Individual Differences, Expertise and Outcome Bias in Medical Decision Making

Aron Liaw[1], Matthew B. Welsh[2], Hillary Copp[1] & Benjamin Breyer[3]

([aron.liaw; hillary.copp; benjamin.breyer] @ucsf.edu, matthew.welsh@adelaide.edu.au)

1. Benioff Children's Hospital, UC San Francisco, 1975 Fourth St. San Francisco, CA 94158 USA
2. Australian School of Petroleum, University of Adelaide, North Terrace, Adelaide, SA 5005 Australia
3. Zuckerberg San Francisco General and Trauma Center, 1001 Potrero Ave, San Francisco CA 94110 USA

Abstract

Outcome bias describes the tendency of people to alter their rating of a decision's quality according to whether the outcome is good or bad – despite equivalencies in available information and decision processes – which has the potential to undermine learning about causal structures and diagnostic information in many fields, including medicine. Herein, a sample of 181 doctors and medical students is shown to display outcome bias in medical and non-medical scenarios – with their susceptibility correlating across the domains, r = 0.38. Analyses showed that rational and intuitive decision styles and a medical risk tolerance measure offered little predictive power. Instead, the strongest drivers of bias susceptibility were the Age and professional Level of participants, with more senior personnel showing less outcome bias. We argue that this could reflect improved learning across a doctor's career or result from increasing confidence making them less likely to change their initial judgement of decision quality.

Keywords: medical decision making; outcome bias; individual differences; expertise; decision style.

Introduction

Outcome bias (Baron & Hershey, 1988) describes people's tendency to judge decision quality by outcome rather than the quality of the decision making process. Baron and Hershey demonstrated this across five studies, starting with an experiment where people judged the quality of pairs of decisions about medical treatment that differed only in terms of whether the treatment succeeded or failed. That is, the background information and the decision made remained the same but the outcome differed. The key finding was that almost half of participants rated the decision made in the good outcome scenario as superior to the same decision when a bad outcome occurred (with most of the remainder giving the same rating and a handful rating the good outcome decision as worse). This was despite a within-subjects design, which maximises the chance of participants working out what an experiment is about and remembering their answers to previous scenarios. Participants' own statements also indicated that the outcome *should* not affect ratings of decision quality.

Outcome bias has since been demonstrated in different fields; for example, ethical decision making (Gino, Moore, & Bazerman, 2009), where people's condemnation of ethic breaches is weighted according to the harm done rather than the nature of the ethical breach.

It is distinguished from the similar hindsight bias (Fischhoff & Beyth, 1975) in that outcome bias affects judgements of how good the decision process was, while hindsight affects people's ratings of how likely or predictable the outcome was. (That said, these processes can be linked in situations where, having seen the outcome, hindsight bias leads to the conclusion that the person making the decision should have been able to predict the outcome and thus that their decision making was flawed.)

Of course, judging decisions by their outcomes is natural– particularly given that we often can not access other people's decision-making processes, only the outcomes of their decisions. Thus, we need to *infer* their decision processes (Gino et al., 2009). The fact that people show outcome bias in circumstances when they are specifically made aware of others' decision process and even for their own decisions, however, indicates a problem in decision making - specifically, the overuse of the generally applicable rule that outcomes are linked to decision quality.

Outcome Bias in Medical Decisions

As noted above, the original outcome bias paper used medical scenarios amongst its materials but was conducted on an undergraduate student population. Follow-up work, however, has looked directly at whether medical practitioners are affected by this bias. For example, Caplan, Posner and Cheney (1991) demonstrated anesthesiologists' ratings of the appropriateness of care provided by other medical practitioners was affected by the outcome of that care not just the quality of the decision about treatments.

Similarly, Sacchi and Cherubin (2004) found outcome bias affected doctors' judgements regarding the quality of their own diagnostic decisions and pointed out the difficulties this causes for doctors trying to learn from their own experiences – as good outcomes can artificially inflate confidence while bad luck can deflate it. In either case, background knowledge can be updated incorrectly – inferring causal relationships from random effects.

