Behavioral Pharmacology of Pain

Odd-Geir Berge

Abstract Behavioral methods are extensively used in pain research. Rodent modeling tends to rely on evoked responses but there is a growing interest in behavioral readouts that may capture elements of ongoing pain and disability, reflecting the major clinical signs and symptoms. Clinically, analgesics show greater efficacy in acute pain after standard surgery than in chronic conditions but are never completely effective on a population basis. In contrast, experimental pharmacological studies in rodents often demonstrate full efficacy, but there is variability in sensitivity between models and readouts. Full efficacy is rarely seen when more complex or multiple readouts are used to quantify behavior, especially after acute surgery or in studies of clinical pain in animals. Models with excellent sensitivity for a particular drug class exist and are suitable for screening mechanistically similar drugs. However, if used to compare drugs with different modes of action or to predict magnitude of clinical efficacy, these models will be misleading. Effective use of behavioral pharmacology in pain research is thus dependent on selection and validation of the best models for the purpose.

Keywords Pain models · Neuropathy · Osteoarthritis · Arthritis · Rodent · Rat · Mouse · Dog · Behavior · Wheel-running · Burrowing · Analgesia · Predictive validity · Opioid · Nonsteroidal · Spontaneous pain · Evoked pain · Screening · Validation

O.-G. Berge
Multidisciplinary Pain Center, Uppsala University Hospital, Uppsala, Sweden

O.-G. Berge (✉)
OGBConsulting AB, Ekbacksvägen 45, SE 144 63 Rönninge, Sweden
e-mail: oddgeirberge@gmail.com

DOI: 10.1007/7854_2014_281
© Springer-Verlag Berlin Heidelberg 2014
Published Online: 4 February 2014
1 Introduction

The concept of behavioral pharmacology of pain includes the use of pharmacological tools to study pain-related phenomena like nocifensive mechanisms and pain pathophysiology, as well as the use of behavioral models to predict the clinical efficacy of pharmacological agents. In spite of a heated discussion about the clinical relevance of behavioral rodent models in pain research, particularly with regard to their usefulness in predicting clinical efficacy of novel analgesics (Mao 2009; Mogil et al. 2010; Quessy 2010; Rice et al. 2008), there is evidence that their use is increasing (Mogil et al. 2009). Both the discussion and some failed predictions are due in part to a lack of appreciation for the necessity of choosing appropriate modeling strategies based on the research purpose. Investigating the functional significance of novel mechanisms, characterizing models for defined pain conditions, screening, and selection of compounds at various stages of a drug discovery project are each activities with different requirements for sensitivity (the ability to predict true efficacy) and specificity (ability to detect negative outcomes). Pain is a clinical problem both in man and animals and there is a significant overlap between veterinary and human medicine, for instance regarding procedural or post-surgery pain, which may represent a valuable translational link.

2 Publication Bias and Environmental Factors

Publication bias, i.e., the favoring of publishing positive data over negative or neutral results, has been documented in many areas of research and also in the context of pain modeling (Currie et al. 2013). Reproducibility of results is less than perfect. One study found that only 20–25 % of reported studies could be satisfactorily replicated in a target validation program (Prinz et al. 2011). Data from another group confirmed 11 % of published conclusions and success was tied to quality indicators like adequate controls, evidence of bias reduction, and complete
description of the data (Begley and Ellis 2012). Procedures may be highly standardized in a particular laboratory, especially in the pharmaceutical industry and in other settings where standard operating procedures are in place to provide consistency when models are used routinely. There are nevertheless environmental and procedural factors that create variability within and between laboratories even when all reasonable measures have been taken to enhance reproducibility. A within-laboratory analysis of thermal nociceptive data from genetically diverse mouse strains found that environmental factors, and particularly the person carrying out the experiments, were more important than genotype (Chesler et al. 2002). Season, humidity, cage density, time of day, sex, and within-cage order of testing also affected the outcome and it should be noted that only some of these factors can be controlled by adequate randomization. In a classic paper comparing behavioral outcomes related to motor and cognitive functions, anxiety, and pharmacology in eight strains of mice across three laboratories, several outcomes where highly laboratory-dependent (Crabbe et al. 1999). Factors considered important, e.g., apparatus, housing and acclimatization, age of animals, and timing of experiments, were extremely well controlled. There was an expected genetic effect on behavior, but somewhat surprisingly, less important was sex, or whether the animals were locally or externally bred with associated differences in transportation. The authors concluded that large genetic effects are likely to be reproducible across laboratories while smaller effects may be more susceptible to laboratory-specific variables.

In the context of behavioral pharmacology, it seems reasonable to assume a similar relationship between pharmacological effect and reproducibility. A high degree of standardization may increase the risk of overestimating spurious effects, and so it has been suggested that measures to systematically increase heterogeneity may improve the external validity of results (Richter et al. 2011). It may be worth considering that the use of pet animals as subjects in studies of painful veterinary conditions automatically provide heterogeneity and a degree of randomization that is difficult to obtain in a traditional laboratory setting. The impact of environmental and genetic factors will vary between paradigms. This should always be considered when different results are obtained with seemingly similar methods and before results are generalized from a particular model to clinical conditions in man or animals.

