A meta-analysis of the effects of measuring theory of planned behaviour constructs on behaviour within prospective studies

Health Psychology Review

http://www.tandfonline.com/doi/abs/10.1080/17437199.2014.927722#.U6jE1fmSx8E

Evon Mankarious[1] & Emily Jane Kothe[2]

[1]School of Psychology, University of Sydney

[2]School of Psychology, Deakin University

Abstract

Background: Measurement reactivity effects, such as the mere measurement effect, have been proposed as a reason for behavioural changes in a number of theory of planned behaviour intervention studies. However, it is unclear whether such changes are the result of the mere measurement effect or of other artefacts of intervention study design.

Aim: To determine the size and direction of changes in health behaviours from baseline to follow-up in prospective studies using the theory of planned behaviour.

Method: Electronic databases were searched for theory of planned behaviour studies which measured health behaviours at two or more time points. Change in behaviour was calculated for all studies.

Results: Sixty-six studies were included. Mean effect sizes across all studies were small and negative (d = -.03). Effect size was moderated by behaviour, behaviour type and follow-up length. Sub-group analyses showed significant decreases in socially undesirable behaviour (d = -.28), binge drinking (d = -.17), risk driving (d = -.20) sugar snack consumption (d = -.43), and sun protective behaviour (d = -.18).

Conclusions: Measurement of intention at baseline resulted in significant decreases in undesirable behaviour. Changes in undesirable behaviours reported in other studies may be the result of the mere measurement effect.

Keywords: Measurement reactivity, the mere measurement effect, theory of planned behaviour, meta-analysis, health behaviour.

Measurement of intention at baseline has been shown to result in behavioural changes in a number of different behavioural domains (for example, Greenwald, Carnot, Beach, & Young, 1987; Morwitz & Fitzsimons, 2004; Sandberg & Conner, 2009) and has also been cited as a possible explanation for changes in behaviour within intervention studies (Godin, Germain, Conner, Delage, & Sheeran, 2013; Hardeman, Kinmonth, Michie, & Sutton, 2011). This phenomenon is part of a wider category of measurement reactivity effects (French & Sutton, 2010) and has been referred to as the self-erasing error of prediction (Sherman, 1980), the self-prophecy effect (Spangenberg & Greenwald, 1999), the question-behaviour effect (Sprott et al., 2006), and the mere measurement effect (Chapman, 2001; Morwitz & Fitzsimons, 2004; Morwitz, Johnson, & Schmittlein, 1993). It will be referred to here as the mere measurement effect.

The direction and magnitude of the mere measurement effect has been subject to debate (Ogden, 2003). While some researchers have found that presenting questions relating to intention can increase behaviour (Godin, Sheeran, Conner, & Germain, 2008) others found the converse (Ayres et al., 2012). Previous meta-analyses conducted within intervention studies have suggested that measurement of an individual's expectation to engage in a behaviour may result in moderate changes in future behaviour (French & Sutton, 2010; Spangenberg & Greenwald, 1999). Importantly, these effects have been shown to remain consistent, even when behaviour is objectively measured (Godin, et al., 2008; Sandberg & Conner, 2009).

While it is possible to identify behavioural changes associated with the mere measurement effect using randomized controlled trials with an intention-to-treat analysis, it is often difficult to distinguish behavioural changes, which may be due to mere measurement, from other factors that may be evident within intervention studies, such as a participant's awareness that they are enrolled in an intervention study, selective recruitment of motivated

individuals into intervention studies, the Hawthorne effect and issues relating to poor or inadequate blinding (J. McCambridge, Butor-Bhavsar, Witton, & Elbourne, 2011). These effects, referred to here as "participation effects", are particularly concerning within the context of mere measurement research since factors such as motivation have been found to impact the extent to which the mere measurement effect is observed (Ayres, et al., 2012).

To date, attempts to systematically investigate the magnitude and direction of mere measurement effects have focused on the evaluation of evidence from intervention studies. It is widely acknowledged that well designed experimental studies based on Solomon four group designs would be the ideal method for evaluating this effect (J. McCambridge, et al., 2011). However, there are currently too few studies that have used this methodology to assess whether or not the mere measurement effect is a source of bias within experimental trials (J. McCambridge, et al., 2011). In addition, since the majority of Solomon four group intervention studies have included participants that were aware that they were involved in an intervention (J. McCambridge, et al., 2011), it is difficult to isolate changes in behaviour due to the measurement of intention at baseline from participation effects. As such, the extent to which observed changes are the result of the mere measurement effect or participation effects is unclear.

One possible method of investigating the mere measurement effect in the absence of these participation effects would be to systematically investigate the mere measurement effect across non-intervention (i.e. prospective) studies. Conceptually, it would be expected that the measurement of intention within intervention and prospective studies would have similar effects behaviour. As such behavioural changes, linked to the mere measurement effect, should be observed in prospective studies where intention is measured at baseline. The analysis of studies which do not implement interventions allows behavioural changes

associated with the mere measurement effect to be isolated from behavioural changes associated with participation effects.

To date, no research has attempted to systematically investigate the existence of the mere measurement effect within prospective studies. This is despite the fact that the body of research using the theory of planned behaviour, in which intention is a routinely measured, offers a large set of studies in which this issue could be assessed. While the conduct of additional Solomon four group trials would be the ideal future direction for research in this area, this review will allow for the identification of those behaviours where evidence of change in behaviour following baseline assessment appears to be strongest through the use of an existing body of evidence.

