Fala Cramond
f.cramond@imperial.ac.uk

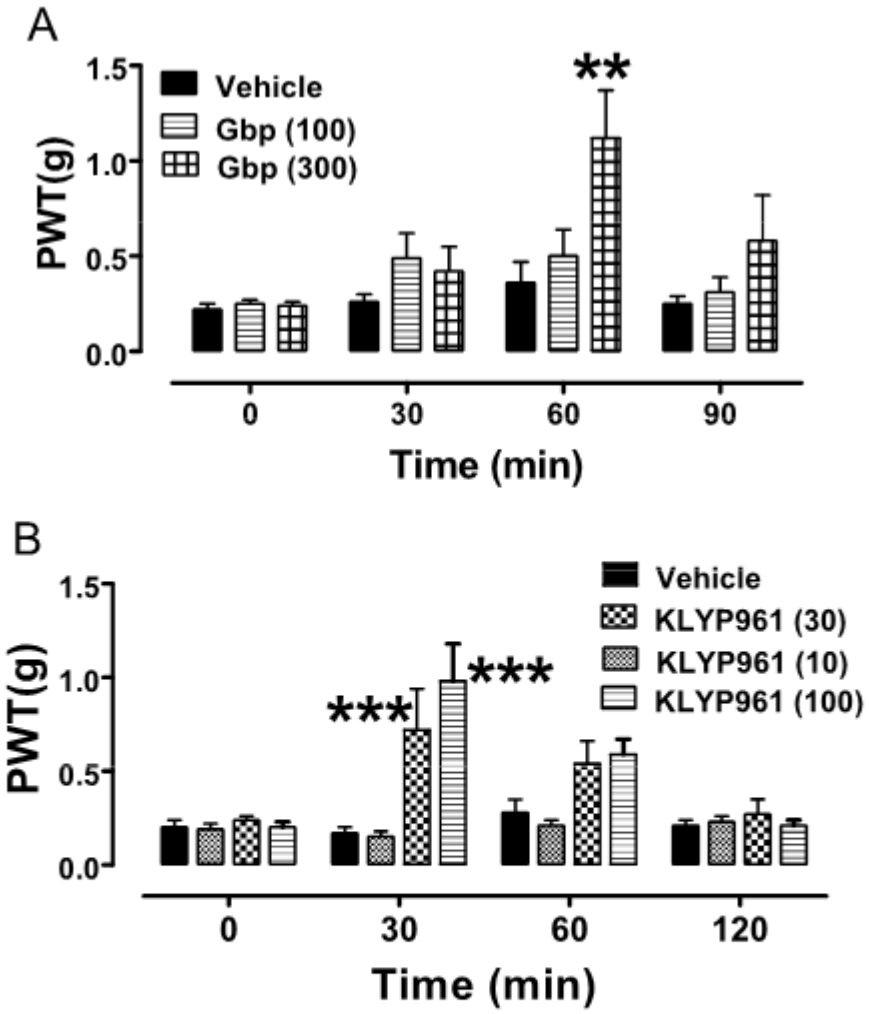


Fig. 5. Orally administered gabapentin (Gbp; A) and KLYP961 (B) attenuate tactile allodynia response after spinal nerve ligation (Chung model) (Kim and Chung, 1992). Tactile allodynia was assessed by monitoring paw withdrawal (PWT) in response to von Frey filaments. Data represent mean ± S.E. (n = 5–6). **, $p < 0.01$ and ***, $p < 0.001$, all relative to vehicle (two-way ANOVA followed by appropriate post hoc test).

Contents

Fala Cramond
f.cramond@imperial.ac.uk
For questions/clarifications or if more data is needed please contact Fala Cramond, f.cramond@imperial.ac.uk.

Introduction

Below is a description of how and what, when extracting data for the neuropathic pain systematic review, I extract for an outcome measure. Firstly, there is a brief overview and explanation of the current extraction process using the CAMARADES (http://www.dcn.ed.ac.uk/camarades/) database, which is the database currently used for neuropathic pain extraction. Next there is a summary of the different data structures which is a brief description of how the relevant data is presented and what the numbers actually are, as explained below it is always mean and a measure of variance. Next, there is a section on the different types of graphs that are presented. I have split them into 5 main categories, based on how they look, rather than the data they contain because I think that is more relevant for the UI development. Fourthly, there are some examples of tables that have relevant data in and lastly, some examples of how data can be presented in the text.