The problem of learning from experience in the face of outcome and hindsight biases has also been raised for nurses (Jones, 1995) and is key to answering the question of whether these biases can be overcome in order to improve medical decision making.

Experience and Individual Differences

A gap in the above research is in the examination of experience and other individual differences on doctors' outcome bias susceptibility. As noted above, outcome and hindsight biases make learning from experience difficult and it is, therefore, valuable to consider whether experience helps eliminate or exacerbates these biases. No previous studies, however, include doctors' experience as a covariate.

A related question is whether the level of outcome bias shown by doctors on medical and non-medical decisions is similar. If so, this would argue for a general propensity within an individual towards (or away from) outcome bias, which could be linked to personal traits. If not, however, it may be that outcome bias is domain specific – its strength determined by prior experience within a field.

A second line of enquiry is whether there are traits that predict susceptibility to outcome bias. While range truncation in such a highly selected population is likely to prevent measures of intelligence from being useful predictors, it is possible that decision styles (a person's preference for how to make decisions; see, e.g., Hamilton, Shih, & Mohammed, 2016) could affect the level of outcome bias shown. Gino et al (2009) argue exactly this in the context of ethical decision making – that a rational mindset helped to overcome outcome bias. This makes sense particularly for a within-subjects design, where more rational participants could be more likely to notice the pairs of outcome bias scenarios and may feel a greater propensity for ensuring that they are consistent across scenarios.

Another possible covariate is a doctor's tolerance for risk (see, e.g., Grol, Whitfield, De Maeseneer, & Mokkink, 1990). While this may not directly affect outcome bias, it could do so indirectly - by pushing a participant's responses towards the floor or ceiling of a rating scale, thereby potentially preventing outcome bias. For example, if a doctor is particularly risk averse, they could judge a scenario as too risky and thus a bad decision even when it has a good outcome, leaving no space for them to judge it as worse when it occurs with a bad outcome.

Aims and Objectives

The aims of this study are, thus, to: compare doctors' susceptibility to outcome bias on generic and medicine-specific questions; explore whether and how this susceptibility is related to individual traits; and to establish whether outcome bias susceptibility varies across different groups of participants in a meaningful way.

Methodology

Participants

Participants were medical students and practitioners, recruited via Facebook and direct emails to ACGME accredited departments of 100 institutions around the US (universities and large medical groups). In total, 181 completed responses were obtained. Table 1 summarises the participant demographics.

Materials

An online survey was developed in UCSF's Qualtrics, asking participants for demographics and measuring predictor variables and outcome bias as detailed below.

Table 1. Participant demographics

Gender	114 F, 60 M & 7 no-response
Level	66 students, 22 residents, 12 fellows, 56 attendings & 25 no-response
Experience	M = 9.1 years (*SD = 13.2);* 16.9 years (SD = *13.8)* excluding students
Age	21 x '18-25'; 40x '26-35'; 27 x '36-45'; 33 x '46-55'; 59 x '56+'; and 1 x no-response

Demographics. Participants provided their gender, age range, level, years of experience and medical specialty.

Predictor Variables. Two measures with the potential to predict bias susceptibility were included in the survey:

Decision Styles Scale (Hamilton et al., 2016). The DSS is 10-item questionnaire that measures people's preferences as to how they make decisions on separate Rationality and Intuition subscales. Scores on each subscale can range from 5-25 and, in both cases, higher scores reflect greater comfort with decisions being made in that style.

Medical Risk Tolerance Scale (Grol et al., 1990). The MRTS is a 5-item response scale assessing medical practitioners' tolerance for risk in medical decisions. Scores range from 5-25, with lower scores reflecting greater tolerance for risks. Herein, however, we have reversed the scoring such that high values reflect higher risk tolerance.

Outcome Bias Questions. Nine decision scenarios were written for this experiment to enable testing for outcome bias – six describing simple, betting scenarios and three describing medical decisions. (While more scenarios could provide a finer measurement of an individual's degree of outcome bias, this was weighed against limiting the length of the survey in order to maximise responses.)