Whereas experimental variability is unavoidable, bias and inadequate reporting of methodological details add avoidable uncertainty to interpretation of the animal research literature (Kilkenny et al. 2009). Recognition of these problems has motivated recommendations and checklists, some that specifically focus on pain research (Rice et al. 2008), while others are of general scope (Kilkenny et al. 2010; Muhlhausler et al. 2013). The greatest impact has been achieved with the ARRIVE guidelines (http://www.nc3rs.org.uk/page.asp?id=1357), currently supported by more than 300 journals (June 2013; http://www.nc3rs.org.uk/page.asp?id=1796). Several pain journals have adopted these guidelines and an example of how they may be implemented can be found in Huang et al. (2013). In the future, hopefully more consistent reporting will facilitate proper meta-analysis of preclinical data in the pain field.
3 Symptoms, Signs, and Readouts

Research in behavioral pharmacology is often motivated by human disease. A thorough description of pain in man is beyond the scope of this chapter, but some points are worth mentioning. Clinically, acute pain is usually studied in the context of surgery. Postsurgical pain is characterized by ongoing pain exacerbated by movements (Brennan 2011). Pressure pain thresholds are reduced around the wound and inversely correlate with spontaneous pain and pain during movement (Møiniche et al. 1997). Minor surgery, such as removal of impacted molars, is frequently used as a human model in drug discovery, typically with readouts reflecting ongoing pain and its reduction. Clinical trials of new analgesic drug candidates are usually carried out in chronic conditions.

In osteoarthritis, the major complaint is pain at rest and during movement, although mechanical and thermal thresholds may be lower both near the affected joints and at more distant sites, possibly reflecting different pathophysiological mechanisms. Lower thresholds have been correlated with higher visual analog scale (VAS) scores and found to be normalized upon effective pain relief (Smith et al. 2012). Functional parameters are often used to quantify treatment effect, but pain reduction does not always result in improved function; in one study, walking speed declined in 20% of patients despite reduced pain (White et al. 2011).

Neuropathic pain is associated with a range of symptoms and signs (Jensen et al. 2001). The common denominator for neuropathic pain is that it is caused by a lesion or a disease affecting the somatosensory system (Treede et al. 2008) but symptoms and pharmacology differ between patients and diagnoses. Spontaneous pain (apparently stimulus-independent), may be continuous or paroxysmal and of different qualities. Stimulus-evoked pain, most commonly involving a mechanical or thermal stimulus, can either be time locked to the stimulus or persistent, representing a sliding transition toward spontaneous pain. Sensory disturbances, either loss of sensation, elevated thresholds, or positive sensory phenomena such as allodynia, hyperalgesia, and paroxysmal pain evoked by suprathreshold stimulation in areas with elevated sensory thresholds (hyperpathia) are not specific to neuropathic pain. Other symptoms such as superficial, ongoing pain, and pain evoked by touch or cold may be more common in neuropathic patients than in non-neuropathic patients (Rasmussen et al. 2004). Spontaneous pain, whose characteristics may differ significantly between patients, is the most frequent complaint and the most commonly used primary readout in clinical studies (Novak and Katz 2010). Increased sensitivity to stimulation in any specific modality is seen in about one-third or less of patients and sensory loss is common (Maier et al. 2010).

In animal models, evoked withdrawal responses are commonly used as surrogate endpoints for pain; however, these measures are associated with a number of methodological problems. First, withdrawal may be either reflexive or voluntary. Second, stimuli that gradually increase in intensity during application will by default first activate low-threshold, non-nociceptive sensory nerves, which may be sufficient to elicit a response. Third, visual or auditory cues may cause motor
responses indiscernible from reflexive withdrawal, particularly when repeated stimulation creates an opportunity for conditioning. Fourth, both application of stimuli and assessment of responses may be difficult to standardize and reproduce between observers and laboratories. If correctly tested, interpreted, and reported, evoked responses may, nevertheless, provide relevant information about sensory functions. For an extensive review and analysis of transient nociception, including discussion of electrical stimuli and vocalization, see Le Bars et al. (2001). This chapter will briefly discuss the use of acute thermal and mechanical stimuli, and then emphasize alternative readouts.

**Heat.** Heat is an adequate pain stimulus that was initially used in human psychophysical studies, (discussed in (Beecher 1957)) and applied to rodents in the tail-flick assay (D’Amour and Smith 1941). In man, progressive heating of the skin causes burning pain followed by stinging pain. Stinging pain exhibits heat response-latency correlations similar to tail-flick and other withdrawal responses in animals (Le Bars et al. 2001). The effective stimulus is determined by heat source, energy supplied, initial temperature of the skin, and skin color (Berge et al. 1988; Luukko et al. 1994; Tjølseth et al. 1989; Wen et al. 2009; Winder et al. 1946). Depending on the tissue temperature achieved and the rate of heating, different classes of nociceptors are sequentially recruited (Le Bars et al. 2001; Schepers and Ringkamp 2009; Yeomans and Proudffit 1996; Yeomans et al. 1996). The current standard for heat stimulation in rodents is the radiant heat paw withdrawal assay, with response latency as the readout (Hargreaves et al. 1988). The assay is sensitive to potentially confounding factors like posture, exact focus of the beam, and the initial temperature of the skin, which may be altered by inflammation or neuropathy as well as by handling, conditions in the testing environment, and by confounding pharmacological effects (Bennett and Xie 1988; Dirig et al. 1997; Luukko et al. 1994). Preheating the floor may allow better consistency of the stimulus function. Starting from a higher temperature favors slower heating rates, a factor that may tune the assay toward C-fiber-mediated responses.