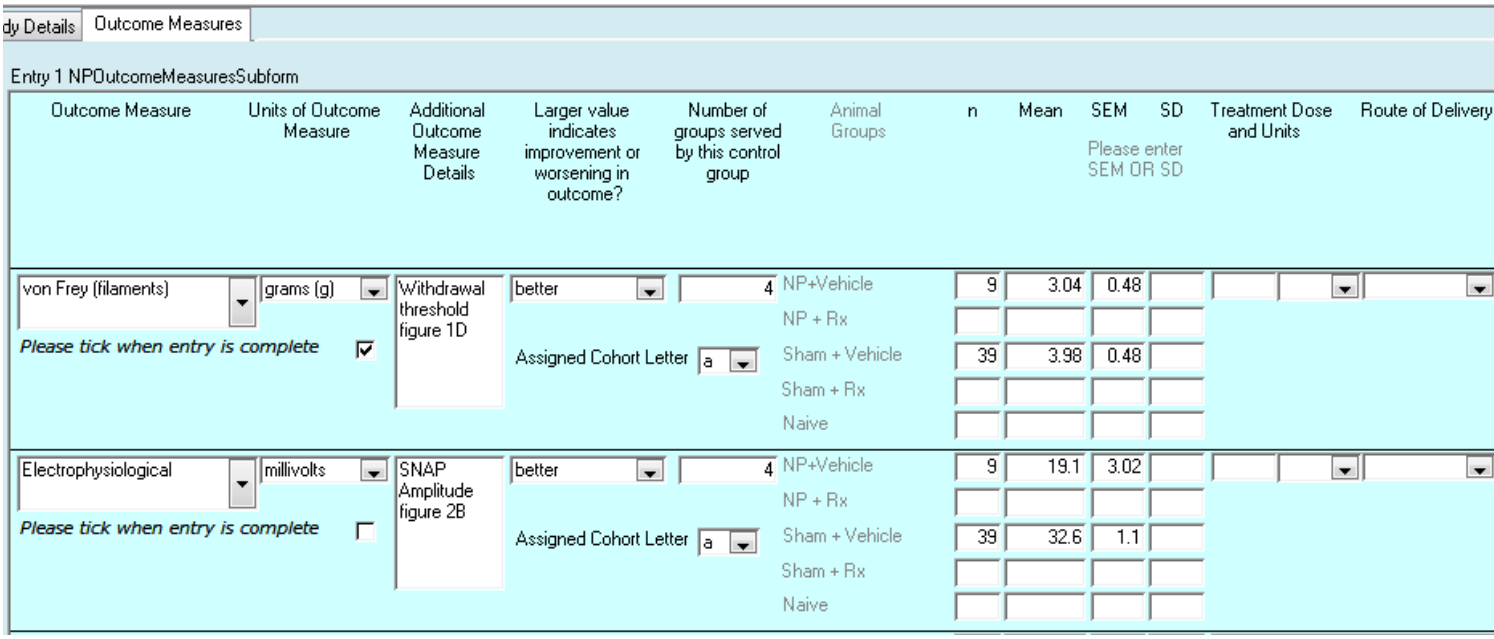
Current Process of Data extraction

Below is a screen shot from the current CAMARADES database, which shows the outcome measure form that is currently filled in by reviewers. The following explanations refer to the red numbers on the screen shot:

1	Name of outcome measure test and unit it is measured in
2	Here it is recorded which figure number the data is from (for ease of checking it when we third screen) and any other relevant info about the figure/ outcome measure. It is a free text box that isn't analysed but can aid other reviewers
3	To get effect sizes from meta-analysis whether a larger number shows an increase or decrease in pain needs to be recorded, especially as the same experiments (e.g. von frey) could be presented with larger values having EITHER better or worse effects
4	This is also needed for meta-analysis – here is a sentence that is copied from the neuropathic pain protocol, and similar explanations appears in other CAMARADES protocols "Where a single control group serves multiple treatment groups, the size of the control group entered to the meta-analysis will be adjusted by division by the number of treatment groups served," hence this is where this data is recorded.
5	Where data is entered, the stars indicate model control, model group, intervention control and intervention group (explained in 'Data Structures' section). Rx = intervention. Sham + Rx = a model control which is then given the intervention to see if the effect seen in the intervention group is independent of the pain model. We do not usually extract this for NP

Fala Cramond
f.cramond@imperial.ac.uk

	as it is specified in the protocol that model = NP + vehicle vs sham + vehicle, however, in the future other projects (especially if specifically looking at one particular drug) might extract this. Naïve = animals that have not received ANYTHING – no model, no vehicle, no sham surgery. This can be used a control for model if there is no sham but if there is a sham/vehicle then that must be taken over the naive animals.
6	The details of the intervention being tested (only applicable in intervention studies) as one paper might use the same drug at different concentrations or at different times relative to the pain model being induced. Surgery = any intervention to induce pain, which might not necessarily have been surgical.
7	Record what time the outcome measure was taken relative to the first time point of induction of the model
8	Weight is also recorded on the front page of data extraction, so I usually leave this blank
9	This is the last time point of outcome assessment – used to see how long the model of pain was used for



Data Structures

For outcome measure data to be extracted for Systematic Review and, crucially, Meta-Analysis the number (n) of animals/ samples, the mean and a measure of variance need to be presented.

Therefore, for the neuropathic pain dataset, the types of data that can be extracted are:

Mean & standard error of the mean (SEM)
Mean & standard deviation (SD)
Mean & confidence intervals or confidence limits (CI or CL)
Median & CI or CL

The above Venn diagram illustrates the frequencies of reporting of each of these structures. Usually, if there is a measure of variance but it is not reported whether it is SD, SEM or CI (n=4 in this sample) then for extraction it is assumed it is SEM, the more conservative measure. However, for this purpose they have been left blank.

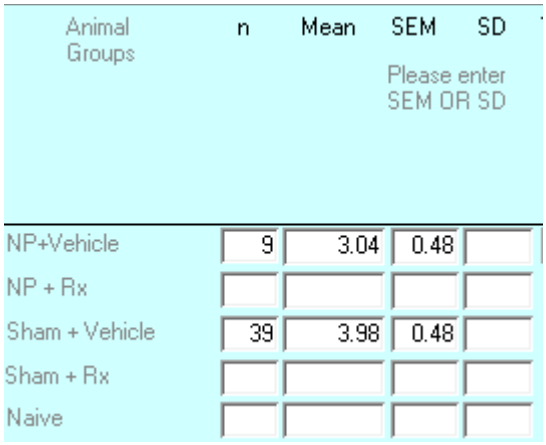
Here are some pasted examples of how the data structures could be described in the text, either in the main body or in the figure legends:

SEM	SD	CI
Data are expressed as mean ±	Data represent mean S.D.	dotted lines represent 95%

Fala Cramond
f.cramond@imperial.ac.uk

S.E.M.		confidence intervals
Variance is expressed as standard error of the mean (sem)	mean S.D.	(CI 95%: 809.0 to 3096)
Values are mean ±SEM.	(mean ± SD)	Data are presented as median ±95% confidence limits
(bars: SEM).		0.7 (95% C.L. = 0.5–1.0)
		(95% confidence limits, dotted lines)

The current process of entering this type of data is shown in the blue screen shot below, as also shown above. The UI would need to be similar in regards to the components it has for the data (n, mean, and variance (SEM, SD and CI).



When entering this data, the reviewer has to specify what type of preclinical experiment it is, namely whether it is a model characterising or intervention testing outcome measure. This is done above by the data being entered in the relevant rows, however this is not a very efficient way of recording this and it is easy to accidently enter it into the wrong row

Model characterising experiments are used to test the validity of a model. Examples of animal models of neuropathic pain are where animals have undergone chemotherapy, a nerve ligation, a spinal cord injury or injection of HIV **(NP + Vehicle)**. The amount of neuropathic pain induced in these animals is then compared to an appropriate control **(Sham +Vehicle)**, namely an animal that has received the sham equivalent (e.g. surgery but no ligation, saline instead of the drug).

