EHPS

volume 20 issue 3
August | 2018

www.ehps.net/ehp

The European Health Psychologist

Bulletin of the European Health Psychology Society

editorial

Teresa Corbett
University of Southampton
Floor M. Kroese
Utrecht University

Health Psychology is a broad topic area that focuses on how psychological, behavioral, and cultural factors influences physical health and illness. There are many questions to be asked about a wide range of issues, and a variety of ways in which to answer them. With such complexity, how are we sure that we are measuring the correct construct or using methodology that is best for the question we have?

As scientists, we are constantly challenged to learn more about methodologies and novel approaches to conducting research. With so much to learn, it is easy to become overwhelmed and be tempted to opt-out of engaging with novel- and often confusing descriptions of- complex methodological approaches. However, as Robert Pirsig famously wrote "*The real purpose of the scientific method is to make sure nature hasn't misled you into thinking you know something you actually don't know.*" Using inappropriate research methods can lead to erroneous or incomplete answers. It is crucial that we endeavour to keep up-to-date with methodological advances, so that the field can continue to develop.

In this special of the European Health Psychologist, we sought to introduce you to some of the exciting and useful methods used in Health Psychology research today.

In the first article, with a focus on qualitative approaches, Morrison et al describe the Person-Based Approach to planning, optimising, evaluating and implementing behavioural health interventions. In this article, the group shares an insight into their successful methods for intervention development, as the University of Southampton's Centre for Clinical and Community Applications of Health Psychology celebrates a decade of the LifeGuide research programme. The group has spent the last 10 years developing numerous interventions that have proved consistently engaging and effective.

Noone et al have contributed a primer on how to use Network Meta-analysis in Health Psychology, introducing the key conceptual issues regarding NMA and a step-by-step tutorial on how to conduct a NMA.

Peters and Crutzen have written an introduction and tutorial on the use of CIBER (confidence interval based estimation of relevance) to help researchers to ensure that they are targeting the most important determinants of behaviour in their studies.

And finally, Gillebaart describes current examples of combining explicit and implicit measures when measuring health dilemmas. The author also explains some of the potential and pitfalls of applying implicit measures to health dilemmas.

These articles aim to introduce you and familiarise you with some of the current methodological challenges and opportunities in Health Psychology. We thank the authors for their contributions and hope that the readers find this issue useful, challenging, engaging, and above all, inspiring.

Teresa Corbett, PhD
Research Fellow- National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care (NIHR CLAHRC), University of Southampton
T.K.Corbett@soton.ac.uk

Floor M. Kroese
Assistant professor, Utrecht University, Faculty of Social Sciences, Department of Social, Health and Organisational Psychology
F.M.Kroese@uu.nl

original article

The Person-Based Approach to planning, optimising, evaluating and implementing behavioural health interventions

Dr Leanne Morrison
University of Southampton
Dr Ingrid Muller
University of Southampton
Prof Lucy Yardley
University of Southampton & University of Bristol
Katherine Bradbury
University of Southampton

This year the digital health research team at University of Southampton's Centre for Clinical and Community Applications of Health Psychology is celebrating ten years of the LifeGuide research programme (www.lifeguideonline.org). This research programme was initiated by developing the unique LifeGuide software, which has enabled researchers to create, modify and adapt digital interventions quickly and efficiently, without needing input from programmers. Over the course of a decade of developing numerous interventions that have proved consistently engaging and effective (e.g. Little et al., 2013; Little et al., 2016; Little et al., 2015), we have come to realise that the most important output from this research for the wider research community is not the LifeGuide software (which will soon be superseded by newer technology) but our successful methods for intervention development. We refer to these methods as the 'Person-Based Approach' (PBA, Yardley, Morrison, Bradbury, & Muller, 2015) to intervention development, which we see as an essential complement to theory- and evidence-based approaches.

The Person-Based Approach adapts methods from user-centred design, using in-depth qualitative research (informed by behavioural theory and analysis) to understand the behavioural aspects of user engagement with interventions – both digital and non-digital (see Figure 1). It is an iterative process of collecting data to obtain a deep understanding of user views, context and experiences of the intervention and using this understanding to design, adapt and optimise the intervention to ensure it is maximally meaningful, feasible and engaging for all users. As the Person-Based Approach has evolved we have published a series of papers describing how to apply it; the following sections provide an introduction to the approach.