**Cold.** As for heat, both rate of temperature change and actual tissue temperature determine the recruitment of different cold fiber types and thus the mechanisms activated by a cold stimulus (Belmonte et al. 2009; Foulkes and Wood 2007). Commonly used in rodent neuropathic pain modeling, the application of acetone produces a cold sensation even in man and may evoke allodynia in a minority of chronic pain patients (Rasmussen et al. 2004). A testing sequence is initiated by application of a fixed volume of acetone to the plantar skin and the response, defined as brief withdrawal of the paw, may be scored after a predefined number of evenly spaced applications (Choi et al. 1994). While the tactile component may be insignificant in the conventional procedure (Taylor et al. 2007), the use of acetone or ethyl chloride spray adds a mechanical component to the cold stimulus. In these cases, the outcome variable is usually the accumulated duration of paw withdrawal during a defined poststimulus interval (Dowdall et al. 2005; Gustafsson and Sandin 2009).
Tactile, Non-Noxious. Mechanical sensitivity in rodents is most commonly determined as a response threshold or response frequency to application of von Frey filaments of different stiffness. The response criterion is usually a brisk withdrawal. The method is associated with a number of drawbacks. First, the applied pressure changes during application: contact area and geometry depend on degree of bending and compliance of the tissue; force generation varies with application speed; and unpredictable off-axis forces are created as the fiber bends (Bove 2006). Second, conventional filaments are hygroscopic and the bending properties, particularly of thinner filaments, will change rapidly in response to normally occurring fluctuations in relative humidity (Ångeby Möller et al. 1998). These factors lead to increased variability and even systematic errors if randomization is inadequate. Mechanically stable filaments with a fixed tip size can circumvent some problems (Frühstorfer et al. 2001; Song et al. 1999) and force transducer-based devices facilitate standardization of stimulus application (Ångeby Möller et al. 1998). Protocols designed to reduce inter- and intra-observer variability have been suggested, taking into consideration duration of application, stimulus interval, degree of bending, and site of stimulation (Hogan et al. 2004; Song et al. 1999).

Tactile, Noxious. Mechanical hyperalgesia by the pinprick method is tested by application of a sharp object to the skin, causing indentation but no penetration. The response, defined as a brisk withdrawal or a “hyperalgesia-like” response with sustained elevation, licking, and grooming of the paw, can be quantified as duration of evoked behavior (Erichsen and Blackburn-Munro 2002; Hogan et al. 2004). Sensitivity to deep pressure is usually tested by application of a constant or gradually increasing stimulus to a paw with withdrawal, struggling, or vocalization as endpoints and the pressure at which the response occurs as readouts (Pradhan et al. 2010; Randall and Selitto 1957; Whiteside et al. 2008). It should be kept in mind that these endpoints may represent different sensory and emotional experiences.

4 Readouts that Do Not Involve Experimenter-Applied Stimuli and Evoked Responses

Alternatives to evoked responses have been introduced to model the subjective experience of pain and also to model pain-relevant disability. These measures can be categorized as general behaviors that are depressed by pain or disability, as behavioral manifestations that preferentially reflect pain, or as operant behaviors.

General Behaviors. Analysis of general behavior to quantify pain and analgesia is common in veterinary medicine. Therapeutic intervention in individual subjects requires readouts that allow immediate assessment, while studies of therapeutic efficacy are better served by data collection over longer periods of time. Some parameters are recorded by standardized, objective, automated procedures. This
can be done in the home cage, thus minimizing observer–subject interaction and allowing continuous data collection (Miller et al. 2011a; Roughan et al. 2009). Food and water consumption as well as motor activity may be used to characterize postsurgical pain and disability (Liles and Flecknell 1993; Jacobsen et al. 2012). Even experimental models may use automated registration of locomotion and rearing (Cho et al. 2013). Thigmotaxis (the tendency to stay near the walls, e.g., in an observation chamber) may be used as an index of increased anxiety caused by a presumably painful manipulation (Huang et al. 2013).

**Wheel-running and Burrowing Activity** are other innate behaviors suppressed by pain and disability. Access to a running wheel may be free, allowing continuous registration of activity, or restricted to certain time slots. Several parameters can be registered, including distance traveled, speed, and in the case of continuous monitoring, circadian activity pattern. Both in relative and absolute terms, mice are more active and regular runners than rats, and, therefore, mice may in effect get a substantial exercise effect of voluntary running (Allen et al. 2001). Mice show great individual, sex and strain differences in wheel-running time (Lightfoot et al. 2010; Siepka and Takahashi 2005), but activity is stable within subjects (Knab et al. 2009). Compared to spontaneous locomotion in the home cage or in an open field, much greater distances are covered with free access to a running wheel (Costello et al. 2010). The high level of activity in some mouse strains provides a solid basis for studying the effect of presumably painful procedures and may increase the impact of movement-related pain. Most studies find that a run-in period of a few days is sufficient to stabilize the activity level in mice, whereas a much longer time may be required for rats (Stevenson et al. 2011). Similar to other spontaneous behaviors, wheel-running in mice is generally reduced by surgery and in experimental models of inflammation (Adamson et al. 2010; Clark et al. 2004; Cobos et al. 2012; Krug et al. 2009; Miller et al. 2011b; Sluka and Rasmussen 2010). Wheel-running is also reduced in spontaneously occurring osteoarthritis in aged mice, where reduced performance is correlated with disease severity (Costello et al. 2010).