Drug experiments are used to test the efficacy of an intervention (drug compound, genetic alteration, diet) and this is done by comparing animal models of neuropathic pain who are given the intervention **(NP + Rx)** with animal models of neuropathic pain who are not exposed to the intervention **(NP + Vehicle)**, and instead often given a control equivalent (saline, vehicle, normal diet).

As model and intervention experiments are analysed separately, the UI needs a function for the reviewer to define and set the outcome measure as a model or drug comparison. As, for other studies, there might be other categories that need to be separately analysed, a function allowing the reviewer to set their own experiment type would also be useful.

The frequency of model or intervention studies are usually consistently, similar. In the existing data on chemotherapy-induced neuropathic pain there are 854 model comparisons vs. 1146 intervention comparisons.

What would be desirable for UI:

As mean and variance are always entered together – need the boxes to enter data next to each other with an easy way of flicking from one to the next (tab button?).

Could perhaps have it automatically set to mean and SEM so it doesn't have to be selected manually each time, but an easy way to change it to a different type of structure would also be necessary.

A way to distinguish between type of study – i.e. model and intervention

To always have the control entered in the same place, for example, the control could always be entered in the right hand box, because the current system means sometimes the control is the third row (for model studies) or the top row (for intervention studies) and this room for error can cause discrepancies between reviewers.

Graph types:

Below are some examples of the different types of graphs. It is difficult to present a completely holistic overview of the graph types that might be in a publication as it is very heterogeneous, but I have included an exemplar graph of every type I came across when coding the initial 34 papers, and these can be put into 5 main categories.

How I labelled the graphs could be interpreted as being a little arbitrary, as I haven't taken into account what the data they are presenting is saying and whether or not they would actually be included in data extraction. However, I have labelled them according to how they look and the key criteria/ identifying features that I used are listed and depicted below:

Line graphs (A) have lines connecting each point. The x axis is often (not always) time, and therefore the graph would be depicting continuous change of y over time.

Scatter graphs (B) are, simply put, line graphs minus the line.

Dot plots (C) are graphs that show each individual animal/ neuron/ experimental unit. They might still show the mean/median and a unit of variance but they demonstrate how each animal scored on that particular experiment.

Bar charts (D) and box and whisker plots (E) are more easily distinguished and fairly self-explanatory.

There are also data presentation images marked as 'representative images'. These contain no quantitative data, and therefore for extraction purposes they can be mostly ignored. However, they are often presented in figures alongside the quantified data, which is extracted

Fala Cramond
f.cramond@imperial.ac.uk

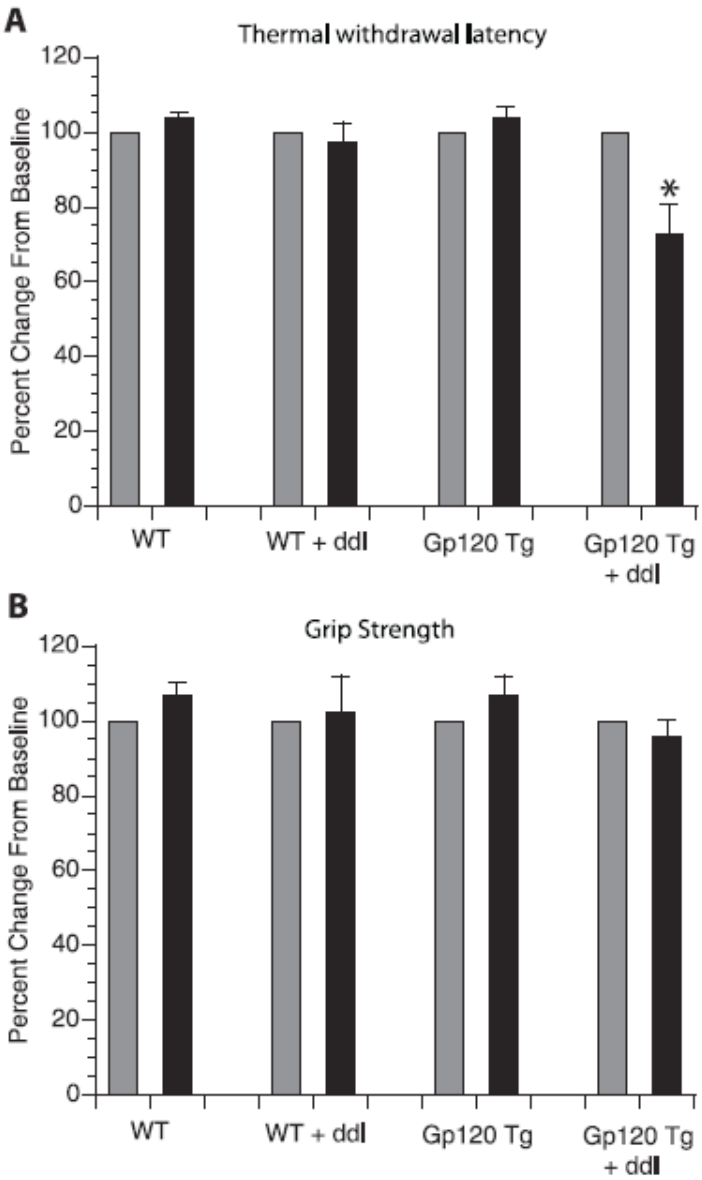


Figure 4. Thermal sensation and motor strength in gp120 transgenic mice treated with DDI. ***A***, Paw withdrawal latency to thermal stimulation was done at baseline and after 4 weeks of oral DDI treatment. Only DDI-treated gp120 transgenic mice developed thermal hyperalgesia. *$p <$ 0.05 compared with gp120 transgenic mice. ***B***, Motor function using grip strength testing was done at baseline and after 4 weeks of oral DDI treatment. In both graphs, the results are expressed as a percentage change from the baseline. Gray bars denote baseline, and black bars represent repeat testing after 4 weeks of DDI or control. Error bars indicate SE.