Intervention planning

The PBA draws on mixed methods research of users' views and experiences to inform the design and planning of an intervention, to ensure that it is engaging and persuasive. Qualitative research can provide rich data on the contextual factors that may influence target users' engagement with the intervention or the behaviour change process (e.g. what are their lives like? What do they value? What are their prior experiences of engaging with the behaviour? What concerns do they have?). Published qualitative and mixed methods research can be scoped and if appropriate, a systematic synthesis can identify the key barriers, facilitators and contextual issues relevant to the target behaviours (e.g. Corbett et al. 2018). If the existing literature is limited in scope or quality, primary qualitative research with target users is conducted.

Insights from these analyses are then used to formulate guiding principles for intervention development. Guiding principles specify the core

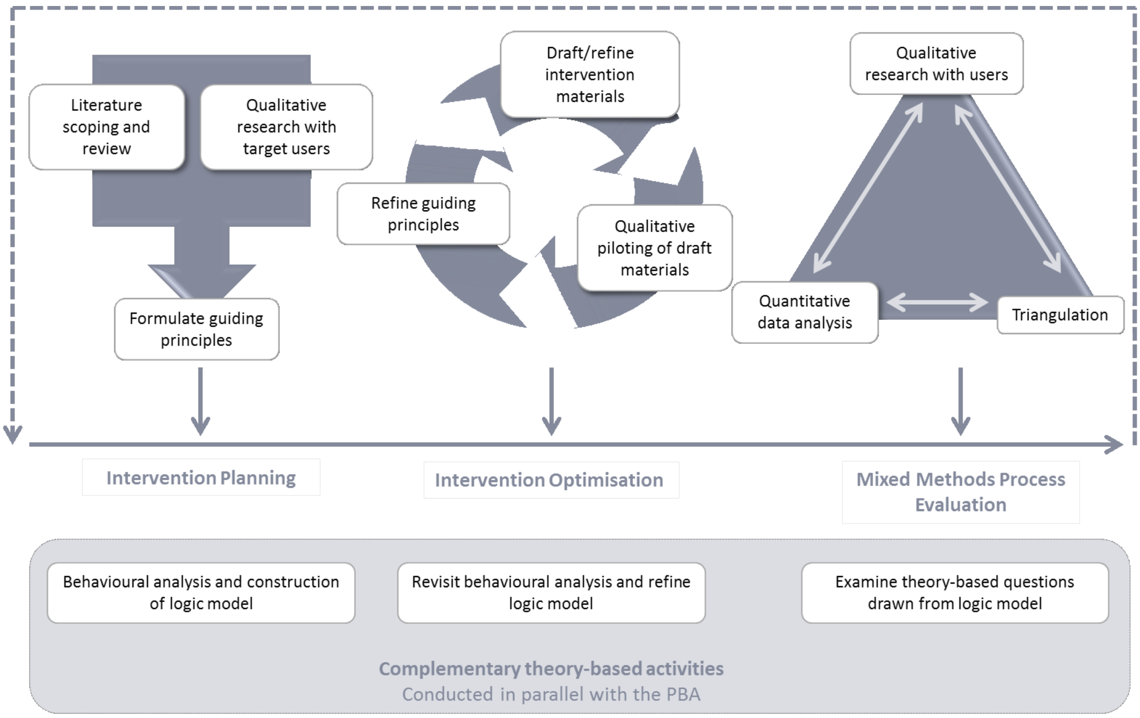


Figure 1. Overview of the person-based approach

design objectives for the intervention and the key intervention features that will support achievement of the design objectives. The design objectives specify what the intervention must do in order to address the needs of the target user (identified from the qualitative research) and enhance engagement with the intervention. For example, a core design objective guiding the development of our app-based stress management intervention was to provide a positive, useful and rewarding experience for users. This was informed by our qualitative work indicating that Smartphone users preferred apps that provided a clear and immediate personal benefit (e.g. practical, entertainment) that could be accessed in brief moments of free time (Dennison, Morrison, Conway, & Yardley, 2013).