Betting Scenarios. The basic structure of the betting scenario questions was as follows, with participants responding on a 5-point, 'Very bad' to 'Very good' scale.

Your friend is playing a simple game. He has the choice to not bet on a coin flip, and automatically win $10, or bet on the coin flip and win $15 if it comes up heads, but nothing for tails. He chooses to bet. The coin comes up heads and he wins, gaining $15. In your opinion, how good a decision was this?

In all variant scenarios, the friend ignores the certain $10 and bets on the coin toss. The pay-offs and whether the outcome was good or bad varied as shown in Table 2.

This gives three pairs of questions with the same decision quality (good, neutral or bad, based on simple, economic calculation when compared to the certain, $10 option). Differences between responses to these pairs thus reflect the impact of the outcome of people's responses (outcome bias).

An individual's level of outcome bias is measured as the sum of these differences - that is: (GDGO-GDBO) +

(NDGO-NDBO) + (BDGO-BDBO), yielding scores from -12 to 12 with scores above zero reflecting outcome bias.

Table 2. Decisions scenarios

Code	Outcomes of bet	Decision	Actual outcome
GDGO	$0 or $40	Good	$40
GDBO	$0 or $35	Good	$0
NDGO	$0 or $20	Neutral	$20
NDBO	$0 or $20	Neutral	$0
BDGO	$0 or $15	Bad	$15
BDBO	$0 or $15	Bad	$0

Note: the codes are anagrams. E.g., GDGO = good decision, good outcome. The difference between payoffs in two good decision scenarios was an uncorrected error but analysis suggested it had little impact on results.

Medical Scenarios. Three scenarios were written for this study. In each, a patient opts for a surgery rather than non-surgical management of their condition. Given their length, they are summarized in Table 3 rather than described in full.

Table 3. Medical Scenarios

Patient	Surgery	Risk	Outcome
♀24yr	Pacemaker	Low	Successful surgery
♀42yr	Panniculectomy	Low	Major complications
♂72yr	Hip replacement	High	Successful surgery

As these were written to be realistic for a sample of medical professionals, they are not as easily categorized as the simple, betting scenarios – with the riskiness (and thus the 'goodness' of the decision) depending not on simple probabilities but interpretations of patient history. However, the authors' view (on writing them) was that they corresponded most closely with GDGO, GDBO and BDGO situations, which allows two comparisons: (GDGO-GDBO) as per the above; and also (BDGO-GDBO), which represents the strongest test of outcome bias. As for the simple outcome bias, medical outcome bias was calculated from sum of these two scores, yielding a possible score of -8 to 8, with scores above zero reflecting outcome bias.

Procedure

The Facebook and email invitations included a direct link to the survey, allowing participants to take part without direct contact with the experimenters. The survey started with a standard consent request before proceeding to demographics, then the DSS and MRTS. Finally, the nine outcome bias questions were presented – intermixed to limit direct comparisons between the betting scenarios.

Results

Figure 1 shows participants' mean ratings of decision quality on the Betting scenarios with 95% confidence intervals. This serves as an initial proof of concept – demonstrating that participants recognised differences between good, bad and neutral decisions but were also affected by outcome bias – as scenarios with good outcomes are consistently rated higher than their matched, bad-outcome scenarios. (NB – the smaller number and greater difficulty in designating good versus bad in the medical scenarios meant that a similar figure would not be helpful.)

Figure 1. Mean responses on Betting questions

Descriptive Statistics

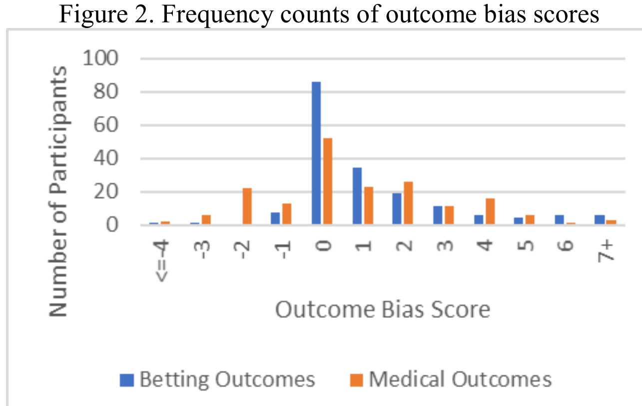
Table 4 summarises descriptive statistics for the individual differences measures and the two measures of outcome bias.