Fala Cramond
f.cramond@imperial.ac.uk

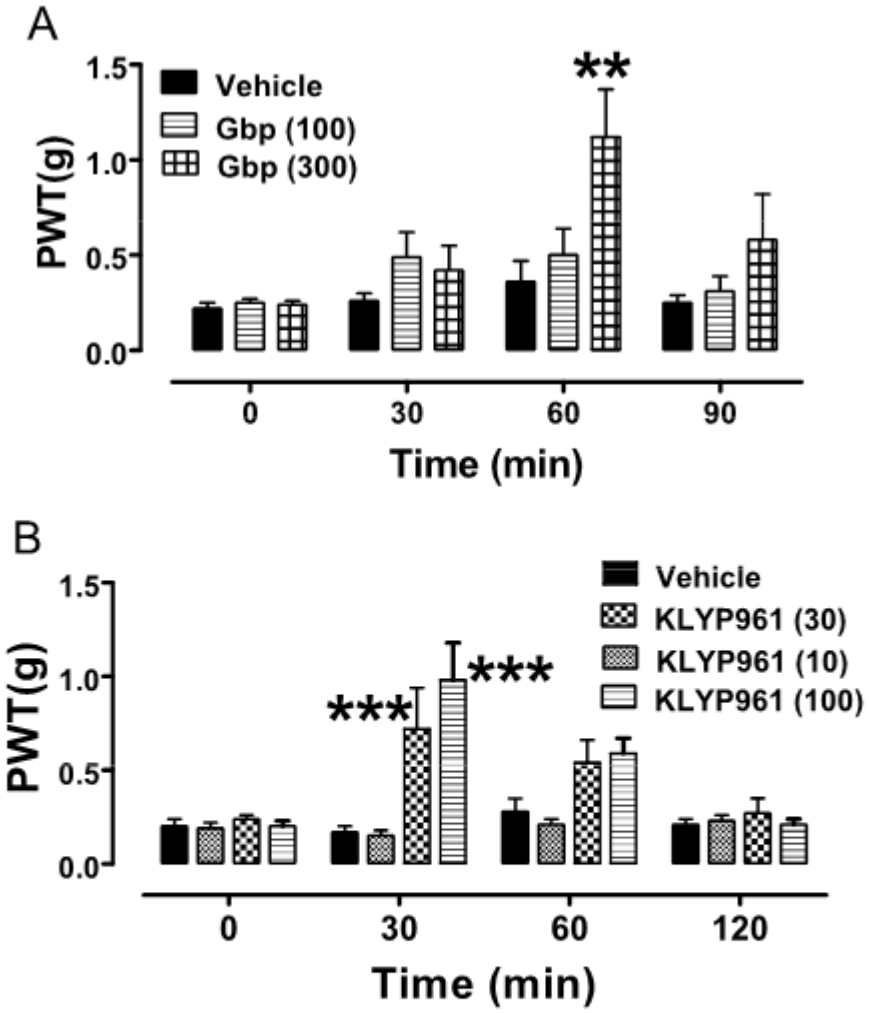


Fig. 5. Orally administered gabapentin (Gbp; A) and KLYP961 (B) attenuate tactile allodynia response after spinal nerve ligation (Chung model) (Kim and Chung, 1992). Tactile allodynia was assessed by monitoring paw withdrawal (PWT) in response to von Frey filaments. Data represent mean ± S.E. (n = 5–6). **, $p < 0.01$ and ***, $p < 0.001$, all relative to vehicle (two-way ANOVA followed by appropriate post hoc test).