The key intervention features specify how the design objectives may be achieved in practice. Intervention features can specify behaviour change techniques (BCTs), but also broader aspects of intervention delivery that suggest how specific BCTs may be implemented to ensure they are optimally persuasive and engaging (e.g. tone, language, structure, intended frequency of use, mode of delivery etc.). For example, to promote an immediately rewarding experience for our app users we designed for content to be accessed in less than three minutes with every app interaction offering the opportunity to see or unlock new content (Morrison et al., 2017).

Once formulated, the guiding principles can offer a succinct summary of the crucial ways in which the intervention is intended to support change in behaviour by improving engagement with the intervention content. We have found that a succinct, accessible summary of the intervention plan can also enhance communication across different disciplines and audiences to support multidisciplinary collaboration and facilitation of stakeholder events, Patient and Public Involvement (PPI) consultations etc. Since guiding principles only identify the crucial evidence-based design objectives they can be used as a quick check-point during intervention development to prioritise tasks and changes to the intervention (see Intervention

Optimisation).

Because guiding principles are grounded in a deep understanding of the users' context they are useful for guiding *how* theory- and evidence-based intervention content are delivered and communicated. This makes them a distinct but complementary tool that can be used alongside other theory-based approaches to intervention planning (e.g. behavioural analysis, construction of logic models, see Band et al., 2017).

Intervention Optimisation

The PBA is particularly valuable for intervention optimisation through inductive qualitative or mixed methods research to elicit detailed user feedback that enables researchers to understand people's views and experiences of using the prototype intervention and the various ways that people may choose to use it. Interventions are modified based on user feedback and then further research is carried out to ensure the modifications have achieved the desired effect of making the intervention and behaviour change elements acceptable, persuasive, and easier to use and adhere to. Guiding principles can also be refined as researchers gain more insights into the experiences and motivations of target users.

We normally use qualitative think-aloud methods to optimise interventions. This interview technique allows researchers to observe participants using the intervention while saying all their thoughts out loud, thus giving valuable insights into their experiences and views of the intervention. This is particularly useful in the earlier stages of intervention development as it can provide insights into every aspect of the intervention, ensuring it is persuasive, useable and acceptable to the people who will use it. In the later stages of intervention development, longitudinal studies can be useful for optimising interventions. This is where people are given an intervention to try on their own before being interviewed about their experiences of using the intervention. This method is particularly useful for assessing people's experiences of behavioural changes or techniques that may require practice.

Intervention optimisation provides insights beyond assessing the acceptability of interventions. In our Diabetes Literacy project, this stage of the PBA was crucial for improving the feasibility of intervention components (Rowsell et al., 2016). We developed a brief web-based intervention to promote physical activity in people with type 2 diabetes and low health literacy. One of the key features of the intervention was a physical activity planner, designed to help people find achievable ways to build on their current activity level. Observational think aloud interviews illustrated early on that people were vastly overestimating their current activity level when completing the planner, leading to participants with sedentary lifestyles receiving inappropriate tailored feedback congratulating them on being active enough. Observing participants complete the planner provided valuable insight into how and why people were incorrectly filling it in. It also highlighted ways the planner needed to be modified. Changes to the activity planner were made iteratively, enabling subsequent think aloud interviews to assess the impact of each change until the intervention was deemed feasible for evaluation in a clinical trial (Muller et al., 2017).

We find it helpful to systematically document all our sources of evidence and feedback and how these feed into optimising the intervention. User feedback from qualitative studies can be entered into a table, together with other sources of evidence such as PPI and expert input or other relevant evidence, to comprehensively record, categorise, and prioritise all changes to an intervention. See Bradbury et al., 2018 for a detailed description and illustration of this approach to qualitative data analysis and criteria for deciding when to implement intervention modifications.

