paper, we propose a method for assessing the predictive validity of several animal models. Using rosiglitazone, a widely used pharmaceutical to treat type 2 diabetes mellitus (T2DM), we aimed to provide data on the predictive validity of different animal models of diabetes. Type 2 diabetes was chosen as animal models are widely used in research into this disease (13), but their predictive validity has never been statistically studied. Rosiglitazone is an ideal case study to test the predictive validity of diabetes animal models: it is widely used in human patients and a preliminary PubMed search showed it to be the most commonly used pharmaceutical in animal studies into type 2 diabetes.

The main objective of this paper was to determine the predictive validity of animal models of diabetes, by analysing which models perform most similarly to humans during rosiglitazone treatment in terms of changes in standard diabetes diagnosis parameters (glycosylated haemoglobin [HbA1c] and fasting glucose [FG] levels). A further objective of this paper was to explore the impact of four covariates on the predictive capacity of animal models. These covariates are methodological issues that often differ across studies, namely (i) diabetes induction method; (ii) drug administration route; (iii) sex of animals and (iv) diet during the experiments.

Materials and methods

Literature review

Both human and animal studies were searched between September and December 2012. Studies reporting rosiglitazone monotherapies with information on glucose and/or HbA1c, published in English, were included. All references were downloaded and managed in Endnote. Two authors (O. E. V., N. Z.) assessed studies and extracted data into an excel table (Microsoft Office Excel 2010). Data on study design elements including time, route and dose of the drug administration, the species and strain of the animal, age and sex/gender of subjects, diets, diabetes induction method, outcomes (i.e. FG and HbA1c levels – number of observations, mean, variability measure) in each study group were extracted. In those papers where data were only reported graphically, a digital online ruler (14) was used to gain numerical values.

Animal studies were identified from Pubmed and Web of Science using the following algorithms. Pubmed search: ('animal experimentation'[MeSH Terms] OR 'models, animal' [MeSH Terms]) AND 'rosiglitazone' [Supplemen-

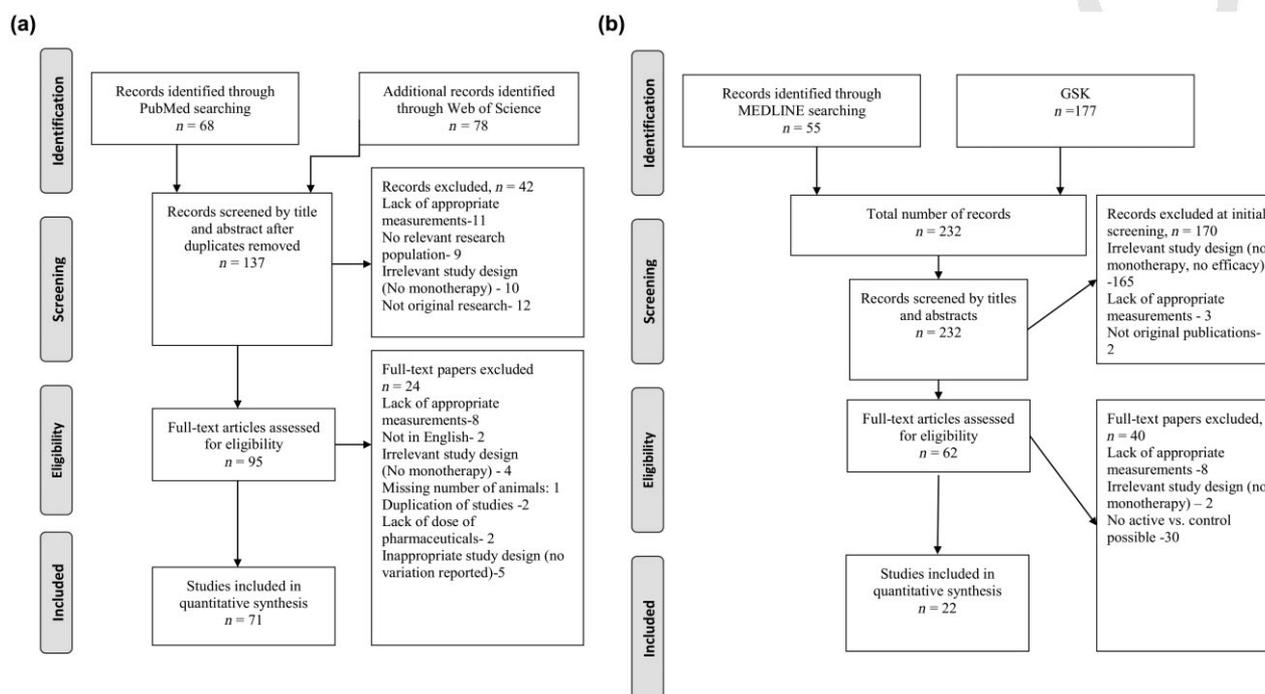


Figure 1 Procedure on selection of studies. (a) Procedure chart on how animal studies were selected. (b) Procedure chart on how human studies were selected.

tary Concept] AND 'blood glucose' [MeSH Terms]. Web of Science: Topic = (rosiglitazone) AND Topic = (blood glucose) AND Topic = (animal) AND Topic = (search filter suggested by Carlijn R Hooijmans (15)). The selection method with exclusion criteria is presented in Fig. 1a. A total of 71 studies were included.

