

1 **Title:** Predicting animal to human translation: A proof of concept study using qualitative comparative
2 analysis

3
4 **Short title:** translation QCA

5
6 **Authors:** Cathalijn H.C. Leenaars^{1-2*}, Steven Teerenstra³, Franck L.B. Meijboom² André Bleich¹

7
8 **Affiliations:**

- 9 1. Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany
- 10 2. Department of Animals in Science and Society – Faculty of Veterinary Sciences, Utrecht
11 University, Utrecht, the Netherlands
- 12 3. Department for Health Evidence (section biostatistics), Radboud Institute for Health Sciences,
13 Radboud University Medical Center, Nijmegen, The Netherlands

14
15 **Keywords:** Prediction, extrapolation, validity, modelling, scoping review, qualitative comparative
16 analysis, regression analysis

17
18 **Abstract**

19 Drug development suffers from high attrition rates; promising drug candidates fail in clinical trials.

20 Low animal-to-human translation may impact attrition. We previously summarised published
21 translational success rates, which varied from 0% to 100%. Based on analyses of individual factors,
22 we could not predict translational success.

23 Several approaches exist to analyse effects of combinations of potential predictors on an outcome. In
24 biomedical research, regression analysis (RGA) is common. However, with RGA it is challenging to
25 analyse multiple interactions and specific configurations (\approx combinations) of variables, which could
26 be highly relevant to translation.

27 Qualitative comparative analysis (QCA) is an approach based on set theory and Boolean algebra. It
28 was successfully used to identify specific configurations of factors predicting an outcome in other
29 fields. We reanalysed the data from our preceding review with a QCA. This QCA resulted in the
30 following formula for successful translation:

31 $\sim\text{Old}^*\sim\text{Intervention}^*\sim\text{Large}^*\text{MultSpec}^*\text{Quantitative}$

32 Which means that within the analysed dataset, the combination of relative recency (\sim means not;
33 >1999), analyses at event or study level (not at intervention level), $n < 75$, inclusion of more than one
34 species and quantitative (instead of binary) analyses always resulted in successful translation ($>85\%$).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

35 Other combinations of factors showed less consistent or negative results. An RGA on the same data
36 did not identify any of the included variables as significant contributors.
37 While these data were not collected with the QCA in mind, they illustrate that the approach is viable
38 and relevant for this research field. The QCA seems a highly promising approach to furthering our
39 knowledge on animal-to-human translation and decreasing attrition rates in drug development.
40

41 INTRODUCTION

42 While the debate on the relevance and acceptability of animal experimentation remains polarized [1-
43 4], animal experiments are still hard to avoid in the process of new drugs reaching the market.

44 However, the predictive value of animal experiments has limits, and poor translation from animal
45 experiments to humans may contribute to the high attrition rates in drug development [5].

46 Explaining attrition can contribute to more efficient drug development, which is one of the reasons
47 why we analyse translational success. Another one is animal welfare; we cannot defend using
48 animals for translational experiments that do not provide relevant information.

49 While the most common approach to evaluating translation is mechanistic and qualitative, we
50 started focussing on quantitative studies in a scoping review of reviews [6]. In that review, we
51 observed translational success rates from 0% to 100% (median: 64%; interquartile range (IQR): 44-
52 79%). To identify factors contributing to translational success, we visualised these data by several
53 potentially predictive factors, which are explained further below. Relevant for this paper are:
54 definition type (binary vs. continuous definitions of translation), unit of analysis (were the results
55 analysed at the event, intervention or study level), species, the number of included observations
56 (events, interventions or studies), and the year of publication. There was no apparent relationship
57 between any of these individual factors and the percentage of translational success. However, the
58 effects of combinations of these potential factors on translational success could still be relevant. We
59 thus performed additional analyses on our previously-collected data, which are described in this
60 paper.

61 The following five factors were further analysed based on their theoretical relevance: publication
62 age, unit of analysis, analysis size, inclusion of multiple species and type of definition for translation
63 (binary vs. continuous). Publication age was included as the state of science and the quality of animal
64 models are thought to improve over time, and because animal-to-human translation is getting more
65 attention in the last decade, which may result in improvements.

66 The unit of analysis was previously extracted as a categorical variable with 3 possible values: event,
67 intervention or study. Events were mainly specific adverse events observed in animals, humans or
68 both, where the observed translation (e.g. the percentage of adverse events observed in both
69 animals and humans) depends on study size and chance; larger studies have a larger chance of
70 picking up rare events [7]. Interventions were mostly specific drugs, where certain groups of
71 interventions may translate better than others [8], and the observed translational success rate
72 depends on the sampling, e.g. which group of drugs was analysed. Analyses at the level of individual
73 studies show translation as results that corresponded between animals and humans, which makes
74 the observed translational success rate depend on multiple factors, including population and
75 experimental design of the compared studies [9].