Table 4. Descriptive Statistics

Measure	Range	Mean	SD
Rationality (DSS)	10-25	19.7	3.3
Intuition (DSS)	5-24	11.3	3.4
Risk Tolerance (MRTS)	5-25	12.7	3.7
Outcome Bias (Betting)	-8-11	1.2	2.2
Outcome Bias (Medical)	-4-8	0.8	2.2

NB – the DDS and the MRTS are measured on 5-25 scales. The outcome bias measures are measured on -12 to 12 and -8 to 8 scales for Betting and Medical respectively.

Figure 2. Frequency counts of outcome bias scores



Looking at the table, one can see that both outcome bias measures, although low, are positive – as expected after seeing Figure 1. This impression is strengthened on examination of Figure 2, depicting the distribution of individual's scores. While a number of participants score around zero, there is a right skew, with more participants scoring above zero than below. Overall, 86 of the 181 participants have positive scores reflecting outcome bias for each of the Betting and Medical outcome bias measures (with 86 and 52 scores of zero and 9 and 43 scores below zero, respectively). These are similar proportions to those reported by Baron and Hershey (1988).

Outcome Bias

To test the significance of the above observations, single sample t-tests compared participant results to the expected score of zero if no outcome bias were present. These confirmed that outcome bias scores in both cases are significantly higher than zero, $t(180) = 7.39$ and 4.89, for the Betting and Medical questions respectively, $p < .0001$ in each case. A Pearson correlation was also calculated between the two outcome bias measures, indicating a moderate correlation, $r(179) = 0.38$, $p < .0001$, suggesting a stable tendency for people to show outcome bias (or not) regardless of the scenarios used. This suggests that, despite differences between scenarios, overall outcome bias susceptibility could be calculated in future work.

Individual Differences

Table 5 shows the Pearson correlations between the three individual difference measures and the Betting and Medical outcome bias scores.

Table 5. Pearson correlations between predictor and outcome bias variables

	1	2	3	4	5
1. Rationality	-		*		
2. Intuition	-.02	-	*	*	
3. Risk Tolerance	**-.17**	**-.19**	-		*
4. Betting	-.05	**.16**	-.09	-	***
5. Medical	-.10	.07	**-.19**	**.38**	-

* - sig. at .05 level, 2-tailed; *** - sig. at .001 level, 2-tailed

In Table 5, relationships between the predictor variables and the outcome bias measures are weak but a number are statistically significant. Specifically, Intuition correlates positively with people's Betting outcome bias score, while Risk Tolerance correlates negatively with Medical outcome bias. That is, people with more belief in their own intuitions and less tolerance for risk (or a greater desire to consult with others) seem to have a weak tendency to show more outcome bias (in the Betting scenario).

Overall, however, the results provide little hope for those seeking to use these individual differences to predict levels of outcome bias, with the strongest relationship explaining less than 4% of the variance in outcome bias scores.

Finally, analyses looked at participants' raw responses on the 1-5 ratings across both the Betting ($M = 2.95$, $SD = 0.65$) and Medical ($M = 3.63$, $SD = 0.58$) questions. This established that participants tended to think the medical decisions were better overall but is reassuring in that the majority of results are clear of floor and ceiling in both cases. Comparison of participant's mean ratings with their Risk Tolerance also found no correlation - $r = .04$, $p > .05$ in both cases - undermining the suggestion that risk tolerance might contribute to floor or ceiling effects.

Group Differences

Further analyses were undertaken to determine whether demographic differences between the participants predicted outcome bias or differences in the predictor variables.

Gender

Table 6 shows the data divided by gender.

