Health care providers' judgments in chronic pain: the influence of gender and trustworthiness

Gráinne Schäfer[1,2], Kenneth M. Prkachin[3], Kimberley A. Kaseweter[3,4], Amanda C de C Williams[1]

1 Research Department of Clinical, Educational & Health Psychology
University College London
Gower Street, London WC1E 6BT, UK

2 Neonatal Unit, Queen Charlotte's and Chelsea Hospital, Du Cane Road, London W12 0HS, UK

3 Department of Psychology,
University of Northern British Columbia,
3333 University Way, Prince George, British Columbia, Canada, V2N 4Z9

4. Department of Psychology,
University of British Columbia—Okanagan,
3333 University Way, Kelowna, British Columbia, Canada, V1V 1V7

Corresponding author
Amanda C de C Williams
Research Department of Clinical, Educational & Health Psychology
University College London
Gower Street, London WC1E 6BT, UK
Tel: +44 2076791608
e-mail: amanda.williams@ucl.ac.uk

5 tables, 1 figure
5 supplemental files

Abstract

Estimates of patients' pain, and judgments of their pain expression, are affected by characteristics of the observer and of the patient. Here we investigated the impact of high or low trustworthiness, a rapid and automatic decision made about another, and of gender and depression history on judgments made by pain clinicians and by medical students. Judges viewed a video of a patient in pain presented with a brief history and rated his or her pain, and the likelihood that it was being exaggerated, minimized or hidden. Judges also recommended various medical and treatment options.

Contrary to expectations, trustworthiness had no main effect on pain estimates or judgments, but interacted with gender producing pervasive bias. Women, particularly those rated of low trustworthiness, were estimated to have less pain and to be more likely to exaggerate it. Unexpectedly, judgments of exaggeration and pain estimates were independent. Consistent with those judgments, men were more likely to be recommended analgesics, and women to be recommended psychological treatment. Effects of depression history were inconsistent and hard to interpret. Contrary to expectations, clinicians' pain estimates were higher than medical students', and indicated less scepticism. Empathy was unrelated to these judgments. Trustworthiness merits further exploration in healthcare providers' judgments of pain authenticity, and how it interacts with other characteristics of patients. Further, systematic disadvantage to women showing pain is of serious concern in healthcare settings.

Keywords: pain expression, judgment study, vignette study, observer bias

Introduction

Many factors influence the communication of pain experience by verbal and nonverbal behaviours and its interpretation by receivers of the communication, including healthcare staff [18,34,36,42]. A social communication model [28] systematizing these processes refers to observers' knowledge, biases, and beliefs about pain (and about particular human characteristics). It relates these to observers' estimates of the likelihood and severity of pain. It examines the implications of those judgments for decisions on patients' eligibility or suitability for particular treatments [9,53]. Suspicion about authenticity of complaints of pain is common [15,35]. Healthcare staff bring biases to their judgments: they make lower estimates of pain than laypeople [53], more so with longer healthcare experience [34]. Higher levels of trait empathy in healthcare staff are associated with higher estimates of others' pain [25] and a diminished tendency to blame patients for unsuccessful outcomes [54].

Patient characteristics that affect observers' pain estimates include ethnicity, skin color, sex, age, attractiveness, likeability, manner [1,5,14,26], and presence or absence of medical evidence, e.g. [10,15]. In clinical studies, women's pain is underestimated compared to men's [2,11,53]; effects are less consistent in experimental studies [29,47]. Clinical decisions tend to be consistent with this, so women may receive less analgesia [30,31,33,40], and their pain attributed to psychological rather than medical problems [8 but see 49,58]. Evidence of depression may encourage attribution of pain or its impact to psychological causes [24] or inadequate care [17,60], so that depressed patients with chronic pain are more likely to be prescribed opioids than those who are not depressed [31,50,52].

One feature not studied in relation to pain evaluation is trustworthiness. Judgment about others' trustworthiness occurs automatically and rapidly, based on facial traits [57,63]. It influences reactions to others [45] and may be implicated in judgments of patients' authenticity or intentions. The clinical setting of pain assessment provides ample opportunity for judgments of pain to be influenced by concerns over the

4

trustworthiness of patient presentation; it is replete with suspicions about alternative motivations, such as "secondary gain," fraud or the desire to obtain opioids [51]. Consequently, understanding how judgments of patient trustworthiness affect decisions about others' pain is important to pursue.

In the present study, we evaluated the influence of patient trustworthiness and depression history on participants' judgments of men and women with chronic pain. We systematically varied their trustworthiness and depression history in vignettes accompanying videos of real patients [43] of high or low perceived trustworthiness. We compared participants with very different exposure to pain patients: UK medical students in clinical years of study, and pain clinicians in the UK; we also assessed their empathy. Each participant evaluated several patients with accompanying information, and estimated the patient's pain; the likelihood that the patient was exaggerating, minimizing, or hiding pain; and the likelihood of the participant prescribing opioids, non-opioid analgesics, antidepressants, or a pain management program. We evaluated a series of a priori hypotheses concerning these outcomes in relation to patient trustworthiness, gender and depression history, and participant experience and empathy.

5

Method

Participants and setting

Doctors working in UK pain services who were members of the International Association for the Study of Pain were invited to participate in the study via direct email; medical students in their 4th, 5th or 6th (clinical) years of study were invited to participate via advertisements in the weekly medical society newsletter and announcements before lectures. Both invitations contained brief information about the study and the link to the study website. Ethical approval was obtained from the University Ethics Committee (ref. 4714/001), and informed consent obtained from all participants.