76 The number of included observations was counted at the level of the unit of analysis, and could thus
77 reflect a number of events, interventions or studies. It was included in the current analyses as a
78 proxy for power, as underpowered studies can result in erroneous conclusions which may impact
79 translation [10]. While some authors argue that species differences introduce uncertainties that
80 seriously limit their validity [11], investigating multiple species can at least theoretically improve
81 translation, as successful transfer of a first species barrier may be predictive of crossing a second.

82 The definition type for animal to human translation could be binary, i.e., there was successful
83 translation or there was not, or continuous, which could refer to a percentage success, a correlation
84 coefficient between animal and human data, a percentage overlap in confidence intervals, etc. Binary
85 definitions can of course be expressed as percentages success, but the type of definition may impact
86 the observed translational success.

87 As multiple roads lead to Rome, multiple combinations of these factors may lead to translational
88 success. In scientific terms, there possibly is causal complexity; comprising equifinality (i.e., there are
89 multiple routes to success) [12] and conjunctural causation (i.e., combinations of factors may be
90 involved instead of individual factors [13]). Besides, causation may be asymmetrical [14]; while the
91 presence of a factor may contribute to success, its absence does not necessarily result in failure (and

92 vice versa). Thus, we have a configural research question: “Which factors, individually or in
93 combination, are necessary or sufficient for successful animal-to-human translation?”
94 To analyse effects of multiple potential predictors on an outcome, regression analysis (RGA) is
95 common. However, RGA is not specifically suitable for research questions comprising multiple
96 interactions and specific configurations (\approx combinations) of predictors. Qualitative comparative
97 analysis (QCA) is an approach developed for configural research questions [15]. It is based on set
98 theory and Boolean algebra. QCA is increasingly used to identify specific configurations of factors
99 predicting an outcome in other fields [16, 17]. We reanalysed the data from our preceding review
100 with a crisp-set QCA (csQCA) [18]. To test the added value of this QCA-approach, we compared it
101 with a classical regression analysis (RGA).

102 **METHODS**

103 **Data collection and selection**

104 We reanalysed the data published in our systematic scoping review [6] for this study. This preceding
105 scoping review was an umbrella review of reviews that addressed animal-to-human translation
106 quantitatively, comparing the results of studies including at least 2 species with one being human.
107 Data were extracted from the included publications to Microsoft Excel. When an included paper
108 described multiple studies or analyses on different data, all those compliant with the inclusion
109 criteria were included as a separate “case” into our analyses. When the original authors did not
110 express translation as a percentage, but provided the data needed to do so, we calculated the
111 percentage and added it to the respective case.
112 From these already published data, we selected the following factors as theoretically relevant for
113 further combined analyses (as explained in the introduction): definition type (binary vs. continuous),
114 unit of analysis (event, intervention or study), species, the number of included observations and the
115 year of publication. Cases with missing data for any of the analysed factors were excluded from the
116 analyses (numbers are mentioned in the results).

117 All analyses were performed in R [19], version 4.0.3 ("Bunny-Wunnies Freak Out"), via RStudio
118 (Version 1.3.1093). Data were imported from excel using the Readxl package [20]. Where needed,
119 data were selected with functions from the Dplyr package [21].

120 Qualitative Comparative Analysis

121 We selected csQCA as it allows for more straightforward definitions and a conservative approach to
122 calibration [18]. (In the field of data analysis, "conservative" means cautious to prevent false positive
123 conclusions.) We performed two separate QCAs; the first for translational success, and the second
124 for translational failure. All cases with calibrated translational success or failure (see below) were
125 included in both QCAs.

126 Data Calibration & data matrix

127 We calibrated all data to create so-called crisp sets as described in table 1. Data were calibrated on
128 theoretical grounds and based on expert opinions from within our network. For example, cut-offs for
129 old publication age at around the start of the century were based on the use of the internet
130 becoming increasingly common in research. Also, >75 observations is considered a large study in the
131 qualitative field. Data distributions were known at the time of calibration, and informed set
132 definitions to the extent that (near) empty sets were consciously prevented.