Consideration of the mere measurement effect in non-intention studies also has the benefit of addressing criticisms of measurement reactivity literature relating to the issue of publication bias (French & Sutton, 2010). Given that null results are less likely to be published than statistically significant findings, it has been argued that the measurement reactivity literature might over-estimate the magnitude of measurement reactivity effects such as the mere measurement effect (French & Sutton, 2010). However, in the case of prospective theory of planned behaviour studies, the decision whether or not to publish those studies was taken separate from any consideration of whether each study supported a mere measurement account of behaviour change. As such, analysis of prospective studies which have measured theory of planned behaviour constructs at baseline may minimise the influence of this potential publication bias by evaluating the potential for the mere measurement effect in studies not specifically designed to test this phenomenon.

Present study

The purpose of the present meta-analysis is to determine the extent to which changes in behaviour consistent with the mere measurement effect (i.e. increases in behaviour

between time 1 and time 2) occur in prospective theory of planned behaviour studies. In particular, the current study considered changes in health behaviour within prospective theory of planned behaviour studies that included a measurement of health behaviour at two or more time points.

Increasing follow-up length has been cited as a way of reducing the effects of the mere measurement effect (Chapman, 2001). Since short lengths of follow-up may increase the probability of detecting behavioural change as a result of the mere measurement effect that dissipates over time (Chandon, Morwitz, & Reinartz, 2004; Chapman, 2001). In order to investigate this suggestion, length of follow-up was considered as a potential moderator of effect size. On this basis of previous research, it was hypothesised that increasing follow-up length would reduce the magnitude of change in behaviour from baseline to follow-up.

Researchers have suggested that the direction and magnitude of behavioural change may be dependent upon a number of different factors, including the nature of the behaviour being measured (French & Sutton, 2010; Sprott, et al., 2006). This may be observed as differences in the magnitude and direction of mere measurement effect related changes in behaviour between studies investigating different behaviours, or may be reflected in differences related to broader characteristics of behaviour. One behavioural characteristic that may be related to the mere measurement effect is that of the social desirability of behaviour. Research conducted on socially desirable health behaviours has typically found that measurement of intention at baseline increases the performance of these behaviours (for example, Godin, et al., 2008; Sandberg & Conner, 2009). However, while some studies investigating socially undesirable behaviour have found increases in socially undesirable behaviours following the measurement of intention (Williams, Block, & Fitzsimons, 2006) others have shown that measurement of intention at baseline decreases socially undesirable behaviours (Todd & Mullan, 2011). To date, no study has systematically investigated the

impact of behaviour on the magnitude of the mere measurement effect. This was investigated in two ways in the present study, type of behaviour (socially desirable vs. socially undesirable) and the specific behaviour measured in each study, were both investigated as potential moderators of measurement activity. This allows for consideration of behaviour type as well as the impact of specific behaviours.

Method

Inclusion and exclusion criteria

Only studies which applied the theory of planned behaviour questionnaire to the prediction of health behaviours were included. Health behaviours were defined as any behaviours which currently impact, or have the potential to impact an individual's health in either a positive or negative way (Gochman, 1988; Green, 1984). Studies must have measured all theory of planned behaviour constructs *and* behaviour (or past behaviour) at baseline and have measured behaviour at one or more subsequent time points using the same, or comparable, scales. Cross-sectional, qualitative, intervention studies were excluded, as were articles published in languages other than English.

Literature search

The literature search and data extraction phases were performed in August 2013. The search strategy used here was modelled on a recent meta-analysis on the theory of planned behaviour in prospective studies conducted by McEachan, Conner, Taylor and Lawton (2011). An electronic database search was conducted using the PsychINFO (OVID), MEDILINE (OVID) and Web of Science (ISI web of knowledge) databases. The search terms used were attitud* and norm* and control and intention*; theory of planned behavi*; planned behavi*; and Ajzen. A full search strategy of the PsychINFO (OVID) database can be viewed in Appendix A. A citation search was conducted using the Web of Science (ISI

web of knowledge) database. All studies conducted on health behaviours that cited Ajzen, 1991; Armitage & Conner 2001; and/or Godin & Kok, 1996 were also included in the search.

Following the search of electronic databases, duplicate studies were removed. Two reviewers conducted a three phase screening process to determine relevant studies. A title screening was first conducted, followed by abstract and full-text screening. At each stage, studies which did not meet the selection criteria were removed. Consultation between the authors resolved issues regarding the eligibility of questionable studies.

Data extraction

Data extraction for included studies was independently completed by two authors (see Appendix B). Mean, standard deviation, and sample size at both baseline and follow-up were extracted from each study to allow for the calculation of effect sizes. Follow-up sample size was not reported in two studies (Chatzisarantis, Frederick, Biddle, Hagger, & Smith, 2007; Chatzisarantis, Hagger, & Smith, 2007). Estimated sample size for these studies was imputed on the basis of average attrition across all included studies. Cohen's *d* was calculated to determine the standardised mean difference for behaviour from baseline to the first follow-up measurement for each study.