Fala Cramond
f.cramond@imperial.ac.uk

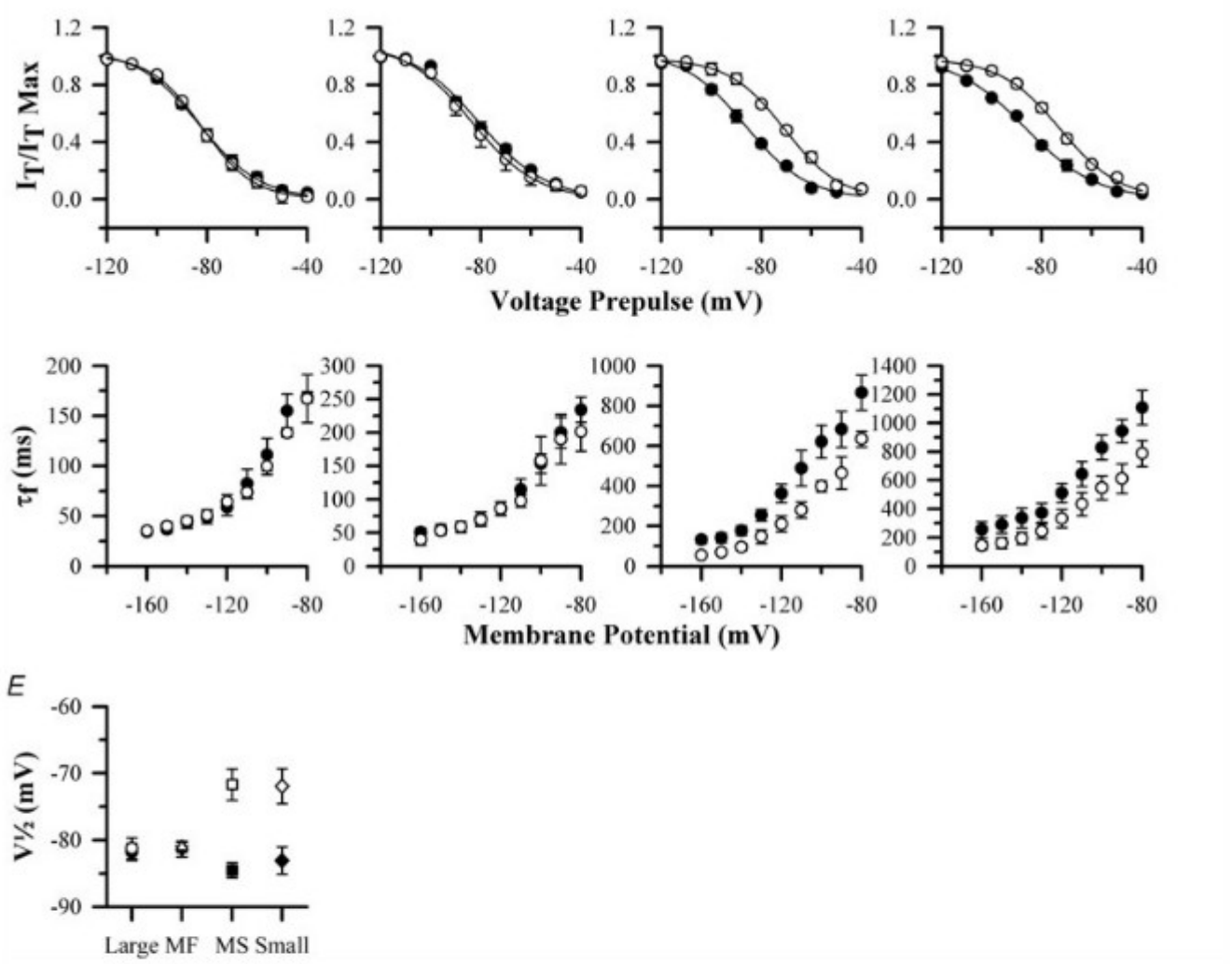
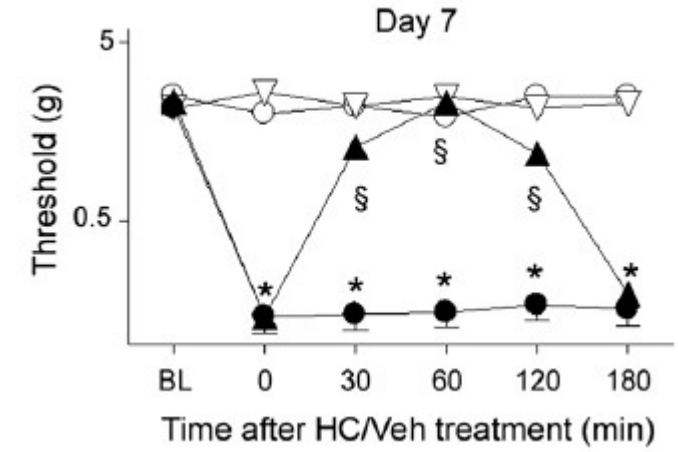
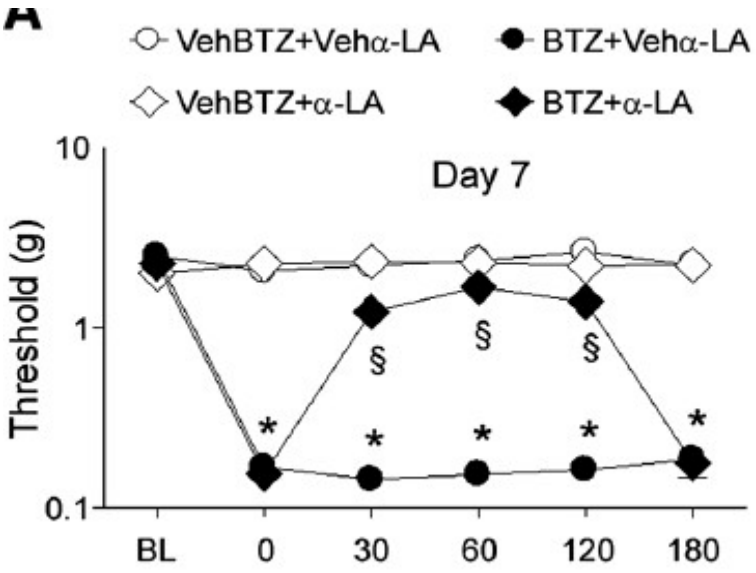