Human studies were identified from three sources. A PubMed search that used the algorithm: 'Blood Glucose' [Mesh] AND 'Hemoglobin A, Glycosylated' [Mesh] AND 'Diabetes Mellitus, Type 2/drug therapy' [Mesh] AND 'rosiglitazone' [Supplementary Concept] AND 'Thiazolidinediones/therapeutic use'[Mesh] AND 'Randomized Controlled Trial' [ptyp] AND English [lang]. In addition, one study (16) from a published metaanalysis on the efficacy of thiazolidinediones (the class of drugs that include rosiglitazone) in the Asian population (17) was included. Finally, all monotherapy studies were identified and included in the analysis from the website of the pharmaceutical company GlaxoSmithKline (GSK) who made all studies on rosiglitazone available through the company's website (18). Details along with exclusion criteria are shown in Fig. 1b.

Parameters and factors analysed

To evaluate the predictive capacity of different animal models, human and animal HbA1c and FG levels were selected as outcome measures. HbA1c is the primary

laboratory test for diabetes in human studies and reflects average blood glucose for the preceding 60–90 d, whereas FG level is a very common parameter to monitor diabetes.

To evaluate the effect of certain factors that often differ across studies and may cause methodological issues, the following were considered: (i) diabetes induction method; (ii) drug administration route; (iii) sex of animals and (iv) diet during experiments. Rationale of selecting these covariates is given later.

1. A number of diabetic animal models have been developed over the years, based mostly on rodents; these models can be classified into two broad categories: (i) genetically induced spontaneous diabetes models and (ii) experimentally induced (non-spontaneous) diabetes models (19). The second consists of several subtypes: streptozotocin (STZ)/alloxan models, partial pancreatectomy models, high-fat (HF)/high-sucrose diet-fed models, HF diet-fed STZ models, and intrauterine growth retardation models (19).

2. Several drug administration methods have been used to introduce the chemical substance. Oral administration, subcutaneous administration and intraperitoneal injection of substances are common procedures in scientific experiments (20).

3. As diet has a significant impact on diabetes's induction and progress, diet is a crucial part of the experimentation, whether spontaneous or experimentally induced models are used (21).

4. The sex of animals has a well-documented impact on diabetes mellitus progress, which may imply gender-specific clinical treatment of diabetes (22).

Statistical modelling and analysis

Information from all included studies was extracted and entered into a single database with as many observations for each study as the number of FG and/or HbA1c outcome measurements reported at distinct follow-up times in that study. Outcome level means and their standard errors were logarithmically transformed (natural base) using appropriate formulas to derive the expected value and standard error of the transformed variable working from the mean and standard error of the source variable. Treatment effect estimates at each observation time were calculated as the between-groups (rosiglitazone versus placebo) difference in transformed outcome levels. Treatment effect standard errors were calculated as the square root of the sum of group-specific squared standard errors. For the analysis of species effect, only rats and mice were included, whereas the remaining analyses were done on the complete dataset, including the two single studies on hamster and gerbil. For this reason animal strains in Fig. 2a,b, are referred to as 'Rat' and 'Mouse'.

Linear regression models were fitted on human observations for both outcomes. The dependent variable was the treatment effect; explanatory variables included dose, time into exposure (natural logarithm of week) complete with a quadratic derivative, and interaction terms between dose and time. Observations were weighted proportionally to their precision (reciprocal of squared standard error of treatment effect). The distribution of dose values was bimodal; a low- vs. high-dose categorization cut-off of 6.5 mg was observed to coincide with a fairly wide gap between the two modes, close to the sample 50th percentile. To verify that this categorization did not cause substantial loss to model fit and/or changes to results and conclusions, the analysis was also completed using dose as a continuous variable, transforming rat doses (mg intake) to a human-comparable scale dividing by 2.25, and non-rat murine doses by dividing by 11.25, in accordance with FDA guidelines regarding interspecies dose conversion (3). Model fit was evaluated using normality tests of residuals and Ramsey's regression specification-error tests, neither of which indicated any insufficiency of fit (all $P \geq 0.195$).

Coefficients derived from the human models were used to calculate expected values of non-human treatment effects. For each included study, and for each level of strain, diabetes induction method, special diet (yes or no), sex and drug administration route within the study if applicable, differences between observed and expected treatment effects were squared, summed, divided by the number of