Type of behaviour was classified as being either socially desirable or undesirable. Socially desirable behaviour was defined as any behaviour which has positive outcomes for an individual for example, physical activity (Harsanyi, 1986). Socially undesirable behaviour was defined as any behaviour which has the potential to have negative consequences for an individual for example, binge drinking (Harsanyi, 1986).

Meta-analysis procedure

This meta-analysis is reported in accordance to the PRISMA statement (Liberati et al., 2009) and was conducted using the metafor package for R (Viechtbauer, 2010). Cohen's *d* was calculated directly from the means, standard deviations, and sample size reported in each

study to quantify the standardised mean difference score for baseline and follow-up time points.

Analyses were conducted using Hunter and Schmidt's (1994) methods for meta-analysis to correct effect sizes for sampling error variability. Estimates of heterogeneity in effect sizes (the Q-statistic) were presented together with I^2 statistics and their confidence intervals. The Q-statistic provides a formal test of the relative homogeneity of effect sizes (Cochran, 1952). However, this statistic cannot be easily compared across meta-analyses; the I^2 statistic was used an alternative to the Q-statistic to provide a more easily interpretable metric for quantifying homogeneity in effect sizes (Higgins & Thompson, 2002). Using this statistic, values of I^2 greater than 25% represent heterogeneity in the effect size that may indicate the influence of unaccounted for moderator variables (Huedo-Medina, Sánchez-Meca, Marin-Martinez, & Botella, 2006). Length of follow-up and type of behaviour were examined as moderators of the mean effect size. Sub-group analyses were also conducted to determine the mean effect size for behaviour that had been investigated in more than one study. Significant moderators of the effect size were identified through the Q-statistic. Forest plots are presented for each meta-analysis. Risk of publication bias across all behaviours was subjectively assessed through funnel plots and quantitatively assessed through regression tests for funnel plot asymmetry using the metafor package.

Results

Literature search

The database search resulted in a total of 7835 potential studies (after duplicates were removed). After title, abstract and full-text screening a total of 66 studies were identified for inclusion. Figure 1 (Appendix C) shows the complete study selection process.

Study characteristics

The duration from baseline to follow-up behavioural measurement ranged from one week (Cooke, Sniehotta, & Schüz, 2007; Fulham & Mullan, 2011; Gardner, de Bruijn, & Lally, 2012; Hamilton & White, 2008; Kor & Mullan, 2011; Norman, Armitage, & Quigley, 2007; Norman & Conner, 2005; Thomson, White, & Hamilton, 2012) to fifteen years (Plotnikoff, Lubans, Trinh, & Craig, 2012). The distribution of study duration was bimodal, with peaks at two and six months. Baseline sample sizes for the included studies ranged from 64 (Shankar, Conner, & Bodansky, 2007) to 2916 (Elliott & Thomson, 2010) and follow-up sample sizes ranged from 57 (Shankar, et al., 2007) to 1513 (Elliott & Thomson, 2010).

Of the included studies, 47 measured desirable behaviour and 19 measured undesirable behaviours. With respect to behaviours, 29 studies measured physical activity; eight measured binge drinking; seven measured safe sex behaviour; six measured risky driving behaviour; three measured sun protective behaviour; two measured eating well-cooked meat; dental hygiene; and sugar snack consumption. Single studies measured each of dieting; smoking; hygienic food handling; breast self-examination; sexual risk behaviour; and sleep behaviour hygiene. Sub-group analyses were conducted for behaviours that were measured in more than one study (i.e. physical activity, binge drinking, safe sex, risky driving behaviour, sun protective behaviour, eating well-cooked meat, dental hygiene, and sugar snack consumption).

Participant characteristics

The youngest participants ranged from 9-12 years of age (Martin, Oliver, & McCaughtry, 2007) while the sample with the oldest participants ranged from 18-91. Participants were recruited from schools, universities, general practitioner offices and by other means.

Meta-analysis of changes in behaviour across all studies

The average change in behaviour from baseline to follow-up across all studies was small and negative (d = -.03, 95% CI [-.04, .11]; see Figure 4; Appendix C). Effect sizes for individual studies ranged from -.79 to .92 with considerable heterogeneity in effects. Visual inspection of the funnel plot did not appear to show any bias (see Figure 13; Appendix C). Quantitative regression analyses confirmed that there was no significant bias across the included studies (p = .24).

Follow-up analyses revealed significant effects of behaviour type, behaviour, and length of follow-up on the pooled effect size. Length of follow-up was a significant moderator of the pooled effect size (Q_M = 4.49, p = .03), such that the change in behaviour between baseline and follow-up increased as length of follow-up increased.

There was a significant difference in change in behaviour from baseline to follow-up between studies investigating socially desirable (k = 47, d = .07, 95% CI [-.009, .15]; I^2 = 92.14%, Q = 605.52, p < .0001; see Figure 2; Appendix C) and socially undesirable behaviours (k = 19, d = -.28, 95% CI [-.37, -.18]; I^2 = 85.91%, Q = 143.48, p < .0001; see Figure 3; Appendix C). On average, studies of socially desirable behaviour found a small increase in behaviour from baseline to follow-up, whereas studies of socially undesirable behaviour found a small, but statistically significant, decrease in behaviour.