It is not clear how useful this assay is in rats. Voluntary wheel-running was reduced following carrageenan injection into the hind paw but not the tail (Loram et al. 2007). This indicates that it is possible to dissociate pain related to limb usage (after paw inflammation) from the general discomfort of inflammation (after tail inflammation). Monosodium iodoacetate-induced monoarthritis also reduced wheel-running, but the effect was small compared to decreased weight-bearing and increased sensitivity to von Frey (Stevenson et al. 2011). Stavudine is an anti-retroviral drug used in the treatment of HIV where it is associated with painful neuropathy and also used to induce neuropathy in a rodent model (Huang et al. 2013). Long-term treatment with stavudine did not change voluntary wheel-running, food intake, and weight gain in rats, even though it increased sensitivity to pressure (but not heat) applied to the tail (Weber et al. 2007).

Wheel-running and other forms of motor activity may interact with nociception, but the data are not consistent; prior running has been reported both to increase and
reduce the responsiveness in the rat tail-flick assay (Kanarek et al. 1998; Spradley et al. 2012). Wheel-running exercise for 2 h, 24 h prior to testing increased bilaterally the response to plantar von Frey stimulation in mice with a unilateral muscle inflammation but not in normal mice, assumed to be an effect of muscle fatigue (Sluka and Rasmussen 2010). Effects of exercise on nociception are not limited to wheel-running paradigms. For example, repeated swimming exercise reduces the response to formalin and decreases the behavioral hypersensitivity in rat and mouse models of peripheral neuropathic pain (Kuphal et al. 2007).

**Burrowing.** Burrowing has been used to study neurobehavioral functions in several species, including rats and mice (Deacon 2006, 2009). The assay measures the amount of material removed from a tube in the home cage during a defined time period and does not require observer interaction. Recently, the behavior has been implemented as a readout in mice after laparotomy and in colitis (Jirkof et al. 2010, 2013). In rats, burrowing is reduced in several traumatic neuropathy models (Andrews et al. 2012; Lau et al. 2013). Although mechanical responsiveness was increased in these models, there was no correlation between burrowing deficits and mechanical response thresholds. This indicates that the measures reflect different aspects of pain. Burrowing was also reduced in stavudine-induced neuropathy (Huang et al. 2013) and in the rat model of inflammation induced by intraplantar injection of Freund’s complete adjuvant (FCA; Andrews et al. 2012; Rutten et al. 2013a, b).

**Ongoing Pain-Like Behaviors.** Behaviors assumed to be evoked by pain, such as limping, licking, or favoring of a paw, are well established as readouts in models like the formalin test and in models of unilateral paw or joint inflammation induced by various agents (Berge 2013). Alternatives are, however, needed for quantification of behavioral changes after certain types of surgery or in multifocal pain. A comprehensive analysis of over 150 individual behaviors and behavioral sequences observed after laparotomy identified transient back arching and horizontal stretching followed by abdominal writhing, and twitching while the animal was inactive as the more robust and quantifiable activities (Roughan and Flecknell 2001). Roughly similar behaviors were affected by vasectomy in mice (Jacobsen et al. 2012). Some similar behaviors are observed in other testing paradigms, e.g., flinching in the formalin test. Other behaviors were sensitive to drug treatment, independent of surgery. These measures included behaviors associated with grooming, among others face grooming following paw licking which is a sequence typical for heat dissipating behavior in the rodent (Berger et al. 1983). The licking-grooming behavior may be a confound when forepaw licking is used as an endpoint in the hot plate test, and is frequently observed in the formalin test, where it should be differentiated from paw lick occurring independent of grooming. In mice, other visual scoring systems used to evaluate pain after surgery are based on the overall condition of the animals, the condition of the skin, fur, or eyes, and on motor coordination and posture (Adamson et al. 2010; Clark et al. 2004).

**Facial Expression.** Arguably, the most innovative approach to readouts reflecting pain in laboratory animals is the analysis of facial expression using grimace scales (Langford et al. 2010). These were inspired by scales used in
nonverbal humans (Williams 2002). Observers tend to focus on the face, which may be an advantage in grimace analyses but a challenge in behavioral paradigms that involve other, or multiple body parts (Leach et al. 2011).

The mouse grimace scale (Langford et al. 2010) is based on scoring of facial expression from sampled video frames. In the original publication, positive scores were synchronized with conventional readouts in inflammatory and nociceptive models of moderate duration and in surgical models. The grimace scores did not correlate to responses in models using shorter-acting stimuli nor in any of three common neuropathic models 1, 7, or 14 days after lesioning. The authors concluded that noxious stimuli of moderate duration and origin in deep structures were most likely to be associated with a “pain face.” The lack of effect in the neuropathy models might indicate that these models primarily display hypersensitivity rather than ongoing pain. At the one-day readout, one would, however expect significant postoperative pain, and the data appear to be in conflict with the results from surgical models in this and another study in mice (Matsumiya et al. 2012) and from a rat study using an adaptation of the mouse scale (Sotocinal et al. 2011). In the rat study, a useful readout was obtained also in the FCA-induced paw inflammation model and in monoarthritis induced by intra-articular injection of kaolin and carrageenan. It should be noted that a baseline grimace score is obtained even in presumably pain-free subjects and the readout is not strictly specific to pain.

Accepting that the experience of pain is multidimensional, with sensory, emotional, and cognitive components, any single readout is unlikely to reflect the complete experience of the animal, although more complex readouts may integrate several dimensions. Another challenge is to understand how the intensity of the sensation is coded by the measurement variable, e.g., whether there are ceiling effects saturating the response at higher intensities of stimulation. For instance, if weight bearing is used as a readout in a monoarthritis model, the maximum observable response occurs when no weight is put on the affected limb but this will not, a priori, indicate maximal imaginable pain. At the lower end, nonpainful sensations may elicit a response or weakly painful sensations may be insufficient. In a behaving animal, the limits may be modulated by competing neuronal activity.