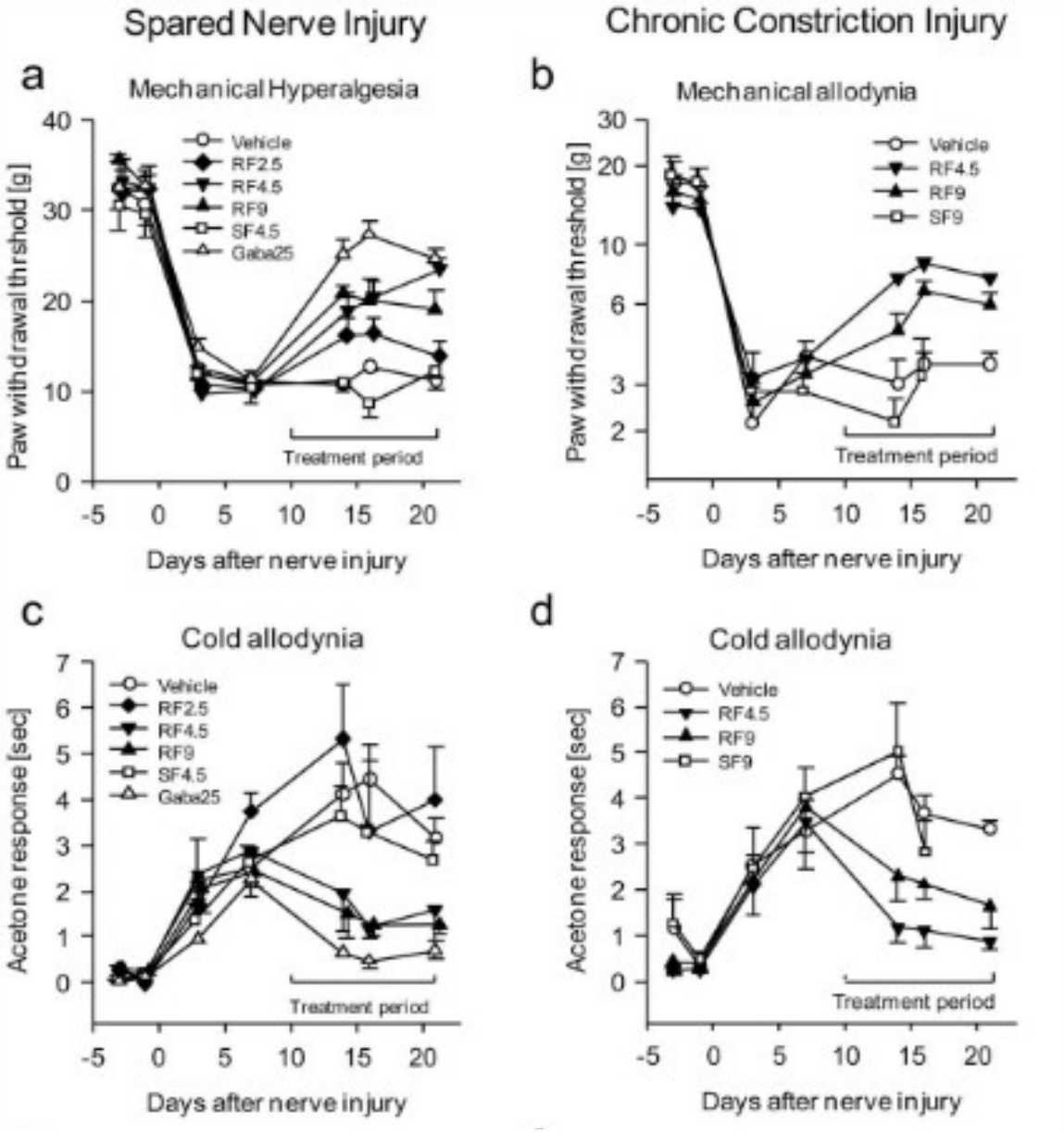
Figure 6. Effect of elevation of cAMP on properties of I_h in rat DRG neurons of different sizes
A, *B*, *C* and *D* show properties of activation of I_h in rat large (*A*), medium fast (MF, *B*), medium slow (MS, *C*) and small DRG neurons (*D*). Upper panels show representative current traces in response to a voltage pulse from −60 mV to −90 mV in the absence (●) and presence (○) of 50 μM FK. Middle panels show steady-state activation curves of I_h as a function of membrane voltage over the range −120 mV to −40 mV in the absence (●) and presence (○) of 50 μM FK. Lower panels show τ_f in absence (●) and presence (○) of 50 μM FK. *E*, effect of 50 μM FK on values of $V_{½}$ in the four populations of rat DRG neurons shown in *A*–*D*. Absence and presence of 50 μM FK is shown by filled or open symbols, respectively. Experiments repeated n = 10–15 times in each neuronal population.