There was a significant effect of behaviour on the pooled effect size (Q_M = 76.35, p <.0001), although there was significant residual heterogeneity remaining even after the effect of behaviour was taken into account (Q_E = 493.46, p < .001). These effects were investigated in planned subgroup analyses reported in the next section.

Physical activity

Average change in physical activity from baseline to follow-up was small (k = 29, d = .08, 95% CI [-.03, .19]; see Figure 5; Appendix C). The heterogeneity in the pooled effect size was significant (I^2 = 91.27%, Q = 340.96, p < .0001).

Binge drinking

Average change in binge drinking from baseline to follow-up was small and negative (k = 8, *d* = -.17, 95% CI [-.23, -.11]; see Figure 6; Appendix C). No significant heterogeneity was found in the pooled effect size (I^2 = 0%, Q = 6.86, *p* = .44).

Safe sex

Average change in safe sex behaviour from baseline to follow-up was small and negative (k = 7, *d* = -.13, 95% CI [-.006, .27]; see Figure 7; Appendix C). There was significant heterogeneity in the pooled effect size (I^2 = 91.12%, Q = 79.29, *p* < .0001).

Risky driving

Average change in risky driving behaviour from baseline to follow-up was small and negative (k = 6, *d* = -.20, 95% CI [-.24, -.15]; see Figure 8; Appendix C). The heterogeneity in the pooled effect size was not significant (I^2 = 0%, Q = 5.24, *p* = .39).

Sun-protective behaviours

The average change in sun protective behaviours from baseline to follow-up were small and negative (k = 3, *d* = -.18, 95% CI [-.35, -.01]; see Figure 9; Appendix C). There was significant heterogeneity on the polled effect size (I^2 = 89.74%, Q = 30.26, *p* <.0001).

Eating well-cooked meat

The average change in eating well-cooked meat was small and negative (k = 2, *d* = -.27, 95% CI [-.63, .09]; see Figure 10; Appendix C). There was significant heterogeneity on the polled effect size (I^2 = 88.23%, Q = 17.02, *p* <.0001).

Dental hygiene

The average change in dental hygiene behaviour was small (k = 2, *d* = .08, 95% CI [-.003, .16]; see Figure 11; Appendix C). The heterogeneity in the pooled effect size was not significant (I^2 = 28.12%, Q = 2.8, *p* = .09).

Sugar snack consumption

The average change in sugar snack consumption was small and negative (k = 2, *d* = -.43, 95% CI [-.86, -.003]; see Figure 12; Appendix C). There was significant heterogeneity on the polled effect size (I^2 = 89.74%, Q = 30.26, *p* <.0001).

Discussion

The purpose of this meta-analysis was to investigate changes in health behaviour from baseline to follow-up within prospective theory of planned behaviour studies. This allowed for the isolation of behavioural changes associated with the mere measurement effect from behavioural changes associated with participation effects.

Whilst changes in overall behaviour were observed, these changes were very small and, therefore, do not appear to support a meaningful change in behaviour associated with the mere measurement effect as reported by other researchers (for example, Godin, Bélanger-Gravel, Amireault, Vohl, & Pérusse, 2011; Godin, et al., 2008). Contrary to previous research showing that the mere measurement effect leads to increases in behaviour following the measurement of intention, the mean effect size within this study was negative. This suggests that on average, the performance of behaviour decreased following the measurement of intention.

There are a number of potential explanations for these effects. First, it should be noted that there was significant heterogeneity in the mean effect size. As a result, the mean effect size is not a reliable estimate of the likelihood of observing mere measurement effect related changes in behaviour within a given study. However, consideration of the effect sizes do show that many studies found either significant decreases in behaviour or very small increases in behaviour that were not statistically significant. As such, these results seem to be broadly inconsistent with a number of mere measurement effect studies which have concluded that mere measurement effect interventions result in a significant increase of behaviour.

It is possible that participation effects, specifically, the effect of poor or inadequate blinding may partially account for results from intervention studies which have found significant mere measurement effects as some studies do not appear to have implemented blinding techniques (Sandberg & Conner, 2009). Thus, it is possible that behavioural changes observed in such studies may have been the result of demand characteristics and not the mere measurement effect.

Motivation, and the selective recruitment of more motivated participants to intervention studies, may also be partially responsible for previous findings. It might be expected that motivation to change behaviour would be less likely to impact the recruitment of participants into a prospective study than into an intervention study. Studies of the impact of motivation on the mere measurement effect have found that motivation is an important determinant of whether the mere measurement effect occurs (Ayres, et al., 2012). Ayres et al. (2012) found significant behavioural changes only occurred when individuals received motivational information as well as receiving questions relating to their intention at baseline. As such, it is possible that, at least for some intervention studies which have attributed behavioural changes to the mere measurement effect, behavioural changes observed in such studies were the result of self-selection bias and not the mere measurement effect. Thus, whilst changes observed may have appeared to be the result of the mere measurement effect, they may have actually been the result of other factors.

However, it must be acknowledged that some high quality studies of the mere measurement effect have sought to address these potential confounders (Godin, et al., 2008). The mere measurement effects observed in these studies cannot be attributed to selective recruitment of motivated participants into intervention studies or to demand characteristics related to poor or inadequate blinding. Instead, more research is clearly needed to determine

the factors that may be responsible for the apparent differences between the prospective studies reviewed here and such mere measurement effect intervention research.