It is well documented that rodents may suppress pain behavior when exposed to novelty factors or stressors, including the presence of a human observer (Kavaliers 1988; Kavaliers and Innes 1988). Transportation of animals from vendor or animal quarters, handling, even when repeated for weeks, and sequential removal from common cages increase stress indicators (Olfe et al. 2010; Prager et al. 2011; Rosén et al. 1992). Other factors that may contribute to variability are single versus multiple housing, cage cleaning, environmental enrichment, and light conditions. Some stress factors are difficult to avoid in a laboratory setting, although many can be minimized by proper routines and randomization.
For practical purposes, a model can be characterized in many ways: by stimulus type and intensity; whether a pathological state is present and if so, type and duration of pathology; and by the observable response. Ethical and practical considerations dictate that truly chronic models are uncommon in rodent behavioral pharmacology; inclusion criteria for a clinical trial in chronic pain is at a minimum 3 months of pain duration and the average in most studies is several years. “Chronic” rodent models typically last for a period of days or weeks after initial insult and only exceptionally are the animals kept for several months. This means that the models as commonly implemented show better face validity for human acute pain than for chronic conditions. Pain intensity may also be lower in animal models than in many clinical conditions and severe pain may in fact be detrimental to behavioral measurements.

**Postoperative Pain.** Postoperative pain is generally considered to be of less medical concern than chronic pain and has not received the same attention from the preclinical research community; however, pain after surgery is surprisingly common (Apfelbaum et al. 2003; Brennan 2011). Treatment often require individualized multimodal approaches which include a combination of drugs, but still renders a large proportion of patients with moderate to severe pain for several days. Some standardized surgical procedures such as third molar extraction resemble animal models in terms of homogeneity of subjects, absence of comorbidity, moderate intensity, and short duration of pain and are frequently used in human experimental studies of new analgesics. They have the advantage of allowing assessment of efficacy without concomitant medication (except rescue medicine). The effect sizes may be higher and the number needed to treat (NNT) values for cyclooxygenase (COX) inhibitors and other compounds may be better than in general surgery and other clinical conditions (Moore et al. 2011). Models of pain due to removal of an impacted third molar is generally considered capable of delivering highly reproducible data (Cooper and Desjardins 2010). Typically, surgery is performed under local anesthesia. Pain builds up over some hours after surgery and remains at a moderate level for about 12 h and gradually subsides over the following days. In a dose response study of valdecoxib given preoperatively, the higher doses reduced average pain scores from moderate to mild (more than 50 % compared to placebo) whereas in bunionectomy, a procedure considered to be more painful, the reduction was approximately 30 % (Desjardins et al. 2002).

In clinical studies of postsurgery analgesia, fixed doses of a nonopioid study drug is usually added to an opioid. The contribution of the nonopioid analgesic to the total analgesic effect is reflected in improved analgesia and reduced opioid consumption—this has been demonstrated for acetaminophen, nonselective COX inhibitors, and COX-2 selective inhibitors (Maund et al. 2011). Acetaminophen and a COX inhibitor may be combined for greater efficacy (Ong et al. 2010). Even gabapentanoids have an opioid-sparing effect (Tiippana et al. 2007; Zhang et al.
Pain is reduced but not abolished by pharmacological treatment in these conditions, in contrast to many animal modeling studies. Standardized surgery in animals can be used to model postoperative pain whether performed for that specific purpose (Brennan 2005), as part of an experimental manipulation (Lascelles et al. 1995) or for medical reasons, e.g., in pet animals (Hansen 2003). In rats and mice, mechanical and thermal sensitivity is increased after incision of the plantar tissue (Brennan et al. 1996; Field et al. 1997; Pogatzki and Raja 2003). Both types of responses are reduced by reference analgesics, but efficacies and potencies may differ within and between studies. For instance, morphine completely reversed thermal and mechanical hypersensitivity in mice at 10 mg/kg while a lower dose only affected the thermal response (Pogatzki and Raja 2003). A similar relationship was found in rats (Field et al. 1997) and in this study gabapentin and pregabalin yielded full analgesic efficacy on von Frey thresholds, but only partial effects on thermal readouts.

In another rat study using two different mechanical assays, COX inhibitors showed efficacy and potency slightly higher on thresholds measured by an electronic von Frey device than on paw pressure. An exception was indomethacin, which was remarkably effective in reducing responses to paw pressure (Whiteside et al. 2004). Even gabapentin, in relatively low doses, was efficacious on paw pressure but reversed the von Frey response by less than 20% at a dose of 30 mg/kg. In this study, morphine was the only compound that completely normalized the response in both assays. A standard veterinary dose of the opioid buprenorphine showed full efficacy on mechanical sensitivity measured by paw pressure and von Frey stimulation, but the latter was sensitive to lower doses (Curtin et al. 2009).

The effects of COX inhibitors and opioids are less consistent in other types of rodent post-surgical pain. In dose response studies, the efficacy of COX inhibitors tends to reach a plateau short of full reversal of responses and readouts differ in sensitivity. For instance, in a mouse vasectomy study where several pain-related behaviors as well as corticosterone levels were quantified, meloxicam reduced surgery-induced behavior with no clear dose response relationship, indicating maximum efficacy even at the lower dose (5 mg/kg); however, the highest dose (20 mg/kg) reduced corticosterone levels, indicating an ameliorated stress response (Wright-Williams et al. 2007). Great individual and strain differences were noted both in pain behavior and corticosterone response. High doses of meloxicam and buprenorphine (20 and 5 mg/kg, respectively) were effective in another mouse study that assessed both general behavioral and grimacing (Leach et al. 2012).