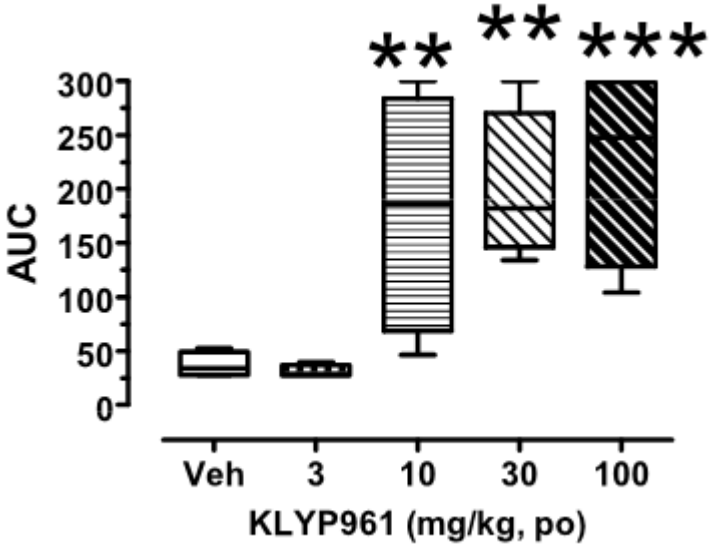
Fala Cramond
f.cramond@imperial.ac.uk



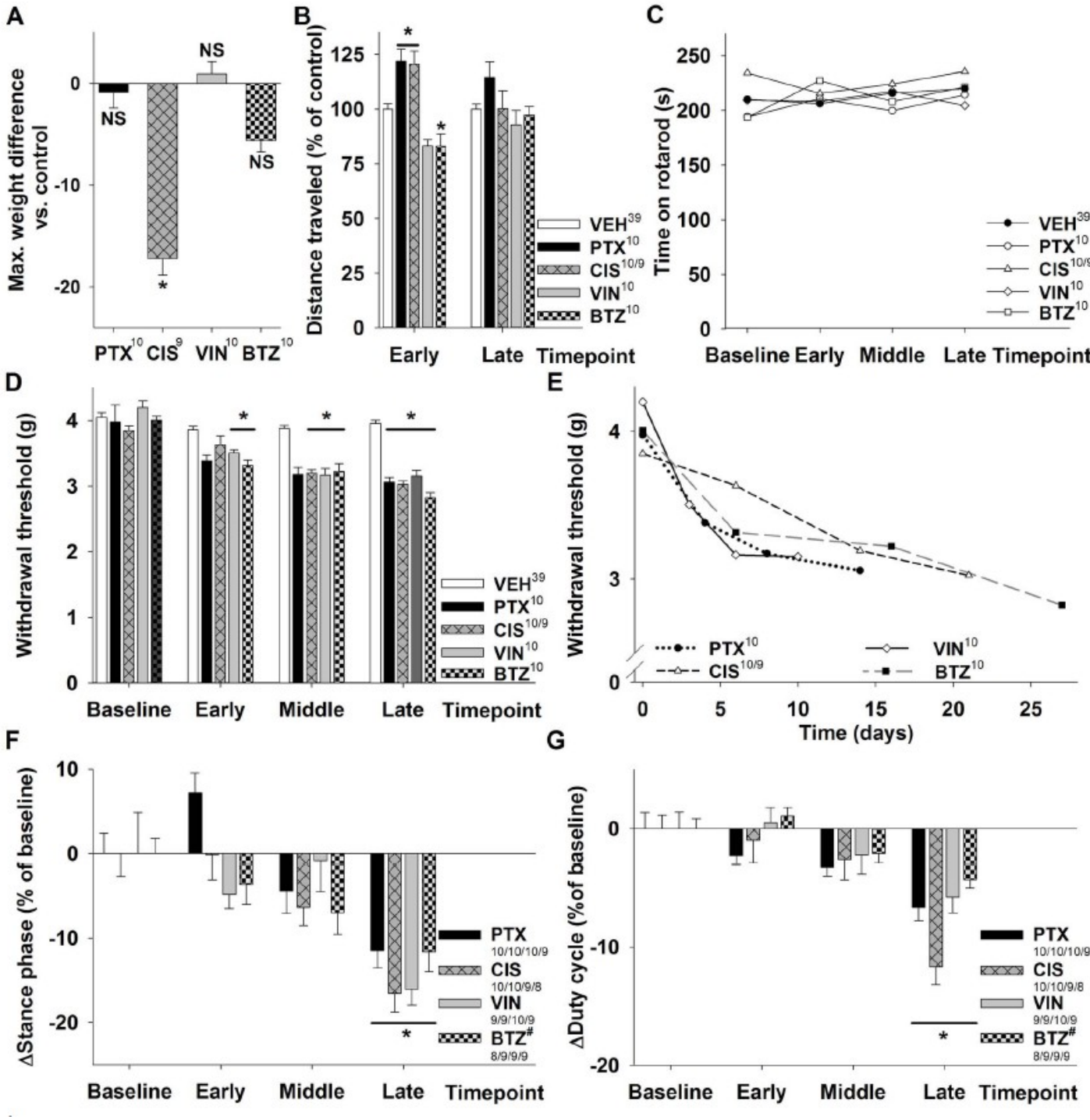


Fala Cramond
f.cramond@imperial.ac.uk





Fala Cramond
f.cramond@imperial.ac.uk



Process of data extraction from graphs

When extracting data from graphs, I use Universal Desktop Ruler to measure and set the axis, and then I measure the model/ intervention group and the control and both sets of errors bars. Relevant graphs that display multiple data points are treated differently depending on the review (and its preceeding protocol). For neuropathic pain, we extract the two points with the biggest difference between them (which can be lengthy if it is unclear which points this is), but other studies will have different criteria for this (e.g. extraction of all time points or, as seen in a recent systematic review in stroke, the latest time point recorded). However, universal desktop ruler causes problems if the

axis are atypical (as shown in some examples above), or if they are log axis. It also can cause inaccuracies because if the reviewers forgets to reset the ruler between graphs with different y axis then the numbers will be incorrect and, from experience, this is one of the discrepancies that takes the most time to correct in third reviewing. Extracting data from graphs is the most time consuming part of data extraction for a systematic review because currently one would measure the axis, measure the data and its error bars manually transferring the numbers from the ruler to the form (i.e. write down the numbers, then switch tabs to the data entry form and type them back in). Therefore there is a lot of switching between tasks, lots of room for error and this means most reviewers only record the numbers to 2 decimal places.

In terms of which graphs are the most simple to extract from, bar graphs are probably the easiest as they are quite straightforward to read. Line graphs can be the most difficult to decipher sometimes because of the number of groups authors can put on there, and some may overlap making it harder to read. However, there is so much variation from publication to publication that this is a very general statement.

Lastly, sometimes authors do write the relevant data in the text and in these instances the figures in the text should be extracted. This is not very common as the authors may not put the right data points that you want, but should be checked first before extracting.

Frequency of each graph types

In the 34 coded publications:

Bar chart	28/34
Line graph	28/34
Scatter graph	7/34
Dot Plot	3/34
Box and Whisker	2/34

What would be desirable for UI:

If I was to rank the graphs in order of importance of having machine assistance I would say bar and line graphs are equally the most important, as they occur as often as each other. However, because bar graphs are usually simpler to read and line graphs have more similarities structure-wise to scatter graphs and dot plots paving the way with UI for line graphs could lead to machine assistance for more types. For an overall ranking of how important the graphs types are I would go with the frequency with which they are used (i.e. bar and line graph are equally important, then scatter graph, then dot plot and then box and whisker being the least important and although it would be great to have machine assistance, should not be a priority).

A tool that can read the axis without it having to be manually entered, however, if this is not possible/ will take a while to develop, it would initially be helpful to have a system that automatically fills in the form after you have taken a Universal desktop (or equivalent program) measurement, as currently you take the measurement then flick between tabs to manually type it in (very time consuming, and much room for error).

However, this would need to have an on/off function, because sometimes it might be necessary to (for example) see which time points have the biggest difference between them before you decided which measurement needs to be taken.

Table Types:

11/34 of the coded publications use tables to display their data. They are very variable in both the data they present, how much data they contain and how they are displayed. It is easier and quicker to extract data from a table than a graph and as the numbers will be the same if they are also in the text, there is no need to look for them there. Some

Fala Cramond
f.cramond@imperial.ac.uk

examples of how tables vary are: size, presentation (are there lines dividing the columns/rows), how variance is presented (brackets or ±) and are independent variables in rows or columns.

As per with graphs, the data points with the biggest difference are taken. See some examples below, with boxes or stars indicating what data would be extracted.

Table 1. Motor and sensory nerve conduction studies

	CMAP (mV)	Motor DL (s)	SNAP (μV)	Sensory CV (m/s)
WT	8.9 (1.5)	1.9 (0.1)	40.8 (9.5)	34.9 (4.4)
WT + DDI	9.2 (0.8)	2.3 (0.3)	34.2 (17.7)	33.3 (2.4)
gp120 Tg	8.0 (1.5)	2.2 (0.4)	43.8 (21.2)	35.8 (5.5)
gp120 Tg + DDI	8.4 (1.6)	2.3 (0.2)	31.2 (11.5)*	36.8 (4.6)

Motor nerve conduction studies were done by stimulating the sciatic nerve at the sciatic notch and recording at the sciatic nerve innervated foot muscles. Sensory nerve conduction studies were done by stimulating at the base of the tail and recording 5 cm distally. Numbers in parenthesis denote SDs. Motor DL, Motor distal latency; sensory CV, sensory conduction velocity; WT, wild-type; Tg, transgenic. *$p = 0.6$ compared with gp120 Tg.