Behavioural changes associated with the mere measurement effect would be expected to dissipate over time (Chandon, et al., 2004; Chapman, 2001). Thus, it has been suggested that shorter follow-up lengths may increase the probability of detecting behavioural changes that are due to the mere measurement effect (Chandon, et al., 2004; Chapman, 2001). It was therefore hypothesised that increasing follow-up length would reduce the magnitude of the mere measurement effect. While follow-up length was a significant moderator of the mean effect size, this effect was not in the expected direction. Results showed that as length of follow-up increased, changes in behaviour associated with the measurement of intention at baseline also increased.

Such results are not consistent with intervention studies which have shown a rapid decay in behaviour as length of follow-up increases (Chandon, et al., 2004) and are not consistent with studies which have shown that changes in behaviour attributed to the mere measurement effect remain consistent between two time points (Chapman, 2001; Godin, et al., 2008). While the results presented here suggest that shorter lengths of time would be more desirable than longer lengths of follow-up, this does not appear to be consistent with theoretical accounts of the mere measurement effect that suggest that changes in behaviour following the measurement of intention are due to an increase in the salience of the target behaviour. Such increases in salience would not be expected to increase over time as appears to be the case in the current analyses. More research is needed to identify why length of follow-up moderated behavioural change in the opposite direction.

A moderator analysis on socially desirable and undesirable behaviour showed a significant difference in the mean effect size between these classifications. This suggests that behavioural change in studies investigating socially desirable behaviours did differ from

studies investigating socially undesirable behaviours. The effect size for change in desirable behaviour was not significant, indicating almost no change in these behaviours from baseline to follow-up. A significant effect for undesirable behaviour was found, with the performance of these behaviours significantly decreasing between baseline and follow-up.

This is important as it suggests that the method of classification not only increases probability of identifying behavioural changes associated with measuring intention at baseline but also indicates that measuring intention at baseline of undesirable behaviours may decrease follow-up behaviour. As the goal of many health psychologists is to increase desirable behaviour or decrease undesirable behaviour, the results of the present study provide evidence that simply measuring intention of undesirable behaviour at baseline could potentially decrease these types of behaviours. Such decreases in undesirable behaviour were found in all behaviours classified here as undesirable (i.e. binge drinking, risky driving, sugar snack consumption) that were analysed as part of planned sub-group analyses.

The decreases in undesirable behaviours broadly, and in binge drinking specifically is consistent with previous research in this domain. The significant decrease in binge drinking behaviour found here is consistent with previous experimental research which has also found a reduction in binge drinking. For example, Todd & Mullan (2011) found that participants who had been presented with theory of planned behaviour based questionnaires consumed less alcohol at follow-up. Similar decreases in binge drinking behaviour have been found by McCambridge and Day (2008) and Lawrence and Ferguson (2012).

To the authors' knowledge, no mere measurement effect intervention study has investigated sugar snack consumption or risky driving. However, the results presented here for these studies provide evidence to suggest that measurement of intention at baseline for these behaviours could significantly decrease performance of these behaviours. Further research is warranted within these behaviours to confirm whether these results are found in

well-controlled experimental studies of the mere measurement effect. Research is also needed in other socially undesirable health behaviours to determine whether the general pattern of reduction in the performance of socially undesirable behaviours is consistent across other health behaviours that have not yet been systematically examined.

Current theories of the mere measurement effect do no provide a satisfactory account of why the mere measurement effect may have differentially affected the performance of undesirable health behaviours. However, there are a number of potential mechanisms for this effect that should be considered in future studies. For example, it is possible that the measurement of intention to engage in undesirable health behaviours induces cognitive dissonance in some individuals (Festinger, 1962). The effects of cognitive dissonance on behaviour have generally been found to be greater for the undesirable behaviours than for desirable behaviours (Freijy & Kothe, 2013). The contrast between the pattern of results between desirable and undesirable behaviours are particularly interesting in light of recent discussions about the tension between the generalizability and utility of theories within health domains (Head & Noar, 2013; Noar & Head, 2013). The results from this review appear to suggest that theories of mere measurement may need to be more behaviour specific in order to properly account for changes in behaviour following the measurement of intention.

Contrary to the overall pattern of results within socially desirable behaviours, sun protective behaviour, classified here as being socially desirable, was found to significantly decrease after the presentation of intention at baseline. To date, no studies have experimentally investigated the mere measurement effect within sun protection behaviours. As such, the cause of this effect is not readily apparent. More work is needed to confirm this apparent effect and determine the factors that may be responsible.

The pattern of results observed for other socially desirable behaviours explored in sub-group analyses (i.e. physical activity, safe sex, eating well-cooked meat, and dental

hygiene) were consistent with mean effect size for all socially desirable behaviours. Namely, changes were very small and not significant. This was surprising given that the mere measurement effect intervention studies on behaviours which could be classified as desirable have found increases in such behaviour when intention was measured at baseline (Godin, et al., 2008).