In rats both pain-evoked and pain-suppressed behaviors after laparotomy were partially normalized after administration of meloxicam, carprofen, and ketoprofen (Roughan and Flecknell 2001; Roughan and Flecknell 2003). In young rats, daily administration of meloxicam (1 mg/kg) restored weight gain after thoracotomy without affecting this parameter in control animals (Brennan et al. 2009). In this study, buprenorphine had a negative effect on growth rate whether the rats had undergone surgery, general anesthesia only, or no treatment. This finding is not unique and underscores the importance of adequate controls to determine any confounding effects of test drugs. More advanced behavioral controls are possible.
in dogs, e.g., after orthopedic surgery where standardized ethograms and pain scales have been developed for veterinary use (Rialland et al. 2012a).

The frequently observed nonanalgesic effects of opioids may interfere with pain assessment (Roughan and Flecknell 2002). A study in mice using an ethographic scoring system in combination with wheel-running and monitoring of food and water intake found no benefit of buprenorphine either alone or in combination with carprofen, presumably due to nonspecific interference of opioid effects with the scoring system (Adamson et al. 2010). However, in another study, repeated administration of buprenorphine or of an extended release formulation of oxymorphone after splenectomy significantly improved ethographic scores, indicating pain relief; oxymorphone also improved wheel-running performance and weight gain (Clark et al. 2004). Mild pain after surgery may in mice have less impact on spontaneous behaviors and appearance, even when changes in autonomic functions registered by telemetry as well as food intake and nest building may be evident and responsive to treatment by COX inhibitors (Arras et al. 2007).

Osteoarthritis. Osteoarthritis is a favored indication for clinical trials and there is a large number of published studies of COX inhibitors, opiates, and novel candidate drugs. In clinical practice, there appears to be little difference in analgesic efficacy between COX inhibitors in osteoarthritis pain (Conaghan 2012). A meta-analysis of responder rates found that 40–50 % of patients reported more than 50 % pain relief and 20–30 % more than 70 % after treatment with COX inhibitors; interestingly, the reduction in pain was only 8–15 mm on a 100 mm VAS (visual analog scale) compared to placebo (Moore et al. 2010). It is worth considering that a standard experimental animal study would not be powered to detect efficacies of this magnitude. Acetaminophen, often seen as first-line pharmacological therapy for OA, is considered less efficacious than nonsteroidal anti-inflammatory drugs (NSAIDS; Lee et al. 2004; Towheed et al. 2006), while opioids may be slightly more efficacious than NSAIDS, but with an unfavorable side effect profile (Nüesch et al. 2009).

Naturally occurring canine osteoarthritis deserves special attention as a possible alternative to rodent models of inflammatory pain, suggested to offer better prediction of clinical efficacy in man (Quessy 2010). Osteoarthritis is common in the dog, representing a significant clinical problem (Johnston 1997). Instruments for quantification of pain and disability are well developed, comprising both subjective rating scales for use by owners and veterinarians/researchers and objective methods for weight bearing, gait analysis, and motor activity (Rialland et al. 2012b). Although in many respects the dog would be closer to man in pathophysiology, species differences in efficacy and therapeutic window are likely. A trial cohort is likely to consist of individuals of different sex, age, strain, and body weight. All of these are factors that may interfere with disease progression, pharmacokinetics, and treatment effects. Depending on inclusion criteria, there may be variability in disease expression and comorbidity. The number and location of affected joints may differ, which could lead to interaction between the joint studied and other affected joints when using force plate and similar methods. The diversity in the population would reduce the likelihood of overestimating treatment
effects, but increases the risk of false negatives. From the practical point of view, the cost of conducting a trial, including synthesis of test compound in case of novel drugs, might be limiting, as may the availability of patients. From an ethical perspective, there is both the advantage of using subjects that may benefit from the scientific advances and the drawback that companion animals may be at risk of suffering side effects or sub-optimal treatment. Naturally occurring disease in pet animals can hardly replace rodent models but may be useful in later stage drug discovery programs and in exploratory work when acceptably safe and inexpensive test compounds are available. COX inhibitors and opioids are effective in dog osteoarthritis but efficacy depends on readout and there may be differences in the impression of efficacy between owners and investigators (Moreau et al. 2003; Vasseur et al. 1995). This is not necessarily negative since the assessment instruments, like different readouts in rodent models, would reflect different aspects of pain. Since some instruments incorporate subjective scales to be used by owners or therapists, placebo effects can be evident, however (Malek et al. 2012).

**Inflammatory Pain.** Rodent models of pain due to inflammation and arthritic disease are usually induced by chemical agents injected into a single joint or in the soft tissue of a paw. Depending on the nature and concentration of the induction agent, symptoms may last from several hours to several weeks, with significant differences in pathophysiology and pharmacology (Ängeby Möller et al. 2012; Ferland et al. 2011). Spontaneous and polyarthritic models may be used for pain studies but are more often applied to research on disease mechanisms (Ameye and Young 2006; Bendele et al. 1999). In the dog, surgically induced models as well as naturally occurring osteoarthritis have been used for pharmacological studies (Gregory et al. 2012).