Table 6. Mean (and *SD*) of measures by gender

	Female (n=114)	Male (n=60)
Rationality	20.0 (*3.37*)	18.6 (*3.63*)
Intuition	10.8 (*3.21*)	10.9 (*3.15*)
Risk Tolerance	12.3 (*3.74*)	13.7 (*4.26*)
Betting	1.54 (*2.39*)	0.58 (*1.62*)
Medical	1.08 (*2.46*)	0.35 (*1.79*)

Looking at Table 3, males and females score similarly on the individual difference traits but show clear differences in terms of the extent to which they show outcome bias, with females showing the bias at higher rates in both the Betting and Medical conditions. Independent samples t-tests were used to assess the significance of these apparent trends. These confirmed that the differences between male and female scores on the DSS Rationality and Intuition measures were not significant. Differences in Risk Tolerance, however, were, $t(172)= 2.3$, $p = .023$ (two-tailed), with males showing higher risk tolerance.

Similarly, the differences in outcomes bias were significant for both the Betting and Medical questions, $t(172) = 2.8$ and 2.0, $p = .006$ and .044, respectively, with males showing less bias in both cases.

Practitioner Level

Table 7 shows the data divided according to the level of the participants (as medical practitioners).

Table 7. Mean (and *SD*) of measures by practitioner level

	Student	Resident	Fellow	Attending
Rationality	20.7	18.7	19.6	19.8
	(*3.0*)	(*3.3*)	(*2.6*)	(*3.3*)
Intuition	11.6	11.6	11.6	10.0
	(*3.4*)	(*2.9*)	(*5.1*)	(*2.6*)
Risk Tol.	11.3	13.8	12.2	14.2
	(*2.9*)	(*3.9*)	(*4.4*)	(*3.9*)
Betting	1.8	1.1	1.1	0.6
	(*3.0*)	(*1.8*)	(*1.1*)	(*1.6*)
Medical	1.2	0.1	1.4	0.3
	(*2.2*)	(*2.7*)	(*2.4*)	(*2.0*)

Note: n = 66, 22, 12 and 56, respectively.

The table shows noticeable differences between the groups on a number of measures. In particular, Attending physicians seem to show less trust in their Intuition, higher risk tolerance and less outcome bias, while Students tend to lie at the opposite extremes on these measures. The Resident group also shows extremely low outcome bias on the Betting scenarios but, given the very small size of this group, the reliability of the result is questionable.

One-Way ANOVAs were conducted in SPSS, comparing the groups' mean performance across all five measures. These confirmed significant differences between groups for: Intuition; Risk Tolerance; and Betting; $F(3, 152) = 2.67$, 7.45 and 2.91, $p = .050$, <.001 and .036, respectively. The other ANOVAs just failed to reach significance $F(3, 152) = 2.58$ and 2.55, $p = .056$ and .058, for Rationality and Medical outcome bias, respectively. Bonferroni post-hoc tests confirmed that significant results were driven by differences between the Attending and Student groups.

Given the effect of practitioner level on results, a χ^2 test was conducted to see whether a relationship between practitioner level and gender was driving the gender effect observed above. This revealed a significant relationship between gender and level, $\chi^2(3) = 10.1$, $p = .014$, with the sample containing more female Students and fewer female Attendings that would be expected based on the overall gender/level breakdown. Thus, multiple regressions (described below) were required to tease these effects apart.

Medical Specialty

Participants listed many specialties – making analysis difficult given space and power constraints. A result that stood out, however, was the difference between surgical and non-surgical specialties. Specifically, despite similar Risk Tolerance scores, surgical specialties (defined as those that make decisions in the operating room on a regular basis, including surgical specialties and anaesthesia) rated the decision to undergo the higher risk surgery (i.e., the bad decision, good outcome Medical scenario) as a worse decision than did non-surgical specialists, Mdiff = -0.42; confirmed as significant by an independent samples t-test, $t(102) = 2.0$, $p = .048$. As a result, the surgeons, overall, did not display outcome bias on this question.

Predicting Outcome Bias

In light of the multiple relationships shown above, linear regressions were run in SPSS using the Forward entry method ($p = .05$ inclusion criterion and $p=0.1$ removal criterion) using Age (converted to a 1-5 scale), Decision Making Training (0 or 1), Experience, Gender (converted to a 0 or 1 scale), Level, Rationality, Intuition and Risk Tolerance on Betting and Medical outcome bias scores. Tables 8 and 9, below, show the models produced for the Betting and Medical outcome bias scores, respectively.