Power analysis was informed by a similar study [16] in which participants' estimates of pain showed an effect size of 0.25. GPower 3.1.5 [21], alpha = 5% and desired power = 80%, estimating correlation among repeated measures to be 0.5, gave a required sample size of 30 per participant group.

Design

The study used the online survey platform Qualtrics [44]. Each participant was exposed to twelve different vignettes and corresponding videos concerning patients with chronic pain (CP). A 3x2x2x2 mixed design was used: history of depression (no history of depression, depression with onset before CP, depression onset after CP); trustworthiness (high, low); gender (all of these were within-subject factors); experience level (clinician, medical student) (between-subject factor). The dependent variables were: participants' estimations of pain; estimations of the probability that patients were exaggerating, minimizing or hiding their pain; and treatment choices.

Materials

Twelve videos (six each male and female) were selected from the UNBC-McMaster Shoulder Pain Expression Archive Database [43] based on ratings of high and low trustworthiness of a larger set. The database contains videos showing faces of 129

6

patients with shoulder pain during painful physiotherapeutic maneuvers (for a full description of patient characteristics, tests and videotape characteristics, see Prkachin & Solomon [43] and Lucey et al. [39]). Trustworthiness ratings were obtained using a method similar to Oosterhof and Todorov [41] (see supplemental file 1 for details, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246): a convenience sample of 55 (14 male) trainee clinical psychologists viewed 51 neutral faces of patients, all Caucasian. Participants accessed the images on line (using the survey platform Limesurvey [38]) after providing informed consent and recording their own gender and year of training. Order of stimuli was randomized and participants completed them in that order, with no return to earlier stimuli, rating each for trustworthiness on a scale from 1 (not trustworthy at all) to 9 (extremely trustworthy). The mean rating of trustworthiness was 5.14 (s.d. 0.85, range 1-9), with female patients rated as more trustworthy than males (female mean 5.49, s.d. 0.61; male mean 4.85, s.d. 0.80; t(49) = -3.14, p = 0.003).

The videos of the three rated as lowest trustworthiness (LT) and three rated as highest trustworthiness (HT) male and female patients were selected as the 12 stimuli for the main study. The database contains patients' expression of pain scored using the FACS system [20] adapted for pain [43]; those chosen were scored as having moderate pain (5 through 9 on a scale from 0 to 16) and were balanced across conditions (details are provided in supplemental file 2, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246). Patients' mean age was 51, range 34 – 67. Videos were edited to between five and ten seconds so that only neutral expressions (before the physiotherapy maneuver) or pain expressions (during or immediately after the maneuver) were depicted. Videos were assigned to history of depression conditions at random.

Twelve corresponding vignettes were generated in the form of a brief primary care doctor's letter requesting an opinion on treatment (see supplemental file 3, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246). Each described the duration of pain (one year), how it affected his or her life (e.g. "She finds

it difficult to drive due to pain"), and that the patient had depression onset before developing pain, or depression onset after developing pain, or asthma (as a neutral condition). For patients with depression, the letters added that they were not on medication or receiving psychological therapy for depression.

Measures

Participants were asked to rate "… how likely do you think it is that the person in the video is exaggerating their pain?" using an integer scale from 0 (very unlikely) to 10 (very likely) by moving a slider bar on the screen. This question was repeated with the terms "minimizing (i.e. downplaying)" and "hiding (i.e. concealing)" pain and the same rating scale. For treatment options, participants were asked to "rate the likelihood that you would consider/recommend management strategies listed below in the care of this patient." The scales were anchored with 0 (very unlikely) and 10 (very likely). The strategies were "Prescription of opioid medication", "Prescription of analgesic medication", "Prescription of antidepressant medication as analgesic", "Referral to a pain management programme" and "Referral to a mental health specialist". After the above questions, they were asked to "rate the amount of pain you think the patient in the video experienced," from 0 (no pain) to 10 (extreme pain). (See supplemental file 4, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246).

To ascertain whether participants agreed with earlier trustworthiness ratings provided by trainee clinical psychologists, they were also asked to rate a still image of a neutral expression for each patient, not taking video and vignette information into account, on a scale of trustworthiness from 0 (not trustworthy at all) to 10 (extremely trustworthy).

Trait empathy of participants was assessed with the Interpersonal Reactivity Index (IRI [12,13]), a self-report questionnaire with 28 statements to which possible responses range from 0 (does not describe me well) to 4 (describes me very well). Cronbach's alpha coefficient for internal consistency of the IRI for this sample was 0.79. Finally, participants were asked to guess the study's purpose, to see whether awareness of purpose affected scores. Responses to this question were provided in an open field text box.

8

Procedure

The website opened with an information sheet and consent statement. Participants who proceeded reported their level of experience, gender, and number of years as a pain clinician or year of study as a medical student. Participants viewed each vignette before watching the corresponding video, then answered questions on the probability that the patient was exaggerating, minimizing or hiding pain; the likelihood of recommending each treatment option. Last they were asked to estimate the patient's pain. Participants could re-read the vignette and review the pain video up to the point when they submitted their responses to that vignette and video. This procedure was repeated for each of the 12 vignette-video pairs.

Vignettes and their corresponding videos were shown in a random order to counterbalance order effects. Participants were then asked to complete the IRI, to provide ratings of trustworthiness of patients, and to record their guess of the study purpose. This took approximately 20 minutes to complete. As an incentive, £2 was donated to *Médecins Sans Frontières/Doctors Without Borders* for each complete study response.