Intervention Implementation

The PBA can also draw on mixed methods process evaluation of the implementation of complex interventions. Here the PBA can be used to understand people's experiences of a fully deployed intervention and highlight modifications which could help an intervention to be more effective in changing behaviour, or more successful in embedding in real-world contexts.

Qualitative process evaluations enable exploration of potential barriers to intervention success or implementation and can be triangulated with quantitative data on health outcomes, behavioural determinants, and intervention usage data, to provide a clearer picture of where the intervention might be working well and how it might need adjusting. Using the PBA, potential barriers to successful outcome or implementation can inform updates to an intervention plan (e.g. guiding principles) and further optimisation of the intervention, drawing on the methods described in the previous section.

Within the evaluation of our weight management intervention (POWeR+) we carried out a PBA qualitative process evaluation (Smith, Bradbury, Scott, Little, & Yardley, 2017) to explore how the intervention might need to be improved to ensure successful implementation in practice. POWeR + is a digital intervention, accompanied by a small amount of nurse support. Within our main trial (N=818) we tested the effectiveness of two types of brief nurse support: face-to-face support and remote support (by phone/email) (Little et al., 2016). Both were equally effective, with mean weight losses comparable to those seen within commercial weight loss interventions. Remote support was the most cost-effective and could be easier to implement at scale as it required less nurse time (Little et al., 2016). However, qualitative interviews with the nurses who provided support to POWeR+ patients highlighted that nurses did not believe that remote support was supportive enough to help patients to lose weight – a potential barrier to implementing this support in practice (Smith et al., 2017). This identified the need for a new guiding principle to be added to our intervention plan: to persuade practitioners that remote support is useful and effective. The key feature that we used to address this was to update our practitioner training materials to persuade practitioners of the value of remote support by showing them the evidence of its effectiveness (comparable to face-to-face support) and its acceptability to patients (through patient quotes).

The PBA advocates taking an inductive approach to collecting qualitative data, asking broad open questions (e.g. about what participants found helpful or unhelpful) in order to ascertain the most important issues or challenges for a participant. If at the evaluation or implementation stage researchers want to include some deductive, theory-based questions they can simply add these after inductive questions have been explored – this way participants' initial answers won't be prompted or influenced by the questions asked.

Conclusions

Although the PBA may seem resource intensive, we find that the time taken to understand users and their views of the intervention means that problems with user engagement are identified and resolved before evaluation and implementation, which avoids wasting resources on evaluating an intervention that will not prove engaging and effective. It is usually possible to persuade funders and collaborators to invest in this work by making this argument! However, the approach is intended to be used flexibly, with whatever methods and resources are available and most suitable. The PBA has evolved over the last decade and continues to evolve as we identify different and better ways of implementing it. For example, we now incorporate PPI input more explicitly and intensively, by forming stakeholder panels that feed in to the whole development process and can also provide rapid feedback and co-design input through regular meetings and consultations (paper in preparation). To reflect this continuous evolution and improvement we are celebrating our ten year anniversary by establishing a website (https://www.lifeguideonline.org/pba) which will provide a living archive and toolbox as we continue to publish papers describing and disseminating our methods.

Competing Interests

We have no competing interests.

References

Band, R., Bradbury, K., Morton, K., May, C., Michie, S., Mair, F. S., Murray, E., McManus, R. J., Little, P., & Yardley, L. (2017). Intervention planning for a digital intervention for self-management of hypertension: A theory- evidence- and person-based approach. *Implementation Science, 12*(25). doi:10.1186/s13012-017-0553-4

Bradbury, K., Morton, K., Band, R., van Woezik, A., Grist, R., McManus, R. J., ... & Yardley, L. (2018). Using the Person-Based Approach to optimise a digital intervention for the management of hypertension. *PloS one, 13*(5), e0196868. doi:10.1371/journal.pone.0196868

Corbett, T., Singh, K., Payne, L., Bradbury, K., Foster, C., Watson, E., … Yardley, L. (2018). Understanding acceptability of and engagement with web-based interventions aiming to improve quality of life in cancer survivors: A synthesis of current research. *Psycho-Oncology, 27*(1), 22-33. doi:10.1002/pon.4566