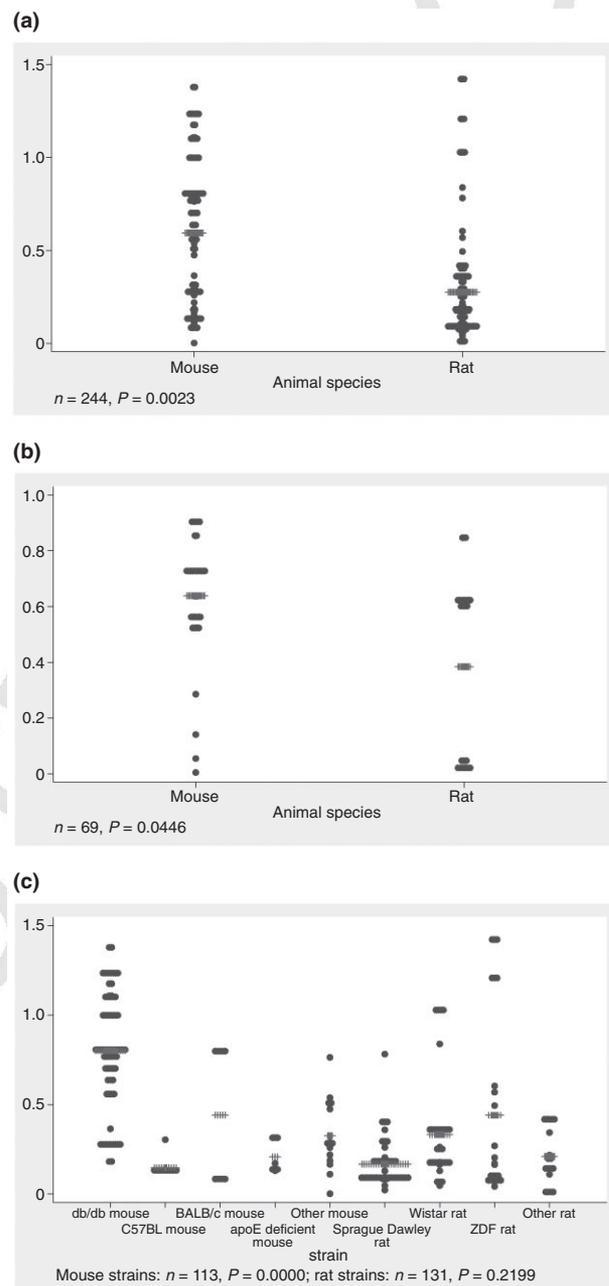


Figure 2 Comparing the animal models with the human reference for glucose and glycosylated haemoglobin (HbA1c) treatment effects. (a) Deviation scores for glucose treatment effects under rosiglitazone monotherapy in mice and rats. Horizontal lines of pluses indicate sample means; n denotes number of observations. (b) Deviation scores for HbA1c treatment effects under rosiglitazone monotherapy in mice and rats. Horizontal lines of pluses indicate sample means; n denotes number of observations. (c) Deviation scores for glucose treatment effects under rosiglitazone monotherapy in rat and mouse strains. Horizontal lines of pluses indicate sample means. Strains rarely used in studies are pooled as 'other mouse' and 'other rat'; n denotes number of observations.

measurement occasions, and taken the square root of. Values so derived were referred to as deviation scores.

The effects of factors raising methodological issues, such as species, induction method, diet, sex and administration route were evaluated by comparing groups in terms of deviation scores. Basic unadjusted comparisons were made using simple linear regression. Adjusted effect estimates were obtained using multiple linear regression. Deviation scores were log-transformed to improve normality. Robust standard errors based on the clustering of observations within studies were used to make the estimation consistent with the presence of non-independence among observations coming from the same study. Explanatory variables with negligible effect estimates on both outcomes and no appreciable role as adjustment or interaction factors were eliminated to ensure model parsimony.

All statistical calculations and analyses were done using the software package Stata version 11. The detailed protocol of data collection and analysis is available as Supporting Information Appendix

Results

General description of dataset

As a result of rosiglitazone treatment, hyperglycaemia ameliorated in both diabetic animals and patients with T2DM.

The dose used to efficiently reduce blood glucose and HbA1c levels in animals varied considerably across studies (6–20 mg/kg). Generally speaking, higher doses are used in animals than humans. In T2DM patients, rosiglitazone reduced fasting blood glucose and HbA1c levels at a daily dose of 4–8 mg. Animal studies differed from studies in human patients in terms of the age of study subjects: the initial age of human patients corresponded to late adulthood whereas that of animals represented adolescence and early adulthood. In animal studies the impact of treatment was detected by a comparison of glucose and HbA1c levels between treatment and placebo groups, whereas 30 out of 62 human studies presented data as compared with baseline measurements (see Fig. 1), that could not be used for our analysis. Table 1 shows main characteristics of study subjects and glucose and HbA1c parameters from human and animal studies.

Comparing the consistency of animal models with the human reference for glucose and HbA1c treatment effects

All analyses shown later refer to models using dose as a categorical factor as no appreciable differences from results obtained with the continuous formulation were observed.

Rodent models roughly agreed with human data, especially for rats, but showed considerably varied accuracy in

Table 1 Characteristics of subjects in the studies

	Human	Rat	Mouse	Gerbil	Hamster
List of strains (number of studies, number of animals)	22, 3076	Brown Norway (1, 18) Dahl SS/JrHsd (1, 20) Goto Kakizaki (1, 13) LETO (1, 10) OLETF (1, 10) Ob-ZSF1 (1, 30) Sprague Dawley (12, 293) Wistar (9, 189) ZDF (14, 251) Zucker lean (2, 32)	(apo)E2 knock-in mice (1, 20) A-ZIP (1, 12) BALB/c (1, 40) C57BL6J (3, 70) DBA/2J (1, 63) FVB/N (1, 12) KKAgamma (1, 12) Ldlr-/-Apob100/100Lepob/ob (1, 23) MKR (1, 12) Swiss albino (2, 30) apoE deficient (2, 96) db/db (14, 214) ob/ob (4, 52)	Gerbil (1, 32)	Syrian Golden (1, 30)
Initial age, weeks – weighted arithmetic mean (SD) years	58.50 (8.93)	8.59 (4.51)	7.05 (2.53)	16.00 (0.00)	9.00 (0.00)
Experimental time, weeks	27.09 (15.10)	5.47 (5.44)	8.53 (7.05)	2.00 (0.00)	5.00 (0.00)
Average dose used, mg kg d ⁻¹	4.87 (2.00)	6.49 (6.95)	11.59 (15.48)	20.00 (0.00)	7.15 (0.00)
Glucose level after treatment, mmol L ⁻¹	8.44 (2.21)	10.58 (2.57)	13.28 (8.72)	6.34 (3.70)	3.50 (2.32)
HbA1c after treatment, %	7.69 (1.13)	4.71 (0.73)	7.19 (2.93)	No data	No data