Analysis

Data collected online were transferred to SPSS 21, IBM Corp. Data were checked for outliers and nine data points adjusted to less extreme values [23]; all variables were checked for normality. Two with positive skew ('likelihood of prescribing opioids' and 'likelihood of referring to a mental health specialist') could not be normalized by transformation, but since ANOVA has been found to be robust to deviations from normality [23], a mixed ANOVA was used. Levene's test found homogeneity of variance was not violated ($p > 0.01$).

Pain estimations and treatment choices were tested using 3 x 2 x 2 x 2 mixed ANOVAs, with history of depression, trustworthiness, and gender of patient as within-subjects factors and with experience level as a between-subjects factor. Sphericity was assessed

using Mauchly's Test, and Greenhouse-Geisser corrected F-values are reported where the assumption of sphericity was violated ($p < 0.05$). Due to the increased chance of significant findings in a 4-way ANOVA, interactions that were not previously hypothesised were evaluated at a stricter level of significance, at $p < 0.01$ [7]. Interactions and main effects were further analysed with Bonferroni-corrected post hoc tests. Effect sizes were calculated using partial eta squared, the proportion of the variability accounted for by a variable that is not explained by other variables in the model [23].

Results

Thirty-four specialist pain doctors and 29 medical students in their clinical years took part in the study (see Table 1 for participant details). There were more men than women among clinicians, and most had practiced for over 20 years. The majority of medical students (19/29: 66%) were female and in their first clinical year, the fourth year of training. The difference in numbers of males and females between groups was significant (χ^2 (1, 63) = 15.15, p < 0.001).

Reliability of trustworthiness ratings

Clinicians' and medical students' mean ratings of trustworthiness were less extreme than ratings made by the trainee clinical psychologists (high trustworthiness, male stimuli: M = 5.93, s.d. = 1.02, range = 4-9; high trustworthiness, female stimuli: M = 5.01, s.d. = 1.07, range = 3-8; low trustworthiness, male stimuli: M = 4.75, s.d. = 0.96, range = 2-8; low trustworthiness, female stimuli: M = 4.43, s.d. = 1.19, range = 1-8), with high use of the mid-point, 5. Despite this, there was a significant difference in trustworthiness ratings between patients rated high or low (F(1, 62) = 81.72, p < 0.001).

Awareness of the study purpose

Only three participants guessed or inferred that the study concerned mental health problems in patients with chronic pain; two others suggested trustworthiness. Twenty-one participants referred to 'bias', 'first impressions' or 'judging by appearance'; 11 thought that the study was about pain perception; eight referred to empathy, four to malingering and three to decision-making. Nine reported that they were did not know the study's purpose. Participants were therefore divided into those who guessed the study concerned bias (26) and those who did not (37), but there were no significant main effects or interactions in any of the analyses (p > 0.05), indicating that participants' responses were not affected by their beliefs about the purpose of the study.

Effects of gender on pain estimations and judgments

Hypothesis 1 predicted that estimates of pain would be lower for women than for men; estimates of probability of exaggeration higher for women than for men, and estimates of minimizing or hiding pain lower; and that women were less likely to be prescribed opioid or non-opioid analgesics and more likely to be prescribed antidepressants.

There was a substantial main effect of gender on estimates of pain, with nearly a one point difference, indicating that participants estimated males as having more pain than females (Table 2). There was also a significant main effect of gender on estimates of exaggerating pain: participants estimated female patients as more likely to exaggerate their pain than males, and less likely to minimize or hide their pain than males (Table 2). All these were consistent with hypotheses. Exploring these relationships showed a further gender difference: among men, exaggeration, minimizing and hiding ratings were unrelated to participants' pain estimates (Pearson's r, all p > 0.10) whereas, among women, exaggeration ratings were inversely correlated with pain estimates (Pearson's r = -0.42, p = 0.001).

Effects of patient gender on pain management decisions

There was a main effect of gender on prescribing, with male patients more likely to be prescribed opioids and non-opioid analgesics than females (Table 3), consistent with our hypotheses. However, these judgments were unrelated to pain estimates. There was no effect of gender on prescribing antidepressants, contrary to our hypotheses.

Effects of trustworthiness

Hypothesis 2: We had predicted that estimates of pain would be lower, and estimates of probability of exaggeration higher, for patients who were judged untrustworthy; and that they would be less likely to be prescribed opioids.

There was no main effect of trustworthiness on estimates of pain (see Table 4), counter to our hypothesis, but there was a significant main effect of trustworthiness on estimations of exaggerating pain, with LT patients rated more likely to exaggerate pain

12

than HT patients. This indicates that observers make distinct appraisals of pain itself and of how it is expressed.

Contrary to our prediction, there was no main effect of trustworthiness on participants' willingness to prescribe opioids: $F_{(1,61)} = 2.81$, $p = 0.099$, $\eta_p^2 = 0.04$, but willingness overall was low. By contrast, there was a significant main effect of trustworthiness on willingness to prescribe any analgesics, with participants more likely to prescribe analgesics for HT than LT patients: $F_{(1,61)} = 7.53$, $p = 0.008$, $\eta_p^2 = 0.11$.

Interactions between trustworthiness and gender

We predicted (Hypothesis 3) that the variables of trustworthiness and gender would interact so that low trustworthiness would increase the difference between men and women on estimates of pain; on judgments of exaggeration, minimizing and hiding pain; and on treatment recommendations.