Set name	Definition IN set	Definition OUT set
Old	Publication date <2000	Publication date ≥ 2000
Intervention	"Interventions" were the unit of analysis	"events" or "studies" were the unit of analysis
Large	k>75	K≤75
MultSpec	At least two animal species were analysed	At most one animal species was analysed
Quantitative	Translation was calculated in a continuous manner	Translation was defined in a binary manner
Translation (outcome)	Success: >80%	Failure: <45%

133 Table 1: Data calibration for QCA

134 Units of analyses in the included reviews could be interventions, publications & studies, and
135 particular (e.g., adverse) events. We distinguished observations at the intervention level from those
136 at the event and study level, as the latter two are both chance processes, while at the intervention
137 level we can imagine a clear distinction between e.g., compounds that do translate well and those
138 that do not.

139 Translational success is difficult to define quantitatively [6]. For our QCAs, we selected the less
140 disputable percentages only: success was defined as >80% correspondence, failure as <45%. We
141 excluded the reviews with percentages from 45 to 80% (which were included in the RGA described
142 below).
143 Set memberships scores were added to the data file in separate columns.

144 Truth table creation & logical minimisation

145 A truth table was created with the truthTable function from the QCA package [22]. We analysed the
146 truth table for sufficiency and necessity of individual factors before addressing combinations of
147 factors. Sufficiency is the presence of the outcome in all cases with the occurrence of a predicting
148 factor, and the factor is never present without the outcome ($F \rightarrow O$, If F then O); necessity is the
149 presence of a factor in all cases with the occurrence of an outcome, but the factor can also be
150 present in cases without the outcome ($F \leftarrow O$, If O then F) Next, logical minimisation of
151 configurations was performed with the minimize function from the QCA package.
152 We anticipated both logical inconsistencies and logical remainders in the truth table. Logical
153 inconsistencies are rows with inconsistent outcomes (i.e., configurations that had both translational
154 success and translational failure). This reanalysis of available data does probably not include all
155 predictive factors relevant to translational success, which would result in perfectly defined sets
156 without logical inconsistencies. Logical remainders are theoretically possible configurations that are
157 not present in the data. Configurations with inconsistent translation and logical remainders were
158 accepted, but they were not used to inform logical minimisation (in the QCA package's truthTable
159 function: incl.cut = 1, n.cut = 1, pri.cut = 0). Also, because of our awareness of missing relevant
160 factors in this proof of concept study, we did not analyse coverage of the solutions; i.e., which part of
161 the cases could be explained with the final formula.

162 Regression analysis

163 All cases with complete data were included for the RGA. The following variables were included in the
164 RGA: definition type (binary vs. continuous), unit of analysis (event, intervention or study), multiple
165 species, the number of included observations and the year of publication. Compared to the QCA, we
166 included more data into the RGA. Because we did not have to dichotomise data into crisp sets, we
167 included all cases with full data, also those with translational success from 45% to 80%, with the
168 original percentage as the outcome. The variables for the number of included observations and the
169 year of publication were also included as numbers instead of dichotomising them. The variables
170 definition type, unit of analysis and multiple species were included as binary variables, exactly like in
171 the QCA.

172 Regression analysis was performed with the `lm` function from R's basic stats package [19]. We tested
173 a single model, including all variables included in the QCA individually. To provide a comparison with
174 the QCA outcome we further added an interaction term for the variables "MultSpec" and
175 "Quantitative".

176 RESULTS

177 QCA

178 In our original review, we included 232 cases from 121 references. From these original cases, the 104
179 without missing data but with clearly successful or clearly unsuccessful translation were included in
180 the QCA. Of these, 50 showed successful translation and 54 did not. The number of cases included in
181 each set as defined in table 1 is shown in table 2.

Set name	N IN set (%)
Old	30 (29%)
Intervention	93 (89%)
Large	15 (14%)
MultSpec	26 (25%)
Quantitative	73 (70%)

182 *Table 2: Number of cases per set*

183 The different observed set configurations with the outcomes are summarised in a truth table (table
184 3). The truth table shows that 9 configurations had inconsistent (both successful and unsuccessful

185 translation) results. For 16 configurations, no cases were observed; these are the so-called logical
186 remainders. None of the configurations was deemed implausible.