The finding that measurement of intention at baseline did not appear to result in significant increases in behaviour at follow-up was particularly surprising within the sub-group of studies that considered physical activity. The mean effect sizes for physical activity were very small, indicating almost no change in physical activity from baseline to follow-up. These findings are not consistent with well-designed experimental studies of the mere measurement effect in the context of physical activity. For example, a study by Godin et al. (2011), which investigated the mere measurement effect in a sample of overweight and obese adults, found that individuals who completed a theory of planned behaviour questionnaire relating to physical activity engaged in more physical activity at three months than individuals who completed a similar questionnaire about fruit and vegetable consumption. Participants in that study were blind to the intent of the study and had been recruited to a longitudinal study of "genetic susceptibility to obesity" and thus were not aware that their participation in the trial was an "intervention". This design means that the mere measurement effect cannot be accounted for by participation effects relating to selective recruitment or inadequate blinding.

Previous physical activity studies have cited the mere measurement effect as a possible explanation of changes in behaviour within intervention studies (Hardeman, Kinmonth, Michie, Sutton, & on behalf of Proactive Project Team, 2009). While more work is needed to reconcile the findings presented here with work by Godin et al. (2011), the

results from the current study do not support the hypothesis that changes in physical activity can be attributed to the mere measurement effect.

Strengths, limitations and future direction

This study was the first to systematically investigate the existence of the mere measurement effect within non-intervention studies. The analysis included studies that have sought to predict behaviour on the basis of measures of intention using the theory of planned behaviour. While the theory of planned behaviour has been the focus of substantial debate within this journal within the last year (Ajzen, 2014; Sniehotta, Presseau, & Araújo-Soares, 2013), the use of studies that had used this theory is advantageous as it makes use of a large body of existing literature that has applied similar measures of intention and allows researchers to separate changes in behaviour believed to be the result of the mere measurement effect from participation effects that may be evident within intervention studies.

The current analysis adds to the current knowledge surrounding the mere measurement effect in a number of important ways. Specifically, the current study addressed the possible criticisms of the mere measurement effect literature pertaining to publication bias (French & Sutton, 2010) by investigating prospective theory of planned behaviour studies. In prospective theory of planned behaviour studies, the decision not to publish is independent of whether a study supported a mere measurement effect account of behaviour change

However, a key inclusion criterion for this analysis was that studies must have reported mean behaviour at baseline and follow-up. In most cases, studies that did not meet this criterion were likely to have measured behaviour at just one time point (follow-up). Studies which did measure behaviour at both baseline and follow-up, but omitted baseline behaviour from the published research were therefore excluded from this analysis. Studies were also omitted due to the inconsistency of behavioural reporting. The current analysis found sixty-six prospective studies which used the theory of planned behaviour to measure

intention at baseline, which is a significantly larger sample of studies than could be amassed by reviewing studies that had experimentally tested measurement. However, a number of studies which could have otherwise been included were not because behavioural measures were not comparable or because baseline data was not available.

A number of behaviours (dieting, safe driving, smoking, hygienic food handling, breast self-examination, sexual risk, and hygienic sleep) were observed once, thus, sub-group analyses were not conducted on these behaviours and mean effect sizes for each of these behaviours is unavailable. However, re-analysis of these behaviours once more studies become available may be desirable in order to quantify these effects. Researchers interested in these areas may wish to consider methods through which to increase the number of studies included in future meta-analyses. In particular, researchers may wish to consider the inclusion of non-intervention studies which use other theories where intention is also key component, for example the health action process approach (Schwarzer, 1992), and the prototype-willingness model (Gibbons, Gerrard, Blanton, & Russell, 1998). Such theories were not included in the current analysis because, to date, much of the mere measurement effect literature has used the theory of planned behaviour and because the measurement of intention in such behaviours often differs from measures typically used within the theory of planned behaviour literature. It should be noted that due to differences in the way intention is measured between studies using different theories the addition of non-theory of planned behaviour studies, may serve to further increase the heterogeneity in effect sizes. Researchers who wish to investigate this issue further should consider the balance between the increase in sample size that might be obtained by broadening inclusion criteria and the corresponding increases in heterogeneity attributable to methodological differences.

An important limitation of the current analysis was that it was assumed that the mere measurement effect would manifest itself as a significant change in the performance of

behaviour within individuals exposed to measures of intention. However, it may be the case that significant differences in behaviour observed in experimental studies are a result of significant decreases in the performance of behaviour among the control group (i.e. the mere measurement effect effects may arrest a decline in the performance of the target behaviour among intervention participants). Unfortunately, the current methods of the current review would not be able to detect such an effect. This should be taken into account when interpreting the effect sizes obtained within this analysis.

Given high rates of attrition that are often observed within prospective studies of the theory of planned behaviour, the use of intention-to-treat analysis would be the optimal method of accounting for the mere measurement effect related changes in behaviour. While some intervention studies has used this method of analysis (Godin, et al., 2013), and the vast majority of researchers within this field have used per-protocol methods of statistical analysis. Thus, the necessary data was not available to allow for analysis of the mere measurement effect using intention-to-treat approach.

The ideal method for detecting the presence of such effects would be by means of a randomized controlled trial (experimental design) where an intention-to-treat analysis is adopted or with experimental studies based on the Solomon four group design. However, a major criticism of such designs (particularly Solomon four group designs) is that they may be resource intensive and expensive to conduct (J. McCambridge, et al., 2011). By identifying those behaviours (or sub-groups of behaviours) where changes in behaviour following presentation baseline measures of intention appear most likely to occur, this review can help to guide more efficient use of resources within future research within this domain.