There was a significant interaction between gender and trustworthiness for pain estimates and for all judgments of pain expression. Post hoc tests indicated that the effect consisted of downgrading pain estimations and judgments for LT females, with no effects of trustworthiness on pain estimations for male patients (see Figure 1). While this established a difference between HT and LT females, with the former attributed more pain, the main effect for gender remained, with males attributed higher pain than females for both levels of trustworthiness ($t(62) = 3.28$, $p = 0.001$; $t(62) = 7.69$, $p < 0.001$).

For exaggerating, minimizing and hiding pain, post hoc tests indicated that the main effects for trustworthiness were on judgments of female but not male patients, and for low not high trustworthiness patients: LT females were rated as more likely to exaggerate, and less likely to minimize or hide their pain than HT females, whereas males were rated similarly regardless of their trustworthiness. Unlike pain estimation, the likelihood for HT females of exaggerating, minimizing or hiding their pain was rated at the level of HT and LT males ($t(62) = 1.45$, $p > 0.1$; $t(62) = 0.58$, $p > 0.5$; $t(62) = 0.44$, p

> 0.5), while LT females were rated disadvantageously in comparison with LT males (t(62) = 8.30, p < .001, t(62) = 5.91, p < .001; t(62) = 6.93, p < .001).

Effects of history of depression on pain estimations and judgements

We had predicted (Hypothesis 4) that estimates of pain would be lower for patients with depression than for those without; and that, for patients who developed depression before rather than after the onset of chronic pain, estimates of pain would be lower, judgments of exaggerating pain would be higher, and treatment decisions would favour psychological over analgesics. However, we found no consistency in results, which led us to believe that our manipulation had failed; that participants had interpreted vignette wording about depression *predating* pain as depression having begun *and ended* before pain onset, therefore suggesting that the patient was not currently depressed, whereas depression following pain onset appeared to have been understood as enduring. Results are not shown here but are available in supplementary file 5 (available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246).

Effects of level of clinical experience on pain estimations

Hypothesis 5 predicted that experienced pain clinicians would provide lower estimates of pain than medical students.

There was a main effect of level of experience on estimates of pain, indicating that pain clinicians gave patients *higher* pain estimates than medical students: pain clinician mean 5.50 (s.e. 0.18); medical student mean 4.68 (s.e. 0.19); $F_{(1,61)} = 9.85$, p = 0.003, η_p^2 = 0.14. Pain clinicians provided similar estimates of pain across differences in years of experience ($F_{(4, 33)} = 0.28$, p > 0.5). Both findings were contrary to our hypotheses.

There was an interaction between gender of patient and level of experience for pain estimations and all pain judgements: $F_{(1,61)} = 30.85$, p < 0.001, η_p^2 = 0.34. Both clinicians and students were more likely to estimate male patients' pain as greater than female patients' pain, but this effect was more pronounced in students' estimates (Table 5).

14

Empathy

Our last hypothesis was that that empathy scores would be positively correlated with pain estimates, and negatively with the likelihood of exaggerating pain. Clinicians scored lower on total empathy than medical students: clinician mean 56.18 (s.d. 10.50), medical student mean 67.86 (s.d. 9.25), t(61) -4.65, p<0.001. Contrary to our hypothesis, estimated pain was not correlated with empathy scores (Pearson's r = -0.21, p > 0.1) for either male or female patients (r = -0.16, r = -0.21, p >0.05), nor was there any relationship between empathy scores and the estimate of likelihood of exaggerating pain (r = -0.01, p > 0.1).

Discussion

This study investigated the effects of five variables on pain estimates, judgements about pain expression, and treatment decisions, of medical students and pain clinicians, for patients with chronic pain. Three variables concerned the patient: trustworthiness, gender, and depression history; two concerned participants: level of experience, and empathy. Our main hypotheses concerned trustworthiness and depression history, but gender emerged as the most influential factor, consistently affecting pain estimates, judgements of expression and treatment decisions, while trustworthiness tended to affect judgements made about women rather than men. History of depression produced varied effects that were hard to interpret, and they are not further discussed. Recasting results by dependent variable, pain estimation, judgment of pain expression and treatment decisions all favoured males; judgment of pain expression was, as hypothesised, related to trustworthiness. Therefore, gender findings will be discussed first, followed by trustworthiness; interactions are discussed only in relation to hypotheses.

Gender

Hypotheses concerning patient gender effects on pain estimates and judgements were strongly supported, and those relating to pain management decisions were partially supported. Males were consistently estimated to have more pain than females, by nearly 0.9/10 units. Females were judged by medical students, but not by clinicians, to be more likely than males to exaggerate pain and less likely to minimise or hide it. Trustworthiness interacted with gender such that for judgments of exaggeration, minimising, and hiding, LT females were adversely judged, while HT females were rated similarly to HT and LT males. Since patients' facial expressions showed no differences in intensity, it is unlikely that differences can be explained by properties of the stimuli other than gender. Effect sizes were large and arguably clinically significant. Consistent with men's pain being taken more seriously, men were more likely than women to be

16

prescribed opioids by medical students, and to be prescribed analgesics by both clinicians and students.

Many studies have found that pain in females is taken less seriously than in males, and is less adequately treated [33,53]. Gender stereotypes represent men as more tolerant of pain than women and less inclined to report it [4,46]. Students' and clinicians' responses were consistent with these stereotypes, so that females' pain was discounted on the basis of presumed lower tolerance and greater inclination to express, even to exaggerate, their pain. Women with chronic pain report these stereotypes in healthcare settings [61].