Configuration	Old	Intervention	Large	MultiSpec	Quantitative	N cases (n successful translation)	Cases
1	0	0	0	0	0	0 (0)	-
2	0	0	0	0	1	5 (4)	Sultan_2017 (5x)
3	0	0	0	1	0	0 (0)	-
4	0	0	0	1	1	2 (2)	Yen_2014 (2x)
5	0	0	1	0	0	0 (0)	-
6	0	0	1	0	1	0 (0)	-
7	0	0	1	1	0	2 (0)	Olson_2000 (2x)
8	0	0	1	1	1	0 (0)	-
9	0	1	0	0	0	10 (3)	Dong_2011, DeBuck_2007, Evans_2006, Sanoh_2014, Tang_2005, Wang_2010, Ward_2005, Ward_2008 (3x)
10	0	1	0	0	1	33 (17)	Akabane_2010, Cao_2006, Cheng_2008 (2x), Chiou_2000a, Chiou_2000b (2x), Chiou_2002 (2x), Jones_2012 (2x), Grime_2013, Jones_2016 (4x), Kalvass_2007 (2x), Lennernas_2007, Ling_2009, Musther_2014, Paine_2011, Rocchetti_2007, Sanoh_2012, Walton_2004, Ward_2009 (2x), Whiteside_2008, Whiteside_2010, Wong_2004 (4x)
11	0	1	0	1	0	4 (2)	Corpet_2005, Fagerholm_2007a, Fourches_2010, Goteti_2010
12	0	1	0	1	1	5 (4)	Mahmood_2001, Mahmood_2004, Mahmood_2013 (2x) Wajima_2002
13	0	1	1	0	0	7 (2)	Monticello_2017 (3x), Nagilla_2004, Weaver_2003 (3x)
14	0	1	1	0	1	2 (0)	Musther_2014 (2x)
15	0	1	1	1	0	4 (2)	Ennever_2003, Fourches_2010, Monticello_2017 (2x)
16	0	1	1	1	1	0 (0)	-
17	1	0	0	0	0	1 (0)	Litchfield_1961
18	1	0	0	0	1	0 (0)	-
19	1	0	0	1	0	1 (0)	Steinberg_1987
20	1	0	0	1	1	0 (0)	-
21	1	0	1	0	0	0 (0)	-
22	1	0	1	0	1	0 (0)	-
23	1	0	1	1	0	0 (0)	-
24	1	0	1	1	1	0 (0)	-
25	1	1	0	0	0	1 (0)	Schein_1973b
26	1	1	0	0	1	19 (10)	Boxenbaum_1982 (4x), Chiou_1998 (2x), Crouch_1979 (2x), Fagerholm_1996, Freireich_1966 (4x), He_1998 (2x), Schneider_1999 (2x), Sietsema_1989 (2x)
27	1	1	0	1	0	1 (0)	Schein_1973a
28	1	1	0	1	1	7 (4)	Bachmann_1989, Mahmood_1996b, Mahmood_1996c, Mahmood_1998a, Mahmood_1998b (2x), Sietsema_1989
29	1	1	1	0	0	0 (0)	-
30	1	1	1	0	1	0 (0)	-
31	1	1	1	1	0	0 (0)	-
32	1	1	1	1	1	0 (0)	-

187 Table 3: truth table. Green: configuration consistent with translational success. Yellow: Configuration with inconsistent
188 results. E.g., in configuration 2, the second line, the first one in yellow, there are 5 cases with this configuration of potential
189 predictors, of which 4 show translational success, and 1 translational failure. Reddish: configuration consistent with
190 translational failure. White: logical remainders.

191 The configurations with inconsistent results indicates that none of the analysed factors was
192 individually sufficient; successful (or unsuccessful) translation was not consistently present with
193 occurrence of any of the analysed predictive factors. We further checked configurations for
194 individual necessary conditions. None of the included factors was always present (or absent) when
195 successful translation occurred, indicating that none of the included factors was individually
196 necessary for translational success. Similarly, none of the included factors was always present (or
197 absent) when translational failure occurred, so none of the individual factors was necessary for
198 translational failure either.

199 Following our conservative approach excluding inconsistent configurations, the only consistent
200 configuration corresponds with the solution from the logical minimisation process, i.e., the following
201 formula for translational success:

202 $\sim\text{Old}^*\sim\text{Intervention}^*\sim\text{Large}^*\text{MultSpec}^*\text{Quantitative} \rightarrow \text{Success}$

203 This means that the combination of relative recency (~ means not), analyses at event or study level,
204 $n < 75$, inclusion of more than one species and quantitative analyses resulted in successful translation
205 (>85%).

206 Further evaluation of the two cases consistent with this formula shows that they were both derived
207 from the same publication; they are two meta-analyses (for different outcomes) including both
208 animal and human data (further described in the discussion). The results from both meta-analyses
209 showed a high degree of overlap between the animal and the human data.