Metabolic parameters measured 24 weeks post induction/onset of diabetes in type 1 and type 2 diabetic and control mice.

Strain and Diet	Glycemic Status	Final Weight	Final Blood Glucose	GHb
Type 1 Models				
B6 5001	Nondiabetic	30.1 ± 0.23, n = 9	112.5 ± 2.4, n =9	4.9 ± 0.10, n =8
B6 STZ 5001	Diabetic	23.9 ± 0.34*, n =8	466.6 ± 7.4*, n =8	12.4 ± 0.29*, n =8
B6Ins2^{Akita} 5001	Nondiabetic	27.4 ± 0.24, n =12	126.3 ± 1.4, n =12	6.7 ± 0.07, n =7
B6Ins2^{WT} 5001	Diabetic	24.3 ± 0.13#, n =11	584.6 ± 3.2*, n =11	15.0 ± 0.37*, n =7
B6-db/db AIN-76A	Nondiabetic	27.6 ± 0.43, n =9	110.3 ± 3.0, n =9	4.2 ± 0.16, n =6
Type 2 Models				
B6-db^{+} AIN-76A	Diabetic	49.2 ± 0.41*, n =11	170.3 ± 13.4, n =11	7.3 ± 0.27§, n =9
B6-db/db 5008	Nondiabetic	25.5 ± 0.13, n =9	139.5 ± 0.87, n =9	5.6 ± 0.03, n =9
B6-db^{+} 5008	Diabetic	50.9 ± 0.82*, n =14	407.1 ± 12.9*, n =14	12.8 ±0.63§, n =6
BKS-db/db 5001	Nondiabetic	32.9 ± 0.17, n =11	106.6 ± 1.7, n =11	4.6 ±0.04, n =11
BKS-db^{+} 5001	Diabetic	43.5 ± 0.40*, n =14	378.7 ± 5.8*, n =13	9.3 ± 0.20*, n =12

Diabetic (STZ injected or db/db) mice had significantly elevated levels of blood glucose and GHb (*$p < 0.001$, #$p < 0.002$). In the type 1 models, STZ treated C57Bl diabetic B6Ins2^{Akita} mice (*$p < 0.001$) compared to vehicle and nondiabetic mice. The db/db mice on both genetic backgrounds and all diets gained asignificant an compared to db^{+} mice.

Fala Cramond
f.cramond@imperial.ac.uk

What would be desirable for UI:

Similar to what would be desirable for graphs, it would be useful to have a system that allowed you to click/highlight the data and it be transferred into the extraction form. As even though here there is no need for a manual measuring software, the process of manually transferring it from PDF to form leaves room for error.

Data in the text:

Most papers (25/34) reported some relevant data in the text, which may sometimes be also represented in graphical format. Where there are both representations, the data in the text should be extracted as there is less room for error. Below are some pasted examples of how data could be reported in the text, in bold I have highlighted key words and symbols that tend to indicate data in the text:

the withdrawal threshold of the AM 251/WIN group (2.1 **±** 0.4 g) **was significantly less than that of** the vehicle/WIN group (15 **±** 0.0 g; $p < 0.05$) and **was not different from the threshold** of the vehicle/vehicle group (2.4 **±** 0.3 g; $p > 0.05$)

In all the injury groups, **the average value for the mechanical threshold** in the contralateral paw was similar (19.1 g **±** 0.2) with values ranging from 17.4 g **±** 1.6 to 20.2 g ± 0.9

similar to the levels that expressed ATF3 (8.9% **±** 0.8% eccentric nuclei of total lumbar DRG neurons in paclitaxel-treated rats vs. 1.7 **±** 0.6% eccentric nuclei of total lumbar DRG neurons in vehicle-treated rats

compared to vehicle-treated rats (Figs. 6E, F) (332 **±** 19 CD68 IR cells)

Mean mechanical responses for contralateral (125.9 G 0.0 g) and ipsilateral (116.5 G 9.4 g)

mean percentage decrease from baseline PWTs was 22.1% **±** 1.02 and 18.2% **±** 1.14

The decrease was 24.8% **±** 7.2 and 37.2% **±** 3.7 at day 7 and day 14, respectively, for the contralateral paw (Fig. 3A and B).

Although the majority of papers report data in the text, from experience, it is fairly uncommon for authors to report the specific data you need to extract in the text, as they often will not state all the necessary components for the necessary time points, especially in neuropathic pain, where not all the time points are to be extracted. However, the reviewer needs to check the text before extracting from a graph, as the authors reported values are obviously the true values. Unfortunately, this is usually a time consuming and redundant exercise.

What would be desirable for UI:

In a similar theme to tables, a feature that allows the reviewer to highlight the numbers and it be transferred into the data entry form.

Perhaps a system of the machine picking up where figures are mentioned in the text (see pasted examples above) and then a way of linking that text with the relevant graph so it is quicker and easier to flick from text to graph and decipher if what the reviewer needs is presented in the text.

Summary

- There is a wide range of data structures and presentations in in vivo neuropathic pain research
- Data extraction for outcome measure is lengthy for four mains reasons
    - Variation of data means it is hard for a reviewer to recognise patterns therefore the manuscript has to be very carefully read before extracting
    - The extraction of data points requires a laborious manual desktop ruler and manually copying it across

    - Data can be in multiple parts of a paper (e.g. in text and graph, or in table and graph)
    - There are many different bits of information that need to be recorded for meta-analysis

- After careful consideration I believe development should be prioritised in the following order:
    - Development of a tool that can read graphs and transfer the reading into a data entry form
        - By graph this should be done first for line graphs, then bar graphs, scatter graphs, dot plots and finally box and whisker plots
    - A machine that can detect where figures are mentioned in the text and link the text with the figure to reduce the amount of time a reviewer has to scroll between text and graph
    - Data entry boxes that indicate the data structure and have the control group consistently in the same place