Information provided in the table is based on all study subjects either treated with rosiglitazone or receiving no treatment, except for the rows 'Glucose level after treatment' and 'HbA1c after treatment' which include only treated groups. Means and standard deviations were pooled across studies by weighted averaging based on sample sizes, means and standard deviations reported for each study. SD, standard deviation.

Table 2 Ratios of glucose and HbA1c levels between the rosiglitazone and placebo arms in human studies at different times into follow-up

	Fasting glucose level ratios, rosiglitazone versus placebo		HbA1c level ratios, rosiglitazone versus placebo	
Low-dose group, median follow-up	3 weeks	0.859 (0.478–1.545)	8 weeks	0.940 (0.908–0.974)
Low-dose group, end of follow-up	52 weeks	0.817 (0.691–0.965)	52 weeks	0.894 (0.811–0.985)
High-dose group, median follow-up	3 weeks	1.056 (0.976–1.143)	8 weeks	1.027 (0.998–1.057)
Low-dose group, end of follow-up	52 weeks	0.870 (0.824–0.918)	52 weeks	0.992 (0.953–1.034)

Parentheses include 95% confidence intervals.

reflecting the efficacy of rosiglitazone in humans. For clarity, data on human-scaled values are provided in Table 2. The table shows how glucose and HbA1c levels differed in the rosiglitazone versus the placebo arms in humans at various time points into follow-up. For example, in low dose treatment at 3 weeks, the average glucose level in the rosiglitazone arms was 85.9% of that in the placebo arms, or, in other words, a reduction of about 14% can be estimated as the treatment effect. For simplicity, data at median and terminal follow-up time points were used to illustrate the tendencies.

Analysis of the 69 publications reporting studies with rats and mice (the single hamster and gerbil study excluded from this analysis) showed that the consistency of animal species-based models with the human reference for glucose and HbA1c treatment effects is highly variable. Glucose and HbA1c treatment effects in rats agreed better with the expected values based on human data than in mice, indicating that rat-based models may have greater consistency than those based on other species. Figure 2a shows that rats had significantly lower scores of deviation from the human reference than mice for glucose treatment effects during rosiglitazone treatment (means: 0.275 vs. 0.594, respectively; $P = 0.0023$, unadjusted analysis). In case of HbA1c treatment effects, rats had the lowest deviation scores again (0.385 on average, vs. 0.639 in mice), and the unadjusted difference was borderline significant (Fig. 2b, $P = 0.0446$).

Strains

The question regarding which strain is the most appropriate to model the clinical efficacy of rosiglitazone could not be conclusively answered. In the analysed studies, 10 rat strains, 11 mouse strains and 2 other species were used. Although Sprague Dawley rats and C57BL/6 mice seemed to be the most consistent animal models for the clinical efficacy of rosiglitazone, no statistical difference in deviation scores was observed among groups. The strain presenting the least consistent results for the clinical efficacy of rosiglitazone (in terms of both glucose and HbA1c treatment effects, with mean deviation scores 0.801 and 0.698, respectively) was the commonly used db/db mouse. Data on rat and mouse models are presented in Fig. 2c.

Impact of animal study covariates – induction method, drug administration route, sex of animals and applied diet

From the point of view of similarity to the human response to rosiglitazone, STZ induction was observed to be the most appropriate induction method, as presented in Fig. 3a. STZ-induced diabetic animal models had the highest average consistency (mean deviation score: 0.237) with the human reference as evaluated in terms of the treatment effects of rosiglitazone on glucose levels. Performing a similar analysis for HbA1c was not practicable because of the low number of relevant studies.

Comparing the three administration methods, oral gavage was associated with the lowest deviance scores (mean: 0.369 for glucose, 0.449 for HbA1c). Of note, the number of observations for peritoneal administration was relatively small, and the unadjusted difference was only borderline significant for glucose (Fig. 3b) and non-significant for HbA1c (Fig. 3c).

To test the hypothesis that humanized (high-sucrose, HF) or other diets could have an impact on diabetes onset and progress, and thereby on the consistency between animal and human models, deviation scores were compared across groups defined by high-sucrose, HF, high-sucrose-HF, and low-fat diets. No link was found between diet and the performance of animal models. Similarly, the sex of the animals was not observed to affect the deviation between animal and human models. Raw data associated with factors representing covariates are shown in Table 3.