210 A separate QCA on the reverse outcome results in the following formula for translational failure:

211 $\text{Old}^*\sim\text{Large}^*\sim\text{Quantitative} +$
212 $\sim\text{Old}^*\sim\text{Intervention}^*\text{Large}^*\text{MultSpec}^*\sim\text{Quantitative} +$
213 $\sim\text{Old}^*\text{Intervention}^*\text{Large}^*\sim\text{MultSpec}^*\text{Quantitative} \rightarrow \text{Fail}$

214 This formula arose from logical minimisation of the 6 configurations with a consistent negative
215 outcome in table 3. It shows that there are 3 combinations of factors that consistently combine with
216 translational failure; first, old, small studies using binary definitions of translation, second, newer

217 large studies at the event or study level analysing multiple species using binary definitions of
218 translation, and third, newer large studies at the intervention level analysing single species using
219 quantitative definitions of translation.

220 RGA

221 From the 232 cases from our original review, the 197 cases without missing data for the analysed
222 variables were included in the RGA. The RGA included five explanatory variables, corresponding to
223 the sets included in the QCA, and an interaction term. Two of these variables were numerical and
224 three categorical-binary. The observed values of the variables are summarised in table 4.

Variable	Corresponding QCA set	Type	N(0)	N(1)	Median (range)
Year of Publication	Old	Continuous			2005 (1961-2018)
Intervention	Intervention	Binary	14	183	
Study size	Large	Continuous			21 (4-951)
MultSpec	MultSpec	Binary	150	47	
Quantitative	Quantitative	Binary	87	110	
Translational success	OUT	Continuous			64 (0-100%)

225 *Table 4: Summary of data included in the RGA*

226 A summary of the RGA is provided in table 5. The effect estimates of the individual variables are
227 relatively modest, but note that the effect sizes for publication age and study size are per year/
228 observation. The interaction term had a relatively large effect estimate, consistent with our QCA,
229 indicating that analysing multiple species in combination with analysing translational success
230 quantitatively may be optimal. However, none of the individual variables, nor this interaction,
231 statistically affected translational success in the RGA.

Variable	Estimate	SE	p
(Intercept)	-77.0	301.6	0.80
Year of Publication	0.07	0.15	0.65
Intervention	-1.3	7.3	0.86
Study size	-0.02	0.02	0.65
MultSpec	1.1	6.6	0.87
Quantitative	2.5	4.3	0.56
MultSpec* Quantitative	10.3	8.9	0.25

232 *Table 5: Summary of the RGA*

233

234 **DISCUSSION**

235 We initiated these analyses as a first exploration and proof of principle of the QCA-method in meta-
236 research of animal-to-human translation. QCAs have successfully been performed on data from
237 systematic literature reviews in other fields [16, 17, 23, 24]. However, to the best of our knowledge,
238 we are the first to perform a QCA with animal metadata, and to use it to analyse animal-to-human
239 translation.

240 Our QCA resulted in a preliminary success formula for translational success at the meta-level; recent
241 small reviews with analyses at the event or study level including more than one species and using a
242 quantitative definition of translation were consistent with successful translation. While the effect
243 sizes and directions of the RGA were consistent with these results, hence supportive of the QCA, the
244 RGA did not identify any of the variables, nor the interaction term, as statistically significant. This
245 shows the strength of the QCA approach.

246 **Cases consistent with the QCA-derived formulae**

247 The formula for translational success was based on 2 meta-analyses, which both came from the same
248 paper [25]. The authors performed an in-depth systematic review on guided tissue regeneration for
249 periodontal infrabony lesions. They included 13 human and 9 animal papers, with varying study
250 quality scores. The approach in their paper can be considered exemplary in synthesising animal and
251 human data; combining them into a sub-grouped meta-analysis of percentages of bone filling,
252 allowing for cross-species comparisons.

253 The formula for translational failure was based on 4 cases from 2 papers including newer studies [26,
254 27], combined with 4 papers combining into a single term for older studies [28-31]. To start with the
255 newer studies; Olson et al. [27] described large analyses of animal studies in dogs, primates, rats,
256 mice and guinea pigs. The authors were fairly optimistic in describing that 71% of human adverse
257 events was somehow predicted in an animal model, but they also detailed low concordance rates in
258 toxicity. Musther et al. [26] described correlational analyses of oral bioavailability, and concluded
259 that bioavailability in animals is not predictive of that in humans. They provided separate data for

260 mice (30 compounds), rats (122 compounds) and dogs (125 compounds), which were individually
261 included in our analyses. Their monkey data (41 compounds) were included in the RGA, but excluded
262 from the QCA because of an intermediate translational success rate.