Examination of these tables shows that both produced significant models (albeit with low proportions of variance explained at 7.9% and 11%) with the same predictors for the Betting and Medical versions of outcome bias - Age and Level. Participant's Medical scores were also affected by their Risk Tolerance score. Specifically, the models suggest that participants at higher Levels tend to show *less* outcome bias despite a tendency for older people to show *more*. Greater medical Risk Tolerance also decreased outcome bias, but only for the Medical outcome bias questions.

Table 8. Regression model for Betting scores

Model
Significant Predictors: Level and Age Formula: Betting = 1.74 – 0.96*Level + 0.77*Age $F(2, 149) = 7.48$, $p < .0001$; Adj $R^2 = .079$
Note: regression conducted using forward entry method. Standardised βs = -.297 (Level) and .204 (Age).

Table 9. Regression model for Medical scores

Model
Significant Predictors: Risk Tolerance, Age and Level Formula: Medical = 1.59 – 0.11* Risk + 0.40*Age - 0.329*Level $F(3, 148) = 7.23$, $p < .0001$; Adj $R^2 = .110$
Note: regression conducted using forward entry method. Standardised βs = -.189 (Risk) .253 (Age) and -.198 (Level).

Interestingly, once the effects of Age and Level are partialled out, the gender differences do not reach significance in either model. Neither is previous decision training or either of the DSS measures (Rationality and Intuition) having a significant effect.

Discussion

The results presented above reconfirm the existence of outcome bias in doctors and medical students and add to this knowledge in a variety of ways.

Firstly, the stability of outcome bias across scenarios of different types was established, with participants' outcome bias on Betting and Medical scenarios correlating significantly together. This supports the idea that there could be particular traits that predict the degree of outcome bias an individual will show.

Our analyses, however, failed to support the finding from Gino et al (2009) that a rational mindset decreases outcome bias. This may, however, simply reflect a range truncation effect, with participants' Rationality scores tending towards the higher end of the scale and none scoring below 10 on the 5-25 range. This is, perhaps, unsurprising, given the need for medical students and practitioners to use rational decision making and reflects a common difficulty in finding predictors of biases in highly selected populations.

The fact that Intuition emerged as a significant predictor in correlations with outcome bias does shine some light on the solution to this – the need to find traits that affect decision making but which are less strongly selected for through medical training. Intuition scores, while as low on average as Rationality scores were high may have been less truncated, spanning almost the scale's full range – from 5 to 24. The combination of range truncation and skew in these measures may also explain the somewhat surprising observation that Rationality and Intuition did not correlate in our sample – unlike in the majority of data presented by Hamilton et al (2016) where a negative relationship is seen.

Overall, amongst the potential covariates examined herein only a handful of weak relationships were shown. Overall, the decision styles and Risk Tolerance measures showed little predictive power for outcome bias and, what little they did, disappeared when demographic variables of participant Level and Age were included in regressions. This suggests that participant Level may be affecting both a person's (trust in their own) Intuition and level of outcome bias rather than Intuition directly affecting outcome bias.

Caveats and Future Research

The fact that Level and Age proved the most consistent predictors of outcome bias, combined with the correlations involving Intuition and Risk Tolerance, could indicate that doctors, across the course of their careers, are learning in such a way as to help them overcome outcome bias. Alternately, however, it may suggest that a measure of confidence (see, e.g., Stankov, Kleitman, & Jackson, 2014) could be useful predictor in future work. The idea being that more senior doctors may be performing better because they are more confident and thus less swayed away from their initial rating as to whether something is a good or bad decision by outcomes. Of course, this might apply differentially in situations where they were rating their own decisions rather than those of others, which would need to be tested as well.