In multiple regression analysis, species, induction method and drug administration route were found to be substantial factors of animal model performance (Table 4). Adjusted for induction method and administration route, rat models performed better: the difference in deviation scores between rats and other species was strongly significant for the FG outcome, although not quite for HbA1c. STZ as an induction method was found to better approximate the relationship between rosiglitazone exposure and treatment effect, observed in humans, than other methods, especially when treatment effect was assessed through HbA1c levels. The data also suggest that intraperitoneal

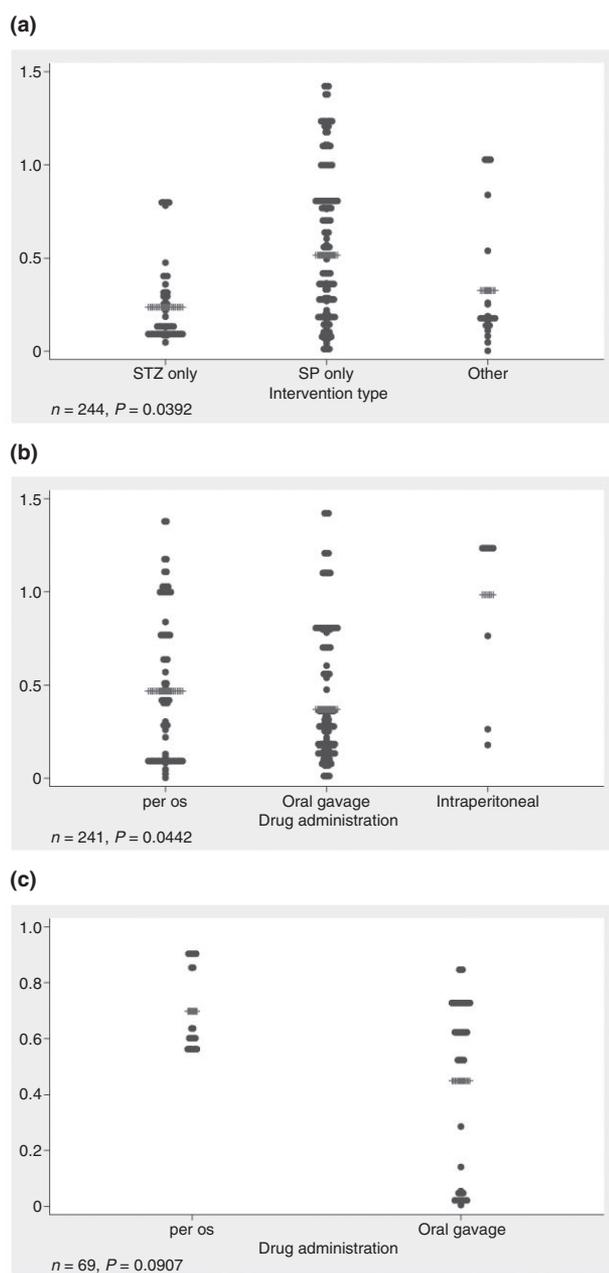


Figure 3 Impact of animal study covariates. (a) Deviation scores for glucose treatment effects under rosiglitazone monotherapy in groups by diabetes induction method. 'STZ only' denotes the use of streptozotocin (STZ) only; 'SP only', the absence of exogenous (non-spontaneous) induction methods; 'other' denotes models that were not used frequently enough for individual analysis (alloxan models, partial pancreatectomy models, high-fat/high-sucrose diet-fed models, high-fat diet-fed STZ models, and intrauterine growth retardation models). Horizontal lines of pluses indicate sample means; n denotes number of observations. (b) Deviation scores for glucose treatment effects under rosiglitazone monotherapy in groups by drug administration route. Horizontal lines of pluses indicate sample means; n denotes number of observations. (c) Deviation scores for HbA1c treatment effects under rosiglitazone monotherapy in groups by drug administration route. Horizontal lines of pluses indicate sample means; n denotes number of observations.

administration may result in poorer consistency between human and animal models than per os administration.

Discussion

Clear understanding on the predictive validity of animal models is in those fields of applied drug testing where several animal models are used. In this study, we presented a method to statistically evaluate the predictive validity of animal model studies. Using rosiglitazone in diabetes as a case study and comparing treatment effects between human and animal studies, we showed that studies in rats are better predictors of results in humans than other animal model studies. Agreement between human and animal studies was further affected by disease induction method and drug administration route. This was the first time that the concept of predictive capacity of animal models was systematically approached in type 2 diabetes pharmaceutical studies.

Our case study of rosiglitazone had two research questions: to compare treatment effect of rosiglitazone in human and animal models in order to identify which is more relevant to humans, and to understand the impact of experimental diet, induction method, sex and administration route of rosiglitazone on the treatment effect of rosiglitazone.

According to the research question on the human relevance of animal models, our findings showed that although the consistency of animal species-based models with the human reference for glucose and HbA1c treatment effects are highly variable, glucose and HbA1c treatment effects in rats agreed better with the expected values based on human data than in other species; rats had significantly lower scores of deviation from the human reference than mice for glucose and HbA1c treatment effects. The question regarding which strain is the most appropriate to model the clinical efficacy of rosiglitazone could only be tentatively answered. There was no statistical difference in deviation scores observed between rat groups; among mouse strains, C57BL/6 showed the most consistent, while db/db showed the least consistent results.