In general, rodent models of paw inflammation and arthritis yield much greater analgesic drug efficacy than would be expected from human data, but both efficacy and potency varies. In plantar inflammation induced by FCA in rats, weight bearing and the response to pressure in a Randall-Selitto test were completely reversed by several COX inhibitors. By contrast, mechanical hypersensitivity measured by progressively increased punctate pressure to the paw was refractory (Huntjens et al. 2009). Several COX inhibitors showed full efficacy at clinically relevant plasma concentrations in a paradigm where gait analysis was used in rats with carrageenan-induced monoarthritis, while the same testing paradigm was much less sensitive when FCA was the induction agent, (Ängeby Möller et al. 2012; Angeby Möller et al. unpublished observations).

In mice, neither analgesic drugs nor prednisolone, when given at doses that completely normalized the running wheel deficit caused by bilateral inflammation, changed mechanical hypersensitivity tested with von Frey filaments after unilateral FCA-induced paw inflammation (Cobos et al. 2012). In this study, however, the efficacy of morphine was unusually high, with an ED50 of less than 0.1 mg/kg. In another study, both rearing and locomotion deficits in carrageenan-induced bilateral paw inflammation was reduced or completely reversed by several COX inhibitors at doses that were devoid of effects on rotarod performance in controls but unfortunately, no control data were presented on spontaneous motor activity.
In contrast, morphine showed some efficacy in the same paradigm at 2 mg/kg, but performance deteriorated at 10 mg/kg, regardless of whether inflammation was present.

Buprenorphine reduced joint tenderness measured as the response to repetitive palpation and partially reversed the running wheel deficit in carrageenan-induced arthritis, whereas a high dose of morphine only reduced tenderness, possibly because the duration of effect was inadequate for the running wheel assay (Krug et al. 2009). In this study, treatment with botulinum toxin type A was effective on both readouts in the FCA model but not after carrageenan.

Neuropathic Pain. Overall, clinical pain of neuropathic origin responds poorly to treatment. In a comprehensive meta-analysis, the combined NNT values for polyneuropathies, the indications most frequently selected for clinical trials, varied from 2 to 3 for opioids and tricyclic antidepressants to 6.4 for gabapentin, while there was a span between 2.5 and 5 for postherpetic neuralgia (Finnerup et al. 2010). There was little evidence for efficacy of drugs other than opioids for pain due to nerve trauma. Furthermore, cancer and HIV chemotherapeutics of different types induce painful peripheral neuropathy by specific pathophysiological mechanisms, but are largely refractory to pharmacological mono-therapy with tricyclic antidepressants and anticonvulsants, including amitriptyline, gabapentin, and pregabalin (Kaley and Deangelis 2009; Phillips et al. 2010).

The more common neuropathic rodent models utilize traumatic injury of a peripheral nerve (Berge 2011, 2013). Consequently, they do not reflect the clinical trial scenario in terms of etiology, but are rather built on explicit or implicit assumption that the results can be generalized. There are, nevertheless, alternative models based on systemic neurotoxic effects of chemotoxic agents, emulating elements of diabetic neuropathy (Obrosova 2009), and neuropathy related to chemotherapy of cancer (Authier et al. 2009) and HIV (Dorsey et al. 2009; Joseph et al. 2004; Wallace et al. 2007b). It is important to keep in mind that different implementations exist and that the use of a particular agent does not guarantee identical pathology across laboratories and species. Although less common, models of postherpetic neuralgia (Fleetwood-Walker et al. 1999) and HIV (Wallace et al. 2007a) have been described.

Reference compounds show variable efficacy and potency in neuropathic pain models but as for other conditions, the data tend to be more optimistic than warranted by clinical reality (Berge 2011). This is not entirely consistent, however. In a review of data obtained in the spinal nerve ligation model, amitriptyline and duloxetine failed to show efficacy at clinically relevant exposures whereas the minimum effective doses for gabapentin and carbamazepine produced peak plasma exposures only 2–3 times higher than the maintenance doses in man (Whiteside et al. 2008). In the rat spared nerve injury model, chronic administration of pregabalin partially reduced hyperalgesia at clinically relevant plasma levels, whereas carbamazepine was ineffective even at doses several times higher than required for anticonvulsive effects (Lau et al. 2013). Whereas the Whiteside study used a conventional evoked readout, the Lau study used burrowing as well. The divergent
data on carbamazepine indicate that either models, readouts, or study designs are differentially sensitive to these drugs and their modes of action.

**Drug Concentration.** Ideally, pharmacological effect should be related to drug concentration in the target tissue, both to allow comparison of efficacy between studies but even more importantly, to establish that a drug is acting at a concentration compatible with its supposed mode of action. But to obtain such data is difficult and beyond the capacity of many laboratories. Plasma level is a reasonable compromise, but it should be appreciated that many compounds, including frequently used reference drugs like morphine, diclofenac, and gabapentin, have a delayed distribution from blood to target organ (Berge 2011; De Gregori et al. 2012; Torres-Lopez et al. 1997). In single administration studies, a drug with a short half-life and delayed distribution may be eliminated before the concentration in the target tissue reaches an effective level. Furthermore, the concentration in more accessible tissues may be sufficient to induce unpredicted effects, perhaps incorrectly interpreted as relevant efficacy or side effect. In repeated administration studies, the possibility for accumulation or induction of metabolism should be considered. The bottom line is that pharmacological data should be interpreted with extra caution when data on tissue or plasma concentration is lacking.