In post hoc exploration to understand results better, we found a distinct gender difference in the *relationship* between ratings of expression of pain and estimates of pain itself. In men they were unrelated, even for those rated as likely to be exaggerating pain. For women, the higher the rated likelihood of exaggeration, the lower the pain estimate. It is not clear whether this difference is itself further gender bias – that even when men are thought to exaggerate pain, it remains credible and is not discounted – or whether there is another explanation that requires specific investigation.

Trustworthiness

Contrary to expectations, trustworthiness did not affect pain estimates. High trustworthiness in women somewhat protected them against the adverse judgements made of pain expression (exaggeration) in LT women, but not against discounting of their pain, with estimates significantly lower than men's whether the woman was perceived to be trustworthy or not. High trustworthiness was in both sexes associated with lower ratings of the likelihood of exaggeration, and greater willingness to prescribe non-opioid analgesics. On reflection, it makes sense that trustworthiness would be more closely associated with ratings of authenticity of pain expression, rather than with

17

attributed pain, but it is interesting that these judgments, as with gender, appear to be somewhat distinct.

Experience and empathy

Consistent with previous findings [34,53], we predicted that medical students (inexperienced in chronic pain) would estimate pain higher than experienced clinicians, but found the reverse. The most likely explanations are the clinicians' (members of IASP) particular commitment to pain care [34,35]; or generational differences between medical students and clinicians.

Clinicians scored lower on total empathy than did medical students, but there were fewer women in the clinician sample. We found no correlation between empathy scores and pain estimates, nor between empathy scores and exaggeration ratings. We were unable to explain this divergence from consistent findings in the research literature [25,48,62].

Judgment processes

Findings may provide further insight into judgment processes. Trustworthiness judgments appeared to be made reliably and consistently across raters, and had a clear effect on clinicians' ratings of pain exaggeration but did not directly affect their pain estimates. This may appear surprising since, having judged that someone is exaggerating pain, observers might be expected to adjust their estimates of pain intensity accordingly. This was in fact the case for women, but not for men. These findings suggest substantial independence between pain estimates and judgments of distortion of pain expression, varying with salient third variables such as gender (in this study), or attractiveness [27].

Strengths and limitations

18

The study has several limitations. We could not assess participants' attention to video or vignette material, or recall when answering questions. Information about timing of depression onset may have been misunderstood (see Depression), and patients in videos did not express depressed mood to any noticeable extent: this may have undermined the depression information in some vignettes. Automated recording of time spent viewing each page, even eye tracking, would have been informative. A further ambiguity may have arisen on analgesics: we asked first about recommending opioids, then about recommending analgesics, implying but not specifying non-opioid analgesics, so participants may have interpreted the question differently. Further, some single item scales (see supplemental table 4, available online as Supplemental Digital Content at http://links.lww.com/PAIN/A246) were of unknown reliability, and the study may have been underpowered for certain interactions. Lastly, gender differences between clinicians and students may have affected responses but are by no means clear [37,56]. Strengths of the study include the use of video rather than still faces, and of real patients experiencing pain, with a design allowing investigation of gender-trustworthiness interactions not usually addressed in judgment studies (e.g.54).

Clinical and research implications

In the present study, gender effects were sizeable, overshadowing those of trustworthiness but, although there was no simple effect, trustworthiness played an important role in several interaction effects. For clinical observer judgments of people in pain, trustworthiness operates in a nuanced manner but is nevertheless relevant, in particular in judgments of women. It warrants re-emphasizing that the trustworthiness differences between stimulus patients were not based in fact but in automatic perceptions of facial features, and concordant across the initial selection and the study proper. That such a variable, in principle irrelevant to patients' motivation, self-presentation or appropriateness for treatment influences judgments of deceptiveness or decisions about treatment, independently or in combination with gender, is sobering, and further evidence of the bias that can distort clinical judgment processes.

19

It was somewhat unexpected to find that appraisals of pain displays (exaggerating, minimizing, hiding) were largely independent of pain estimates. This suggests that the process of evaluating pain communications involves multiple components, with opportunity for modification at stages across components. This complexity may offer routes to modify clinical tendencies to discount pain. Further investigation of the separate determinants of components of the pain communication process and their relationship to variables such as gender that interact with both should be a focus of future investigation.

Altogether, many of the experimental differences observed, particularly those involving gender alone and in interaction with trustworthiness, showed substantial effect sizes, such as the effect of trustworthiness on ratings of pain exaggeration among women (1.94 scale points or a 19.4% difference). The influence of depression on ratings of the likelihood of prescribing opioids, by contrast, was a 2.6% difference). The question of what constitutes clinically significant change in pain is unresolved [6], but for 11-point numeric rating scales for pain, one widely used standard for clinical significance is 30% change [22], with a 10 -20% change deemed of "minimal" importance [19]. By these standards none of the changes we found reaches clinical significance, but pain outcomes were not our concern: rather, participants were rendering judgments about others' pain and associated dimensions, for which to our knowledge there is no empirical basis for determining what amount of change would have a recognizable clinical impact. Despite the burgeoning literature documenting the influence of numerous variables on judgements of others' pain, there is a notable absence of studies that measure *clinical* outcomes associated with differences in observers' judgments. Research addressing this issue is desperately needed.

Conflicts of interest: None of the authors has any conflict of interest to declare.