This could be regarded as a Bayesian explanation of the expertise effects. Specifically, outcome bias among students could reflect weaker priors which are, therefore, more affected by the new evidence provided by the outcome. More experienced people, by comparison, could have stronger priors as a result of that experience. Such an effect could also shed light on the difference between results for the Betting and Medical scenarios. A possibility we did not consider, for example, is whether people assumed that the coin described in the betting scenarios was 'fair'. We intende for them to do so but did not specifically state it and so participants could have, intuitively, been considering the possibility that the coin was not fair – with the result that their prior beliefs were weaker than in the medical scenarios. If true, an explicit statement or demonstration of the fairness of the coin should reduce outcome bias in these cases.

Another potential trait that could be considered is Need for Cognitive Closure (Webster & Kruglanski, 1994), which measures person's tolerance for ambiguity and/or their need to quickly resolve it – the expectation being that people high in NFCC might show less outcome bias as, having made a decision, they are less likely to revisit it once the outcome becomes known. In either case, however, whether medical personnel show a truncated range on such traits will also need to be tested.

Another interesting possibility raised by the data is that surgical and non-surgical specialists interpret surgical risk differently. This could be directly examined in future work.

A potential concern regarding the data is that the lack of control over online survey data may have resulted in errors or deliberate mistakes in personal data. In particular, it was noted that participant age data was strangely distributed – with more medical students selecting an age of 56+ than seems likely at first glance. This is likely to have eroded the predictive power of Age – by adding noise to the data. Given this, it may be that Age would be a stronger predictor in a future study with greater control over participant inputs. An alternative recruitment strategy could also aid in statistical analysis by ensuring equal numbers of participants in all groups.

Additionally, while prior Decision Making training was not a significant predictor of performance in our data, future research could explore this further by requesting further details on the type of training received and when it was received – given work in other areas showing that the durability of such training can be low over the course of years (see, e.g., Welsh, Bratvold, & Begg, 2005).

Finally, while the findings suggest that outcome bias is reduced by medical expertise, additional work is required to see whether these effects replicate when considering experts in other, non-medical fields. This could shed light on which of the possible explanations described above are most likely.

Conclusions

Doctors and medical students showed outcome bias in medical and non-medical decision scenarios, rating decisions with good outcomes significantly better than those with bad outcomes. The degree of outcome bias shown in these different sets of questions was similar and correlated, indicating a stable susceptibility to outcome bias.

The individual differences traits tested herein showed little predictive power, possibly due to range truncation, but outcome bias decreased with the Age and employment Level of participants. This could represent learning across a doctor's career but could also, we suggest, relate to their overall level of confidence, which is likely to inure them against changing their opinion on what the right decision is in light of new information like outcomes.

Acknowledgments

MBW is supported by ARC LP160101460, which includes support from Santos and Woodside.

References

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569.

Caplan, R. A., Posner, K. L., & Cheney, F. W. (1991). Effect of outcome on physician judgments of appropriateness of care. *Jama, 265*(15), 1957-1960.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things.

Organizational Behavior and Human Performance, 13(1), 1-16.
Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.451.913&rep=rep1&type=pdf(* Downloaded Jan 23rd 2019).
Grol, R., Whitfield, M., De Maeseneer, J., & Mokkink, H. (1990). Attitudes to risk taking in medical decision making among British, Dutch and Belgian general practitioners. *Br J Gen Pract, 40*(333), 134-136.
Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment, 98*(5), 523-535.
Jones, P. R. (1995). Hindsight bias in reflective practice: an empirical investigation. *Journal of Advanced Nursing, 21*(4), 783-788.
Sacchi, S., & Cherubini, P. (2004). The effect of outcome information on doctors' evaluations of their own diagnostic decisions. *Medical education, 38*(10), 1028-1034.
Stankov, L., Kleitman, S., & Jackson, S. A. (2014). Measures of the Trait of Confidence. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of Personality and Social Psychological Constructs* (pp. 158-189): Academic Press.
Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049.
Welsh, M. B., Bratvold, R. B., & Begg, S. H. (2005). SPE 96423 - Cognitive biases in the petroleum industry: impact and remediation. *Proceedings of the Society of Petroleum Engineers 81st Annual Technical Conference and Exhibition*.