263 To continue with the older studies; Litchfield [28] concluded that many serious side effects that can
264 occur when a drug is given to humans were not predictable from observations on dogs or rats. The
265 rat data were included in the QCA as clear translational failure, the dog data were only included in
266 the RGA because of intermediate translational success. Steinberg & Schlesselman [31] compared the
267 effects of pancreatitis therapeutics between 13 human studies and 25 animal studies in dogs, pigs,
268 rats and guinea pigs, with low correspondence between the results.

269 Schein co-authored two publications in 1973 that both described multiple analyses included in our
270 RGA, most with translational success rates between 45 and 80%. In one publication Schein and
271 Anderson carefully concluded that combining data from multiple species could reduce false negatives
272 for prediction of human adverse events, but one of their data sets reflected translation below 45%
273 and was included in our QCA [29]. In the other publication, Schein et al. concluded that animal
274 models can predict a substantial part of the adverse events occurring in clinical use [30], but again,
275 translation was low in one of their data sets which we included in the QCA.

276 The term in the formula for translational failure that combines the 4 configurations listing these older
277 studies (Old*~Large*~Quantitative) effectively illustrates the concept of logical minimisation,
278 and thereby the potential of the QCA-method.

279

280 The formula we here present for translational success is restricted to smaller studies, and two of the
281 three terms in our formula for translational failure cover large studies. While translational success
282 being related to smaller studies may seem counterintuitive, we would like to mention that the
283 familiarity with the data and the individual cases may be better with smaller studies, which might
284 benefit the quality of the work, which, in specific configurations, could positively affect translational
285 success.

286 Suggestions for future QCAs

287 The here-presented data were not collected with QCA in mind, and therefore not optimised for this
288 approach. However, they still resulted in preliminary formulae consistent with translational success
289 and failure, based on relatively few consistent configurations. We expected contradictory lines in the
290 truth table; configurations that were not consistent with the outcome. We hypothesise that this is
291 mainly due to not all relevant factors being included in this QCA.

292 Future studies should gather data for more factors, but also on a larger number of cases to fill the
293 logical remainders. A 2-step approach is considered; a first large QCA could comprise multiple factors
294 relating to the meta-level. A second QCA could be restricted to the successful configurations from
295 the first, and address factors at the primary study level. With more cases included, and less concern
296 about logical remainders, multivalued QCAs (mv-QCAs) [32] may well be preferable. Concerning the
297 currently included factors, it would be relevant to distinguish studies at the event and the study level
298 instead of grouping them outside the set of studies at the intervention level. While it may seem like
299 an attractive idea to include a factor for individual species, the resulting truth table would become
300 incredibly large and have many logical remainders for the less frequently used species. However, a
301 category distinguishing e.g., rodents, non-human primates and other mammals could be viable for
302 future work.

303 QCAs can also be applied to other types of data than literature [14, 15, 18], which may make other
304 types of data and variables accessible for analyses. E.g., individual compounds or targets can be
305 defined as cases in a QCA, or communication and consideration of all available data in experimental
306 design can be added as factors. There are indications that animal data are insufficiently considered in
307 the design of human trials [33, 34], which might partially explain translational failure. Data could be
308 gathered from multiple sources comprising also investigators' brochures, ethics applications and
309 patent registrations.

310 Implications and conclusion

311 While our results are not conclusive and need confirmation, analysing multiple species in
312 combination with analysing translational success quantitatively may be the optimal approach for
313 future studies aiming to translate to humans. Analysing animal-to-human translation quantitatively
314 as a percentage of correspondence instead of making simplified binary yes/no distinctions fairly
315 reflects the available data. While we do not encourage to increase the number of animal studies
316 overall, if a study aiming at translation is considered to be necessary, we may need to get used to the
317 idea of testing more than one species.

318 In this paper, we present the first QCAs addressing translational success and failure rates. While the
319 data were not collected with this method in mind, we show that the approach is viable, relevant and
320 promising. Further knowledge on animal-to-human translation may help to improve efficiency in
321 research and drug development, and to focus animal studies to where they are predictive.