References

[1] Alexander J, Manno M. Underuse of analgesia in very young pediatric patients with isolated painful injuries. Ann Emerg Med 2003;41:617–22.

[2] Anderson KO, Mendoza TR, Valero V, Richman SP, Russell C, Hurley J, … Cleeland CS. Minority cancer patients and their providers: pain management attitudes and practice. Cancer 2000;88:1929–38.

[3] Bair MJ, Robinson RL, Katon W, Kroenke, K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163:2433–45.

[4] Bernardes SF, Keogh E, Lima ML. Bridging the gap between pain and gender research: a selective literature review. Eur J Pain 2008;12:427–40.

[5] Birdwell BG, Herbers JE, Kroenke K. Evaluating chest pain. The patient's presentation style alters the physician's diagnostic approach. Arch Intern Med 1993;153:1991–5.

[6] Birnie KA, McGrath PJ, Chambers CT. Pain 2012;153: When does pain matter? Acknowledging the subjectivity of clinical significance.2311 – 2314.

[7] Bishop DVM. Interpreting unexpected significant findings. doi:doi:10.6084/m9.figshare.1030406. (Accessed 18 May 2014).

[8] Chang AM, Mumma B, Sease KL, Robey JL, Shofer FS, Hollander JE. Gender bias in cardiovascular testing persists after adjustment for presenting characteristics and cardiac risk. Acad Emerg Med 2007;14:599–605.

[9] Chapman KR, Kaatz A, Carnes M. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. J Gen Intern Med 2013;28:1504–10.

[10] Chibnall JT, Tait RC, Ross LR. The effects of medical evidence and pain intensity on medical student judgments of chronic pain patients. J Behav Med 1997;20:257–71.

[11] Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, Pandya KJ. Pain and its treatment in outpatients with metastatic cancer. New Eng J Med 1994;330:592–6.

[12] Davis M. Measuring individual differences in empathy: Evidence for a multidimensional approach. J Person Soc Psychol 1983;44, 113–126.

[13] Davis MH. A multidimensional approach to individual differences in empathy. JSAS Cat Select Doc Psychol 1980;10:85.

[14] De Ruddere L, Goubert L, Prkachin KM, Stevens MAL, Van Ryckeghem DML, Crombez G. When you dislike patients, pain is taken less seriously. Pain 2011;152:2342-7.

[15] De Ruddere L, Goubert L, Stevens M, Williams ACdeC, Crombez G. Discounting pain in the absence of medical evidence is explained by negative evaluation of the patient. Pain 2013;154:669–76.

[16] De Ruddere L, Goubert L, Vervoort T, Prkachin KM, Crombez G. We discount the pain of others when pain has no medical explanation. J Pain 2012;13:1198–205.

[17] Druss BG, Rosenheck RA, Desai MM, Perlin JB. Quality of preventive medical care for patients with mental disorders. Med Care 2002;40:129–36.

[18] Drwecki BB, Moore CF, Ward SE, Prkachin KM. Reducing racial disparities in pain treatment: the role of empathy and perspective-taking. Pain 2011;152:1001–6.

[19] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA,. Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105-121.

[20] Ekman P, Friesen WV. A new pan-cultural facial expression of emotion. Motiv Emotion 1986;10:159–68.

[21] Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behav Res Meth Ins C 1996;28:1–11.

[22] Farrar JT, Young JP, LaMoreaux L, Werth JL. Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149-158.

[23] Field A. Discovering Statistics using IBM SPSS Statistics. London: Sage Publications, 2013.

[24] Graber MA, Bergus G, Dawson JD, Wood GB, Levy BT, Levin I. Effect of a patient's psychiatric history on physicians' estimation of probability of disease. J Gen Intern Med 2000;15:204–6.

[25] Green AD, Tripp DA, Sullivan MJL, Davidson M. The relationship between empathy and estimates of observed pain. Pain Med 2009;10:381–92.

[26] Green CR, Anderson KO, Baker TA, Campbell LC, Decker S, Fillingim RB, Kalauokalani DA, Lasch KE, Myers C, Tait RC, Todd KH, Vallerand AH. The unequal burden of pain: confronting racial and ethnic disparities in pain. Pain Med 2003;4:277–94.

[27] Hadjistavropoulos T, McMurtry B, Craig KD. Beautiful faces in pain: Biases and accuracy in the perception of pain. Psychol Health 1996;11(3):411-20.

[28] Hadjistavropoulos T, Craig KD. A theoretical framework for understanding self-report and observational measures of pain: a communications model. Behav Res Ther 2002;40(5):551-[70.

[29] Hirsh AT, George S, Robinson ME. Pain assessment and treatment disparities: a virtual human technology investigation. Pain 2009;143:106–13.

[30] Hirsh AT, Hollingshead NA, Bair MJ, Matthias MS, Kroenke K. Preferences, experience, and attitudes in the management of chronic pain and depression: a comparison of physicians and medical students. Clin J Pain 2014;30:766–74.

[31] Hirsh AT, Hollingshead NA, Bair MJ, Matthias MS, Wu J, Kroenke K. The influence of patient's sex, race and depression on clinician pain treatment decisions. Eur J Pain 2013;17(10):1569-79.

[32] Hirsh AT, Hollingshead NA, Matthias MS, Bair MJ, Kroenke K. The influence of patient sex, provider sex, and sexist attitudes on pain treatment decisions. J Pain 2014;15:551–9.