Since models differ in physiological and genetic relevance, there is no single diabetic animal model which would fit for all scientific purposes; ideally, more than one species or strain are used in each study (13). Three different approaches are used to evaluate the reliability of animal models: the first is phenomenological/pathophysiological similarity of the model to the syndrome it is imitating (face validity), the second compares the etiology of diseases in animal models and humans (construct validity), and the third approach refers to the ability of the model to respond to appropriate medications (predictive validity) (23). The vast majority of reviews on T2DM animal models gives information on the models' face and construct validity,

	Rat	Mouse	Gerbil	Hamster
Induction method (number of studies, number of animals)				
SP	27, 350	21, 208	1, 16	1, 15
SP + STZ	1, 8			
SP/diet	1, 10			
GM		6, 58		
GM + STZ		2, 24		
Alloxan	1, 5	5, 101		
Dexamethasone	1, 6	1, 7		
Diet	2, 14			
Low protein IU diet	1, 7			
Streptozotocin	13, 135			
Surgery	1, 20			
Administration route (number of studies, number of animals)				
Per os	14, 139; 1, 10	12, 129		
Oral gavage	24, 394	15, 250	1, 16	1, 15
Intraperitoneal injection	1, 12	2, 13		
Unknown	1, 10	1, 6		
Diets (number of studies, number of animals)				
High-NaCl	1, 10			
High-energy			1, 16	
High-fat	7, 77	6, 103		1, 15
High-fat-high-sucrose	1, 8			
Low-fat		1, 5		
Normal diet	24, 265	21, 275		
Unknown diet	8, 195	2, 15		
Sex in animals in absolute numbers				
Male	401	318	16	15
Female	36	18	0	0
Both or unknown	118	62	0	0

GM, genetically induced models; SP, spontaneous diabetes models; STZ, streptozotocin models; IU, intrauterine.

Factor	Contrast	Effect	95% CI	<i>P</i>
Outcome: fasting glucose				
Species	Rat versus mouse	-0.790	-1.255 -0.325	0.001
Intervention	SP only versus STZ only	0.505	-0.086 1.097	0.093
Intervention	Other versus STZ only	0.151	-0.722 1.023	0.731
Administration route	Oral gavage versus per os	-0.188	-0.666 0.290	0.434
Administration route	Intraperitoneal versus per os	0.650	0.130 1.169	0.015
Outcome: HbA1c				
Species	Rat versus mouse	-1.577	-3.257 0.103	0.063
Intervention	SP only versus STZ only	3.179	1.020 5.339	0.008
Intervention	Other versus STZ only	4.937	2.494 7.379	0.001
Administration route	Oral gavage versus per os	-0.867	-2.131 0.397	0.159

Additive effects of animal study factors on deviation scores of animal models with respect to the human reference for the relationship between rosiglitazone exposure and treatment effect on follow-up levels of fasting glucose and HbA1c. CI, confidence interval.

categorized by species. The characteristics of often used species such as murine models (24), monkeys (25) canines (26) or pigs (27) are widely discussed. As it was noted, very few studies on T2DM addressed the translatability of

animal research results to humans and how to select animal models with 'higher human relevance'. A recent study has pointed out that genetic similarities between humans and certain species can be useful for appropriate model selec-

Table 3 Description of factors that may cause methodological issues (covariates) in animal studies

Table 4 Additive effects of animal study factors on deviation scores of animal models

tion (28) and another study categorized mouse models by outcome measures that are used in the clinical practice of diabetic nephropathy (29).

A generally good correlation between human and animal experimental outcomes is often assumed in pharmacological studies, not considering the impact of the species effect. For example, the use of the leptin-deficient mouse (ob/ob) in type 2 diabetes research is widely recommended in any pharmaceutical research (30), but our case study does not prove 'high predictive validity of this model' for rosiglitazone efficacy in humans. This example points out that if we want to understand how reliable our animal models are in a particular situation, results must be (re)assessed in the light of human data.

Concerning our second research question, in multiple regression analysis, induction method, and drug administration route were studied, and the induction method was found to be a substantial factor affecting animal model performance. STZ as an induction method was associated with better approximation of the human relationship between rosiglitazone exposure and treatment effect than other methods, especially when treatment effect was ascertained through HbA1c levels. Although this result gives significant input for experimental design, it has to be underlined that these specific comparisons were complicated because HbA1c level, which is the primary outcome of human studies, was less frequently reported in animal studies (30).

There are several established methods for determining glycated haemoglobins (31) and many of them are used in rodents (32). The relation between HbA1C and plasma glucose levels in diabetes animal models has also been well described (33). However, in an ideal case, free plasma concentration of rosiglitazone could be compared between human and animal subjects, and identify how much of the drug is in the blood.

One of the strengths of this study is that to ensure sufficient coverage of relevant literature, it goes beyond traditional information sources; data from unpublished human studies have also been involved in the analysis. A very complex and long lawsuit filed against GSK started in 2007. One of the consequences of the legal action against GSK was that all studies performed by the company were made available through the company's website (18) and thus became available for our analysis. On the other hand, there are specific limitations to this work, such as the presence of different characteristics in human versus animal studies, which impeded the immediate comparability of the two datasets. One of these constraints was that the dose used to efficiently reduce blood glucose and HbA1c levels in animals varied considerably between studies (6–20 mg/kg). Generally speaking, higher doses are used in animals than humans. Additionally, in animal studies, the impact of treatment was detected by a comparison of glucose and

HbA1c levels between treatment and placebo groups; consequently, single-arm human studies that presented data as compared with baseline measurements, i.e. without a placebo control, could not be used in the analysis. Another problem was that the age of the study populations differed: the initial ages of human patients correspond to late adulthood whereas that of animals represents adolescence and early adulthood.