322 References

- 323 1. Baker EJ, Beck NA, Berg EL, Clayton-Jeter HD, Chandrasekera PC, Curley JL, et al. Advancing
324 nonclinical innovation and safety in pharmaceutical testing. *Drug Discov Today*. 2019;24(2):624-8.
325 Epub 2018/11/24. doi: 10.1016/j.drudis.2018.11.011. PubMed PMID: 30468877.
- 326 2. Genzel L, Adan R, Berns A, van den Beucken J, Blokland A, Boddeke E, et al. How the COVID-
327 19 pandemic highlights the necessity of animal research. *Curr Biol*. 2020;30(18):R1014-R8. Epub
328 2020/09/23. doi: 10.1016/j.cub.2020.08.030. PubMed PMID: 32961149; PubMed Central PMCID:
329 PMCPMC7416712.
- 330 3. Herrmann K, Kimberly J. *Animal Experimentation: Working Towards a Paradigm Change*: Brill;
331 2019.
- 332 4. Hobson-West P. The role of 'public opinion' in the UK animal research debate. *J Med Ethics*.
333 2010;36(1):46-9. Epub 2009/12/23. doi: 10.1136/jme.2009.030817. PubMed PMID: 20026693;
334 PubMed Central PMCID: PMCPMC5027911.
- 335 5. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*.
336 2004;3(8):711-5. Epub 2004/08/03. doi: 10.1038/nrd1470. PubMed PMID: 15286737.
- 337 6. Leenaars CHC, Kouwenaar C, Stafleu FR, Bleich A, Ritskes-Hoitinga M, De Vries RBM, et al.
338 Animal to human translation: a systematic scoping review of reported concordance rates. *J Transl
339 Med*. 2019;17(1):223. Epub 2019/07/17. doi: 10.1186/s12967-019-1976-2. PubMed PMID: 31307492;
340 PubMed Central PMCID: PMCPMC6631915.
- 341 7. Jacobson RM, Adegbenro A, Pankratz VS, Poland GA. Adverse events and vaccination-the lack
342 of power and predictability of infrequent events in pre-licensure study. *Vaccine*. 2001;19(17-
343 19):2428-33. Epub 2001/03/21. doi: 10.1016/s0264-410x(00)00467-9. PubMed PMID: 11257373.
- 344 8. Mahmood I. Prediction of human drug clearance from animal data: application of the rule of
345 exponents and 'fu Corrected Intercept Method' (FCIM). *J Pharm Sci*. 2006;95(8):1810-21. Epub
346 2006/06/24. doi: 10.1002/jps.20650. PubMed PMID: 16795002.
- 347 9. Leenaars C, Stafleu F, de Jong D, van Berlo M, Geurts T, Coenen-de Roo T, et al. A Systematic
348 Review Comparing Experimental Design of Animal and Human Methotrexate Efficacy Studies for
349 Rheumatoid Arthritis: Lessons for the Translational Value of Animal Studies. *Animals (Basel)*.
350 2020;10(6). Epub 2020/06/21. doi: 10.3390/ani10061047. PubMed PMID: 32560528; PubMed
351 Central PMCID: PMCPMC7341304.
- 352 10. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
353 Epub 2005/08/03. doi: 10.1371/journal.pmed.0020124. PubMed PMID: 16060722; PubMed Central
354 PMCID: PMCPMC1182327.
- 355 11. Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in
356 preclinical animal research? Why most animal models are bound to fail. *J Transl Med*.
357 2018;16(1):304. Epub 2018/11/09. doi: 10.1186/s12967-018-1678-1. PubMed PMID: 30404629;
358 PubMed Central PMCID: PMCPMC6223056.
- 359 12. Kast FE, Rosenzweig JE. General systems theory: Applications for organization and
360 management. *Academy of Management Journal*. 1972;15(4):447-65.
- 361 13. Aus JP. Conjunctural causation in comparative case-oriented research. *Qual Quant*.
362 2009;43:173-83. doi: 10.1007/s11135-007-9104-4.
- 363 14. Befani B. Between complexity and generalization: Addressing evaluation challenges with
364 QCA. *Evaluation*. 2013;19(3):269-83. doi: <https://doi.org/10.1177/1474022213493839>.
- 365 15. Ragin CC. Using qualitative comparative analysis to study causal complexity. *Health Serv Res*.
366 1999;34(5 Pt 2):1225-39. Epub 1999/12/11. PubMed PMID: 10591281; PubMed Central PMCID:
367 PMCPMC1089061.
- 368 16. Hanckel B, Petticrew M, Thomas J, Green J. The use of Qualitative Comparative Analysis
369 (QCA) to address causality in complex systems: a systematic review of research on public health
370 interventions. *BMC Public Health*. 2021;21(1):877. Epub 2021/05/09. doi: 10.1186/s12889-021-
371 10926-2. PubMed PMID: 33962595; PubMed Central PMCID: PMCPMC8103124.
- 372 17. Roig-Tierno N, Gonzalez-Cruz TF, Llopis-Martinez J. An overview of qualitative comparative
373 analysis: A bibliometric analysis. *Journal of Innovation & Knowledge*. 2017:15-23.