[33] Hoffmann DE, Tarzian AJ. The girl who cried pain: a bias against women in the treatment of pain. J Law Med Ethics 2001;29:13–27.

[34] Kappesser J, Williams ACdeC. Pain estimation: asking the right questions. Pain 2010;148:184–7.

[35] Kappesser J, Williams ACdeC, Prkachin KM. Testing two accounts of pain underestimation. Pain 2006;124:109–16.

[36] Kaseweter KA, Drwecki BB, Prkachin KM. Racial differences in pain treatment and empathy in a Canadian sample. Pain Res Manage 2012;17(6):381-384.

[37] Kret ME, de Gelder B. A review on sex differences in processing emotional signals. Neuropsychologia 2012;50:1211-1221.

[38] Limesurvey https://www.limesurvey.org/en/. (Accessed Nov 11 2015).

[39] Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I. Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. In: Ninth IEEE International Conference on Automatic Face & Gesture Recognition (FG 2011). Santa Barbara, USA: IEEE 2011. pp. 21-5.

[40] Michael GE, Sporer KA, Youngblood GM. Women are less likely than men to receive prehospital analgesia for isolated extremity injuries. Am J Emerg Med 2007;25:901–6.

[41] Oosterhof NN, Todorov A. The functional basis of face evaluation. P Natl Acad Sci USA 2008;105:11087–92.

[42] Prkachin KM, Kaseweter KA, Browne ME. Understanding the suffering of others: the sources and consequences of third-person pain. In: Pickering G, & Gibson S, editors. Pain, emotion and cognition: a complex nexus. New York: Springer, 2015. pp. 53 – 72.

[43] Prkachin KM, Solomon PE. The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. Pain 2008;139:267–74.

[44] Qualtrics. Qualtrics. Provo, Utah: Qualtrics. http://www.qualtrics.com. (Accessed 11 Nov 2015).

[45] Rezlescu C, Duchaine B, Olivola CY, Chater N. Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. PLoS One 2012;7:e34293.

[46] Robinson ME, Riley JL, Myers CD, Papas RK, Wise EA, Waxenberg LB, Fillingim RB. Gender role expectations of pain: relationship to sex differences in pain. J Pain 2001;2:251–7.

[47] Robinson ME, Wise EA. Gender bias in the observation of experimental pain. Pain 2003;104:259–64.

[48] Saarela MV, Hlushchuk Y, Williams ACdeC, Schürmann M, Kalso E, Hari R. The compassionate brain: humans detect intensity of pain from another's face. Cereb Cortex 2007;17:230–7.

[49] Safdar B, Heins A, Homel P, Miner J, Neighbor M, DeSandre P, Todd KH. Impact of physician and patient gender on pain management in the emergency department--a multicenter study. Pain Med 2009;10:364–72.

[50] Sullivan MD, Edlund MJ, Steffick D, Unützer J. Regular use of prescribed opioids: association with common psychiatric disorders. Pain 2005;119:95–103.

[51] Sullivan M, Ferrell B. Ethical challenges in the management of chronic nonmalignant pain: negotiating through the cloud of doubt. J Pain 2005;6:2–9.

[52] Sullivan MD, Edlund MJ, Zhang L, Unützer J, Wells KB. Association between mental health disorders, problem drug use, and regular prescription opioid use. Arch Intern Med 2006;166:2087–93.

[53] Tait RC, Chibnall JT, Kalauokalani D. Provider judgments of patients in pain: seeking symptom certainty. Pain Med 2009;10:11–34.

[54] Tait RC, Chibnall JT, Luebbert A, Sutter C. Effect of treatment success and empathy on surgeon attributions for back surgery outcomes. J Behav Med 2005;28:301–12.

[55] Teh CF, Zaslavsky AM, Reynolds CF, Cleary PD. Effect of depression treatment on chronic pain outcomes. Psychosom Med 2010;72:61–7.

[56] Thompson AE, Voyer D. Sex differences in the ability to recognise non-verbal displays of emotion: a meta-analysis. Cognition & Emotion 2014;28(7):1164-1195, DOI: 10.1080/02699931.2013.875889

[57] Todorov A, Baron SG, Oosterhof NN. Evaluating face trustworthiness: a model based approach. Soc Cogn Affect Neur 2008;3:119–27.

[58] Turk DC, Okifuji A. What factors affect physicians' decisions to prescribe opioids for chronic noncancer pain patients? Clin J Pain 1997;13:330–6.

[59] Van 't Wout M, Sanfey AG. Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. Cognition 2008;108:796–803.

[60] Viron MJ, Stern TA. The Impact of Serious Mental Illness on Health and Healthcare. Psychosomatics 2010;51:458–65.

[61] Werner A, Malterud K. It is hard work behaving as a credible patient: encounters between women with chronic pain and their doctors. Soc Sci Med 2003;57:1409–19.

[62] Williams ACdeC, Morris J, Stevens K, Gessler S, Cella M, Baxter J. What influences midwives in estimating labour pain? Eur J Pain 2013;17:86–93.

[63] Willis J, Todorov A. First impressions: making up your mind after a 100-ms exposure to a face. Psychol Sci 2006;17:592–8.

Figure legend

Figure 1. Gender-trustworthiness interaction in judgments of pain expression.