The results of the present analysis are important as it will allow researchers to implement consistent strategies to control for participation effects and therefore determine the extent to which behaviour changes as a result of the mere measurement effect and due to

participation effects. In doing so there is also a need for future researchers to properly account for past behaviour in order to assess the role factors such as habituation have on performing future behaviours. Through understanding the impact of such factors, researchers will be able to better control for them. Additionally, researchers need to implement appropriate and valid, behavioural measures in order to more effectively assess the mere measurement effect.

The present analysis was the first to systematically investigate the existence of the mere measurement effect within prospective studies. Results presented here suggest that, for behaviours classified as undesirable, measurement of intention at baseline could potentially decrease follow-up behaviour. All sub-group analyses which investigated undesirable behaviours (i.e. binge drinking, risky driving behaviour and sugar snack consumption) showed significant decreases in behaviour. Surprisingly, significant decreases in sun protective behaviour were also found after measurement of intention at baseline. These findings suggest that changes in undesirable behaviours reported in other studies may be the result of the mere measurement effect.

References

Ajzen, I. (2014). The theory of planned behaviour is alive and well, and not ready to retire: a commentary on Sniehotta, Presseau, and Araújo-Soares. *Health Psychology Review*, 1-7. doi: 10.1080/17437199.2014.883474

Ayres, K., Conner, M., Prestwich, A., Hurling, R., Cobain, M., Lawton, R., & O'Connor, D. B. (2012). Exploring the question-behaviour effect: Randomized controlled trial of motivational and question-behaviour interventions. *British Journal of Health Psychology, 18*(1), 31-44.

Chandon, P., Morwitz, V. G., & Reinartz, W. J. (2004). The short- and long-term effects of measuring intent to repurchase. *Journal of Consumer Research, 31*(3), 566-572.

Chapman, K. J. (2001). Measuring intent: There's nothing "mere" about mere measurement effects. *Psychology and Marketing, 18*(8), 811-841.

Chatzisarantis, N. L. D., Frederick, C., Biddle, S. J. H., Hagger, M. S., & Smith, B. (2007). Influences of volitional and forced intentions on physical activity and effort within the theory of planned behaviour. *Journal of Sports Sciences, 25*(6), 699-709.

Chatzisarantis, N. L. D., Hagger, M. S., & Smith, B. (2007). Influences of perceived autonomy support on physical activity within the theory of planned behavior. *Special Issue: Familiarity Impacts Person Perception, 37*(5), 934-954.

Cochran, W. G. (1952). The $\chi 2$ test of goodness of fit. *The Annals of Mathematical Statistics*, 315-345.

Cooke, R., Sniehotta, F., & Schüz, B. (2007). Predicting binge-drinking behaviour using an extended TPB: Examining the impact of anticipated regret and descriptive norms. *Alcohol and Alcoholism, 42*(2), 84-91.

Elliott, M. A., & Thomson, J. A. (2010). The social cognitive determinants of offending drivers' speeding behaviour. *Accident Analysis and Prevention, 42*(6), 1595-1605.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2): Stanford university press.

Freijy, T., & Kothe, E. J. (2013). Dissonance-based interventions for health behaviour change: A systematic review. *British journal of health psychology, 18*(2), 310-337.

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology, 15*(3), 453-468.

Fulham, E., & Mullan, B. (2011). Hygienic food handling behaviors: attempting to bridge the intention-behavior gap using aspects from temporal self-regulation theory. *Journal Of Food Protection, 74*(6), 925-932. doi: 10.4315/0362-028X.JFP-10-558

Gardner, B., de Bruijn, G. J., & Lally, P. (2012). Habit, identity, and repetitive action: A prospective study of binge-drinking in UK students. *British Journal of Health Psychology, 17,* 565-581. doi: 10.1111/j.2044-8287.2011.02056.x

Gibbons, F. X., Gerrard, M., Blanton, H., & Russell, D. W. (1998). Reasoned action and social reaction: Willingness and intention as independent predictors of health risk. *Journal of Personality and Social Psychology, 74*(5), 1164-1180.

Gochman, D. S. (1988). *Health behavior: Emerging research perspectives*: Springer.

Godin, G., Bélanger-Gravel, A., Amireault, S., Vohl, M. C., & Pérusse, L. (2011). The effect of mere-measurement of cognitions on physical activity behavior: A randomized controlled trial among overweight and obese individuals. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 2-6.

Godin, G., Germain, M., Conner, M., Delage, G., & Sheeran, P. (2013). Promoting the Return of Lapsed Blood Donors: A Seven-Arm Randomized Controlled Trial of the Question–Behavior Effect.

Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: Mere measurement effects on frequency of blood donation. *Health Psychology, 27*(2), 179-184. doi: 10.1037/0278-6133.27.2.179

Green, L. W. (1984). Modifying and developing health behavior. *Annual Review of Public Health, 5*(1), 215-236.

Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting behavior by asking people if they expect to vote. *Journal of Applied Psychology, 72*(2), 315-318.

Hamilton, K., & White, K. M. (2008). Extending the theory of planned behavior: The role of self and social influences in predicting adolescent regular moderate-to- vigorous physical activity. *Journal of Sport & Exercise Psychology, 30*(1), 56-74.