Animal models are unique in their predictive value for human drug efficacy. This study aimed to present how the predictive validity of animal models can be assessed retrospectively. Our method shows that regular reassessment of animal models helps to identify 'human relevance of each model' and adapt research design for actual research goals. Although our findings are important, one should be careful with interpretation of results presented here; extrapolation of our results outside the thiazolidinedione class of drugs should be avoided.

Conflict of interest statement

No conflict of interest was declared.

Funding

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 'National Excellence Program'. This research is continuation of a research grant (bolsa de investigação com a referência SFRH/BPD/64149/2009) supported by the Fundação para a Ciência e a Tecnologia. Research period of this project: 2009–2013.

Acknowledgements

We would like to thank the useful comments of Istvan Gyertyan and Laszlo Kardos that greatly contributed to improve the quality of this paper.

Supporting information

Additional Supporting Information may be found in the online version of this article, <http://dx.doi.org/10.1111/obr.12265> 

Appendix S1  Estimating the predictive validity of diabetic animal models in rosiglitazone studies. 

References

- Graham MLBM, Papas KK, Hering BJ, Schuurman H-J. Species incompatibilities in the pig-to-macaque islet xenotransplant model affect transplant outcome: a comparison with allotransplantation. *Xenotransplantation* 2011; 18: 328–342. 100
101
102
103

2. Whiteside GT, Adedoyin A, Leventhal L. Predictive validity of animal pain models? A comparison of the pharmacokinetic-pharmacodynamic relationship for pain drugs in rats and humans. *Neuropharmacology* 2008; **54**: 767–775.
3. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). (2005). Guidance for industry: estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers. [WWW document]. URL <http://www.fda.gov/downloads/Drugs/Guidances/UCM078932.pdf> (accessed 2013).
4. World Health Organization HSI/ESD/ICTRP. (2004). International Clinical Trials Registry Platform (ICTRP). [WWW document]. URL <http://www.who.int/ictcp/en/> (accessed 2013).
5. Walton MK. Addressing and advancing the problem of missing data. *J Biopharm Stat* 2009; **19**: 945–956.
6. Ghersi D, Clarke M, Berlin J *et al.* Reporting the findings of clinical trials: a discussion paper. *Bull World Health Organ* 2008; **86**: 492–493.
7. Vedula SS, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of gabapentin. *PLoS Med* 2013; **10**: e1001378.
8. Janssen H, Speare S, Spratt NJ *et al.* Exploring the efficacy of constraint in animal models of stroke: meta-analysis and systematic review of the current evidence. *Neurorehabil Neural Repair* 2013; **27**: 3–12.
9. van der Worp HB, Howells DW, Sena ES *et al.* Can animal models of disease reliably inform human studies? *PLoS Med* 2010; **7**: e1000245.
10. Percie du Sert N, Rudd JA, Apfel CC, Andrews PL. Cisplatin-induced emesis: systematic review and meta-analysis of the ferret model and the effects of 5-HT(3) receptor antagonists. *Cancer Chemother Pharmacol* 2011; **67**: 667–686.
11. Lashev L, Pashov D, Kanelov I. Species specific pharmacokinetics of rolitetracycline. *Zentralbl Veterinarmed A* 1995; **42**: 201–208.
12. Khan HA. CalcDose: a software for drug dosage conversion using metabolically active mass of animals. *Drug Chem Toxicol* 2003; **26**: 51–58.
13. King AJ. The use of animal models in diabetes research. *Br J Pharmacol* 2012; **166**: 877–894.
14. Rohatgi A (2012). WebPlotDigitizer. [WWW document]. URL <http://www.arohatgi.info/WebPlotDigitizer/app/> (accessed 2013).
15. de Vries RB, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. A search filter for increasing the retrieval of animal studies in Embase. *Lab Anim* 2011; **45**: 268–270.
16. Rahman S, Ismail AA, Ismail SB, Naing NN, Abdul Rahman AR. Effect of rosiglitazone/ramipril on preclinical vasculopathy in newly diagnosed, untreated diabetes and IGT patients: 1-year randomised, double-blind, placebo-controlled study. *Eur J Clin Pharmacol* 2007; **63**: 733–741.
17. Louisa M, Takeuchi M, Takeuchi M, Nafrialdy S, Setiabudy R. A meta-analysis on treatment effects of thiazolidinediones for type 2 diabetes mellitus in Asian populations. *Acta Med Indones* 2011; **43**: 39–52.
18. GSK. (2014). Studies filtered by compound rosiglitazone. [WWW document]. URL <http://www.gsk-clinicalstudyregister.com/compounds/rosiglitazone#rs> (accessed 2013).
19. Islam MS, Loots du T. Experimental rodent models of type 2 diabetes: a review. *Methods Find Exp Clin Pharmacol* 2009; **31**: 249–261.
20. Turner PV, Brabb T, Pekow C, Vasbinder MA. Administration of substances to laboratory animals: routes of administration and factors to consider. *J Am Assoc Lab Anim Sci* 2011; **50**: 600–613.
21. Ricci MR, Ulman EA. Laboratory animal diets: a critical part of your in vivo research. *Anim Lab News* 2005; **4**: 1–4.
22. Choi JW, Aseer KR, Chaudhari HN, Mukherjee R, Choi M, Yun JW. Gender dimorphism in regulation of plasma proteins in streptozotocin-induced diabetic rats. *Proteomics* 2013; **13**: 2482–2494.
23. McKinney WT Jr, Bunney WE Jr. Animal model of depression. I. Review of evidence: implications for research. *Arch Gen Psychiatry* 1969; **21**: 240–248.
24. McMurray F, Cox RD. Mouse models and type 2 diabetes: translational opportunities. *Mamm Genome* 2011; **22**: 390–400.
25. Harwood HJ Jr, Listrani P, Wagner JD. Nonhuman primates and other animal models in diabetes research. *J Diabetes Sci Technol* 2012; **6**: 503–514.
26. Ionut V, Liu H, Mooradian V *et al.* Novel canine models of obese prediabetes and mild type 2 diabetes. *Am J Physiol Endocrinol Metab* 2010; **298**: E38–E48.
27. Lee MS, Song KD, Yang HJ, Solis CD, Kim SH, Lee WK. Development of a type II diabetic mellitus animal model using Micropig(R). *Lab Anim Res* 2012; **28**: 205–208.
28. Yu S, Zheng L, Li Y *et al.* A cross-species analysis method to analyze animal models' similarity to human's disease state. *BMC Syst Biol* 2012; **6**(Suppl. 3): S18.
29. Breyer MD. Drug discovery for diabetic nephropathy: trying the leap from mouse to man. *Semin Nephrol* 2012; **32**: 445–451.
30. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). (2008). Guidance for industry diabetes mellitus: developing drugs and therapeutic biologics for treatment and prevention. [WWW document]. URL <http://www.fda.gov/downloads/Drugs/Guidances/ucm071624.pdf> (accessed 2013).
31. Lenters-Westra E, Schindhelm RK, Bilo AJ, Slingerland RJ. Haemoglobin A1c: Historical overview and current concepts. *Diabetes Res Clin Pract* 2013; **99**: 75–84.
32. De Tata V, Novelli M, Bombara M *et al.* Determination of glycated hemoglobins in the rat: comparison between two different chromatographic methods and application in experimental diabetology. *Res Exp Med (Berl)* 1996; **196**: 9–16.
33. Dan K, Fujita H, Seto Y, Kato R. Relation between stable glycated hemoglobin A1C and plasma glucose levels in diabetes-model mice. *Exp Anim* 1997; **46**: 135–140.

AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remarks
1	AUTHOR: Please check and confirm that the corresponding author's address is correct.	✓
2	AUTHOR: Please confirm that given names (red) and surnames/family names (green) have been identified correctly.	✓
3	AUTHOR: Please provide full developer's information for this software.	done
4	AUTHOR: Stata version 11: Please provide full developer's information for this software.	done
5	AUTHOR: SD has been changed to Sprague Dawley; please check and confirm that this is correct.	✓
6	AUTHOR: Table 3 has been restructured for clarity; please check and confirm that this is correct.	✓
7	AUTHOR: For all URLs, please check the website address/URLs and confirm that they are correct. (Please note that it is the responsibility of the author(s) to ensure that all URLs given in this article are correct and useable).	done
8	AUTHOR: Supporting information has been changed to Appendix S1; please check and confirm that this is ok. Also, please that the legend is correct.	✓
9	AUTHOR: Please supply accessed date for References 3, 4, 14, 18 and 30.	done
10	AUTHOR: Please supply the given name of author Nafrialdi for Reference 17.	done
11	AUTHOR: Reference 21 if this is not a one-page article please supply the first and last pages for this article.	done

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

<i>Instruction to printer</i>	<i>Textual mark</i>	<i>Marginal mark</i>
Leave unchanged	... under matter to remain	Ⓟ
Insert in text the matter indicated in the margin	∧	New matter followed by ∧ or ∧ [Ⓢ]
Delete	/ through single character, rule or underline or ┌───┐ through all characters to be deleted	Ⓞ or Ⓞ [Ⓢ]
Substitute character or substitute part of one or more word(s)	/ through letter or ┌───┐ through characters	new character / or new characters /
Change to italics	— under matter to be changed	↵
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	≡ under matter to be changed	≡
Change to bold type	~ under matter to be changed	~
Change to bold italic	≈ under matter to be changed	≈
Change to lower case	Encircle matter to be changed	≡
Change italic to upright type	(As above)	⊕
Change bold to non-bold type	(As above)	⊖
Insert 'superior' character	/ through character or ∧ where required	Υ or Υ under character e.g. Υ or Υ
Insert 'inferior' character	(As above)	∧ over character e.g. ∧
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	Ƴ or ƴ and/or ƶ or Ʒ
Insert double quotation marks	(As above)	ƶ or Ʒ and/or Ʒ or ƶ
Insert hyphen	(As above)	⊥
Start new paragraph	┌	┌
No new paragraph	┐	┐
Transpose	┌┐	┌┐
Close up	linking ○ characters	⸸
Insert or substitute space between characters or words	/ through character or ∧ where required	⸶
Reduce space between characters or words		⸵