26

Table 1: Information about participants

	n female, male	**%** female, male
Gender:		
Pain clinicians N=34	5, 29	15, 85
Medical students N=29	18, 11	62, 38
Clinicians: years in practice,		
years as pain clinician		
0-5 years	0, 5	0, 15
6-10 years	1, 7	3, 20
11-15 years	6, 5	18, 15
16-20 years	5, 5	14, 15
20+ years	22, 12	65, 35
Medical students: year of study		
4th	18	62
5th	8	28
6th	3	10

Table 2: Means and standard errors for the effect of gender on pain estimates and judgments

	Female	Male	$F(1, 61)$	p	η_p^2
Pain estimates	4.65 *0.15*	5.53 *0.14*	69.61	<.001	.53
Exaggerating pain	3.92 *0.16*	3.21 *0.16*	26.92	<.001	.31
Minimizing pain	3.74 *0.17*	4.41 *0.18*	19.37	<.001	.24
Hiding pain	3.59 *0.18*	4.34 *0.18*	23.87	<.001	.28

Table 3. Means and standard errors for the effect of gender on pain management decisions

	Female	Male	$F(1, 61)$	p	η_p^2
Opioids	**2.16** *0.19*	**2.68** *0.22*	**31.84**	**< .001**	**.34**
Analgesics	**6.12** *0.22*	**6.77** *0.20*	**31.01**	**< .001**	**.34**
Antidepressants	4.26 *0.23*	4.23 *0.25*	0.01	.758	> .01

Table 4: Means and standard errors for the effect of perceived trustworthiness on pain estimate, likelihood of patient exaggerating pain, and on likelihood of prescribing opioids or analgesics
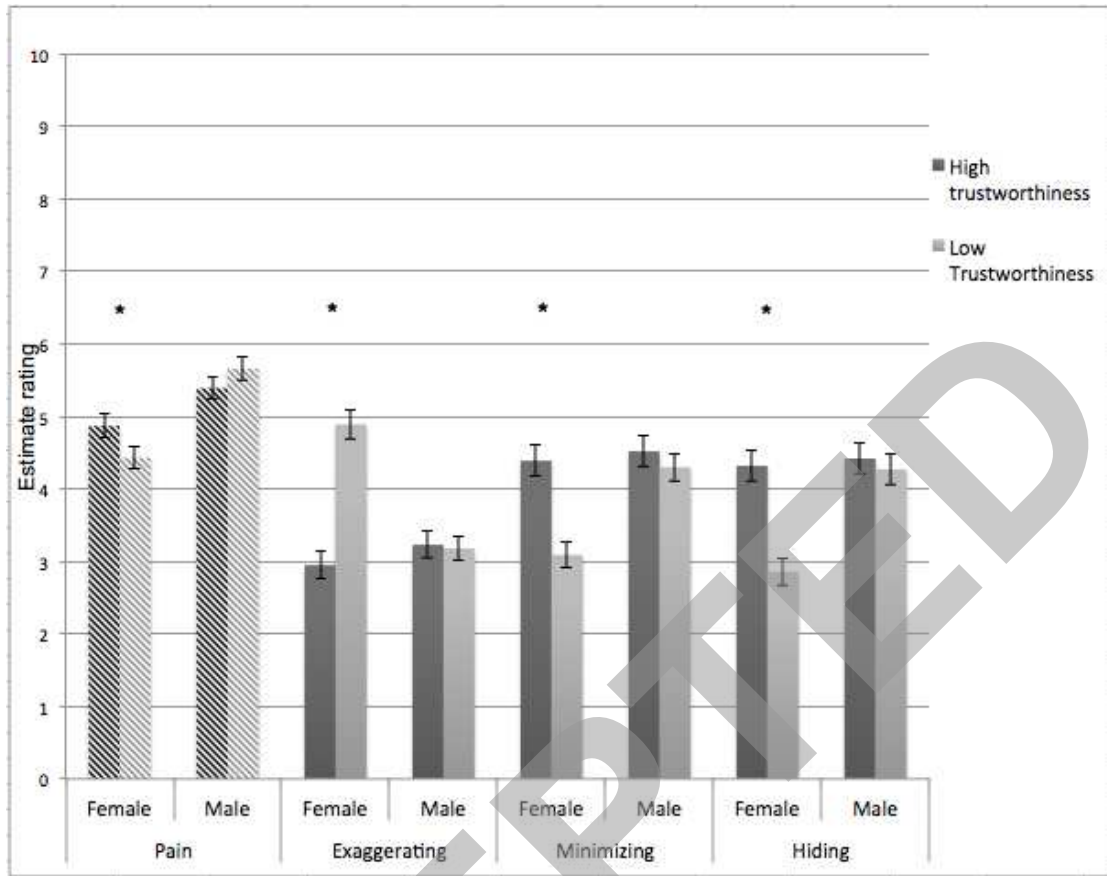
	High trustworthiness	Low trustworthiness	F (1, 61)	p	η_p^2
Pain estimates	5.13 *0.14*	5.04 *0.14*	1.18	.282	.02
Exaggerating pain	**3.09 *0.16***	**4.03 *0.16***	**58.15**	**< .001**	**.49**
Prescribe opioids	2.49 *0.21*	2.34 *0.21*	2.81	.099	.04
Prescribe analgesics	**6.57 *0.19***	**6.32 *0.21***	**7.53**	**.008**	**.11**

Table 5: Means and standard errors for patient gender and participant training level

	Clinicians				Medical students			
	Female patients	Male patients	t(62)	p	Female patients	Male patients	t(62)	p
Pain estimates	**5.35** *0.20*	**5.65** *0.19*	2.06	.044	**3.94** *0.22*	**5.41** *0.20*	9.48	<.001