Hardeman, W., Kinmonth, A. L., Michie, S., & Sutton, S. (2011). Theory of planned behaviour cognitions do not predict self-reported or objective physical activity levels or change in the ProActive trial. *British Journal of Health Psychology, 16*(1), 135-150.

Hardeman, W., Kinmonth, A. L., Michie, S., Sutton, S., & on behalf of Proactive Project Team. (2009). Impact of a physical activity intervention program on cognitive predictors of behaviour among adults at risk of Type 2 diabetes (ProActive randomised controlled trial). *International Journal of Behavioural Nutrition and Physical Activity, 6*(16), 1-10.

Harsanyi, J. C. (1986). *Rational behavior and bargaining equilibrium in games and social situations*: Cambridge University Press.

Head, K. J., & Noar, S. M. (2013). Facilitating progress in health behaviour theory development and modification: the reasoned action approach as a case study. *Health Psychology Review, 8*(1), 34-52. doi: 10.1080/17437199.2013.778165

Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological Methods, 11*(2), 193.

Kor, K., & Mullan, B. A. (2011). Sleep hygiene behaviours: An application of the theory of planned behaviour and the investigation of perceived autonomy support, past behaviour and response inhibition. *Psychology & Health, 26*(9), 1208-1224. doi: http://dx.doi.org/10.1080/08870446.2010.551210

Lawrence, C., & Ferguson, E. (2012). The role of context stability and behavioural stability in the mere measurement effect: An examination across six behaviours. [Article]. *Journal of Health Psychology, 17*(7), 1041-1052. doi: 10.1177/1359105311433346

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7), e1000100.

Martin, J. J., Oliver, K., & McCaughtry, N. (2007). The theory of planned behavior: Predicting physical activity in Mexican American children. *Journal of Sport & Exercise Psychology, 29*(2), 225-232.

McCambridge, J., Butor-Bhavsar, K., Witton, J., & Elbourne, D. (2011). Can research assessments themselves cause bias in behaviour change trials? A systematic review of evidence from Solomon 4-group studies. *PLoS ONE, 6*(10), e25223.

McCambridge, J., & Day, M. (2008). Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. *Addiction, 103*(2), 241-248.

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: A meta-analysis. *Health Psychology Review, 5*(2), 97-144.

Morwitz, V. G., & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology, 14*(1-2), 64-74.

Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does measuring intent change behavior? *Journal of Consumer Research, 20*(1), 46-61.

Noar, S. M., & Head, K. J. (2013). Mind the gap: bringing our theories in line with the empirical data – a response to commentaries. *Health Psychology Review, 8*(1), 65-69. doi: 10.1080/17437199.2013.855593

Norman, P., Armitage, C. J., & Quigley, C. (2007). The theory of planned behavior and binge drinking: assessing the impact of binge drinker prototypes. *Addictive Behaviors, 32*(9), 1753-1768.

Norman, P., & Conner, M. (2005). The theory of planned behavior and exercise: Evidence for the mediating and moderating roles of planning on intention-behavior relationships. *Journal of Sport & Exercise Psychology, 27*(4), 488-504.

Ogden, J. (2003). Some problems with social cognition models: a pragmatic and conceptual analysis. *Health Psychology, 22*(4), 424.

Plotnikoff, R. C., Lubans, D. R., Trinh, L., & Craig, C. L. (2012). A 15-year longitudinal test of the theory of planned behaviour to predict physical activity in a randomized national sample of Canadian adults. *Psychology of Sport and Exercise, 13*(5), 521-527. doi: http://dx.doi.org/10.1016/j.psychsport.2012.02.005

Sandberg, T., & Conner, M. (2009). A mere measurement effect for anticipated regret: Impacts on cervical screening attendance. *British Journal of Social Psychology, 48*(2), 221-236.

Schwarzer, R. (1992). Self efficacy in the adoption and maintenance of health behaviours:Theoretical approaches and a new model In R. Schwarzer (Ed.), *Self Efficacy: Thought Control of Action*. Washington: Hemisphere.

Shankar, A., Conner, M., & Bodansky, H. J. (2007). Can the theory of planned behaviour predict maintenance of a frequently repeated behaviour? *Psychology Health & Medicine, 12*(2), 213-224.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology, 39*(2), 211-221.

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2013). Time to retire the theory of planned behaviour. *Health Psychology Review, 8*(1), 1-7. doi: 10.1080/17437199.2013.869710

Spangenberg, E. R., & Greenwald, A. G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology, 8*(1), 61-89.

Sprott, D. E., Spangenberg, E. R., Block, L. G., Fitzsimons, G. J., Morwitz, V. G., & Williams, P. (2006). The question–behavior effect: What we know and where we go from here. *Social Influence, 1*(2), 128-137.

Thomson, C. E., White, K. M., & Hamilton, K. (2012). Investigating mothers' decisions about their child's sun-protective behaviour using the theory of planned behaviour. *Journal of Health Psychology, 17*(7), 1001-1010.

Todd, J., & Mullan, B. (2011). Using the theory of planned behaviour and prototype willingness model to target binge drinking in female undergraduate university students. *Addictive Behaviors, 36*(10), 980-986.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Williams, P., Block, L. G., & Fitzsimons, G. J. (2006). Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. *Social Influence, 1*(2), 117-127.