- 374 18. Vink MP, Van Vliet O. Not Quite Crisp, Not Yet Fuzzy? Assessing the Potentials and Pitfalls of
375 Multi-value QCA. *Field Methods*. 2009;21(3):265-89. doi: 10.1177/1525822X09332633.
- 376 19. R_Core_Team. R: A language and environment for statistical computing. R Foundation for
377 Statistical Computing, Vienna, Austria.2020. Available from: <https://www.R-project.org/>. Accessed
378 [24-01-2022].
- 379 20. Wickham H, Bryan J. readxl: Read Excel Files. R package version 1.3.1. 2019.
- 380 21. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package
381 version 1.0.3. 2021.
- 382 22. Dusa A. QCA with R. A Comprehensive Resource. : Springer International Publishing; 2019.
- 383 23. Melendez-Torres GJ, Sutcliffe K, Burchett HED, Rees R, Richardson M, Thomas J. Weight
384 management programmes: Re-analysis of a systematic review to identify pathways to effectiveness.
385 *Health Expect*. 2018;21(3):574-84. Epub 2018/03/07. doi: 10.1111/hex.12667. PubMed PMID:
386 29508524; PubMed Central PMCID: PMC5980502.
- 387 24. Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in
388 systematic reviews of complex interventions: a worked example. *Syst Rev*. 2014;3:67. Epub
389 2014/06/22. doi: 10.1186/2046-4053-3-67. PubMed PMID: 24950727; PubMed Central PMCID:
390 PMC4079172.
- 391 25. Yen CC, Tu YK, Chen TH, Lu HK. Comparison of treatment effects of guided tissue
392 regeneration on infrabony lesions between animal and human studies: a systematic review and
393 meta-analysis. *J Periodontol Res*. 2014;49(4):415-24. Epub 2013/10/12. doi: 10.1111/jre.12130.
394 PubMed PMID: 24111550.
- 395 26. Musther H, Olivares-Morales A, Hatley OJ, Liu B, Rostami Hodjegan A. Animal versus human
396 oral drug bioavailability: do they correlate? *Eur J Pharm Sci*. 2014;57:280-91. Epub 2013/08/31. doi:
397 10.1016/j.ejps.2013.08.018. PubMed PMID: 23988844; PubMed Central PMCID: PMC4107270.
- 398 27. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, et al. Concordance of the
399 toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol*. 2000;32(1):56-67.
400 Epub 2000/10/13. doi: 10.1006/rtph.2000.1399. PubMed PMID: 11029269.
- 401 28. Litchfield JT, Jr. Forecasting drug effects in man from studies in laboratory animals. *JAMA*.
402 1961;177:34-8. Epub 1961/07/08. PubMed PMID: 13762703.
- 403 29. Schein P, Anderson T. The efficacy of animal studies in predicting clinical toxicity of cancer
404 chemotherapeutic drugs. *Int J Clin Pharmacol*. 1973;8(3):228-38. Epub 1973/11/01. PubMed PMID:
405 4203921.
- 406 30. Schein P, Davis RD, Cooney DA, editors. Qualitative aspects of drug toxicity in prediction from
407 laboratory animals to man. 5th International Congress on Pharmacology 1973.
- 408 31. Steinberg WM, Schlesselman SE. Treatment of acute pancreatitis. Comparison of animal and
409 human studies. *Gastroenterology*. 1987;93(6):1420-7. Epub 1987/12/01. PubMed PMID: 2445620.
- 410 32. Thiem A. Clearly Crisp, and Not Fuzzy: A Reassessment of the (Putative) Pitfalls of Multi-value
411 QCA. *Field Methods*. 2013;25(2):197-207. doi: 10.1177/1525822X13478135.
- 412 33. Sievers S, Wieschowski S, Strech D. Investigator brochures for phase I/II trials lack
413 information on the robustness of preclinical safety studies. *Br J Clin Pharmacol*. 2021;87(7):2723-31.
414 Epub 2020/10/18. doi: 10.1111/bcp.14615. PubMed PMID: 33068032.
- 415 34. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy
416 studies in investigator brochures: Do they enable risk-benefit assessment? *PLoS Biol*.
417 2018;16(4):e2004879. Epub 2018/04/06. doi: 10.1371/journal.pbio.2004879. PubMed PMID:
418 29621228; PubMed Central PMCID: PMC5886385.

419