6 Modelings Strategies

The examples presented above are meant to illustrate the great heterogeneity between models and readouts in terms of pharmacological sensitivity. It is hardly possible to identify a paradigm that has general translational value on its own. This is commonly seen as a problem, but may in fact be an asset if used to select models optimal for the purpose of the research. A model may be used for screening on the assumption that the mechanism pursued will provide clinical analgesia if adequately addressed. This is a typical strategy in a drug discovery project at a stage where compounds are synthesized and optimized. This strategy is also useful in projects where clinically proven drugs are subject to reformulation and pro-drug approaches and the effect of the modifications needs to be compared to an existing drug. In these cases, it is important to establish values for potency, ideally indicating the exposure level needed for relevant target engagement in man. The primary requirement, however, is that the data allow drugs with the same mode of action to be compared. This is best served by a model that produces a full range of response, from no effect at a low dose to 100 % effect at the high end of the dose–response range. This is achievable such as for opioids in many acute models and for COX inhibitors in carrageenan-induced monoarthritis (Ängeby Möller et al. 2012; Le Bars et al. 2001). In these examples, the former would underestimate clinical potency, while the latter would provide a realistic estimate. Both models would overestimate efficacy in any clinical population (Fig. 1). The readout for screening should be pharmacodynamically linked to target engagement, but does not have to faithfully reflect a relevant pathophysiological mechanism. For
example, in neurology and psychiatry research, various types of motor behavior driven by drug–target interaction but with no or little connection to the clinical symptomatology have successfully been used to develop effective therapies (Kakkola and Teravainen 1990; Ögren et al. 1990). But the use of substitute endpoints does carry a risk of optimizing for an unwanted or irrelevant effect. A classical example would be respiratory depression, once suggested to be the most predictive modeling readout for opioid analgesia (Beecher 1957).

The use of models for screening can be validated with available ligands to the same molecular target. But when the mode of action is unprecedented, good pharmacological tools may be unavailable. Establishing the ability of the model to detect a relevant drug–target interaction would then be part of the model validation process, necessary for any novel target.

The aim of most studies in behavioral pharmacology is to predict clinical potential of novel drugs or pathophysiological mechanisms. This task is more diverse and challenging than screening, and the recent history is a mix of success and failure (Berge 2011). Correct predictions require integration of information from different sources. The simplest, but perhaps most challenging situation would be a black box project where a compound is active in some model, but with an unknown mode of action. A classic example is the inhibitory effect of gabapentin in neuropathic pain. To estimate the potential of such a compound, a possible approach would be to screen it in a battery of pharmacologically characterized models in order to establish a degree of efficacy on different sensory and behavioral parameters, including side effects, and use the information to formulate testable hypothesis concerning target indication. The test battery should include models representing different types of pathophysiology, functions, and modalities.
In a more typical situation, the molecular target is defined and to some extent validated. Ideally, there would be evidence from human studies to evaluate construct validity for both model and target. The development of analgesic therapies based on inhibition of nerve growth factor (NGF) may serve as an example. Before the publication of clinical data on analgesic efficacy in osteoarthritis (Lane et al. 2010), substantial evidence suggested a role for NGF in human nociception and joint pain. NGF and its receptor TrkA is upregulated in human joint tissue from patients with arthritis (Aloe et al. 1992) and injection of NGF causes hyperalgesia or pain in human experimental models (Dyck et al. 1997). There was also a large body of rodent data (McMahon 1996). A suitable model would incorporate some critical features of the clinical condition, e.g., upregulation of NGF in arthritic tissue and a readout reflecting movement-related pain. FCA-induced monoarthritis with gait analysis in the rat would fulfill these criteria and has been used to demonstrate convincing efficacy of an NGF antibody as well as other compounds acting on the same pathway (Ängeby Möller et al. unpublished observations). In this model, even the cyclooxygenase pathway is activated (Finn and Oerther 2010) but COX inhibitors are less efficacious.

7 Conclusions

The examples discussed in this text allow several conclusions. Pain in the clinic, when studied at a population level, is poorly treated with single drug therapies. This is clearly the case in major human indications and in veterinarian settings. Even in many experimental models, readouts vary in sensitivity to analgesics, and so the challenge is to understand how this reflects the perception and experience of the individual. Although high doses of drugs are frequently used, lack of sensitivity is not a universal problem of animal behavioral models. There are combinations of models and readout with excellent efficacy at clinically relevant drug exposure for some drug classes. However, sensitivity does not guarantee validity. In fact, good sensitivity to compounds like gabapentin or ibuprofen would indicate that the model is tuned toward a certain mechanism and may be useless for general prediction. The search for globally relevant models is probably futile—the focus should be on identifying the best models to address specific questions. Translation and prediction then have to be based on the weighted evidence derived from a variety of sources, behavioral pharmacology being just one, albeit important. Undoubtedly, many projects have been progressed into clinical development with poorly validated targets and on the assumption that the animal models were validated by their sensitivity to current analgesics. This is unfortunate, but there are excellent tools in the toolbox of behavioral pharmacology. With proper investment in the validation processes, there is reason to hope for better success rates in the future.
References


Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10:712. doi:10.1038/nrd3439-c1


Winder CV, Pfeiffer CC, Maison GL (1946) The nociceptive contraction of the cutaneous muscle of the guinea pig as elicited by radiant heat, with observations on the mode of action of morphine. Arch Internat Pharmacodyn Ther 72:329–359
Yeomans DC, Pirec V, Proudfoot HK (1996) Nociceptive responses to high and low rates of noxious cutaneous heating are mediated by different nociceptors in the rat: behavioral evidence. Pain 68:133–140
Yeomans DC, Proudfoot HK (1996) Nociceptive responses to high and low rates of noxious cutaneous heating are mediated by different nociceptors in the rat: electrophysiological evidence. Pain 68:141–150