Dennison, L., Morrison, L., Conway, G., & Yardley, L. (2013). Opportunities and challenges for smartphone applications in supporting health behaviour change: Qualitative study. *Journal of Medical Internet Research, 15*(4), e86. doi: 10.2196/jmir.2583

Little, P., Stuart, B., Francis, N., Douglas, E., Tonkin-Crine, S., Anthierens, S., … Yardley, L. (2013). Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. *The Lancet, 382*(9899), 1175-1182. doi:10.1016/S0140-6736(13)60994-0

Little, P., Stuart, B., Hobbs, R. F. D., Kelly, J., Smith, E. R., Bradbury, K. J., ... Yardley, L. (2016). An internet-based intervention with brief nurse support to manage obesity in primary care (POWeR+): a pragmatic, parallel-group, randomised controlled trial. *The Lancet Diabetes & Endocrinology, 4*(10), 821-828. doi: 10.1016/S2213-8587(16)30099-7

Little, P., Stuart, B., Hobbs, R. F. D, Moore, M., Barnett, J., Popoola, D., … Yardley, L. (2015). An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): A primary care randomised trial. *The Lancet, 386*(1004), 1631-1639. doi:10.1016/S0140-6736(15)60127-1

Morrison, L. G., Hargood, C., Pejovic, V., Geraghty, A. W. A., Lloyd, S., Goodman, N., … Yardley, L. (2017). The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. *PLOS One, 13*(5), e0198008. doi: 10.1371/journal.pone.0198008

Muller, I., Rowsell, A., Stuart, B., Hayter, V., Little, P., Ganahl, K., … Nutbeam, D. (2017). Effects on engagement and health literacy outcomes of web-based materials promoting physical activity in people with diabetes: an international randomized trial. *Journal of Medical Internet Research, 19*(1). doi:10.2196/jmir.6601

Rowsell, A., Muller, I., Murray, E., Little, P., Byrne, C. D., Ganahl, K., ... Nutbeam, D. (2015). Views of people with high and low levels of health literacy about a digital intervention to promote physical activity for diabetes: a qualitative study in five countries. *Journal of Medical Internet Research, 17*(10). doi:10.2196/jmir.4999

Smith, E., Bradbury, K., Scott, L., Little, P., & Yardley, L. (2017). Providing online weight management in Primary Care: a mixed methods process evaluation of healthcare practitioners' experiences of using and supporting patients using POWeR+. *Implementation Science, 12*(69). doi: 10.1186/s13012-017-0596-6

Yardley, L., Morrison, L., Bradbury, K., & Muller, I. (2015). The person-based approach to intervention development: Application to digital health-related behavior change interventions. *Journal of Medical Internet Research, 17*(1), e30. doi:10.2196/jmir.4055

Dr Leanne Morrison
Psychology, University of Southampton
L.Morrison@soton.ac.uk

Dr Ingrid Muller
Psychology, University of Southampton & School of Experimental Psychology, University of Bristol
I.Muller@soton.ac.uk

Prof Lucy Yardley
Psychology, University of Southampton & School of Experimental Psychology, University of Bristol
L.Yardley@soton.ac.uk

Dr Katherine Bradbury
Psychology, University of Southampton
kjb1e08@soton.ac.uk

original article

Getting started with Network Meta-analysis in Health Psychology

Noone Chris
National University of Ireland, Galway
Joy Leahy
National University of Ireland, Galway & Trinity College Dublin
Eimear Morrissey
National University of Ireland, Galway
Laure Ngouanfo
University of Limerick
Tamara Navarro
McMaster University
Robby Nieuwlaat
McMaster University
Gerry Molloy
National University of Ireland, Galway

Introduction

Meta-analysis has been an important evidence synthesis methodology in health psychology and indeed many health sciences for several decades now (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018). Standard approaches to pairwise meta-analysis are clearly described in multiple accessible sources and commercial and free software to conduct meta-analyses are widely available (Borenstein, Hedges, Higgins & Rothstein, 2011; Field and Gillett, 2010). The fundamentals of the method are usually covered in post-graduate training in health psychology and frequently in undergraduate psychology courses. In the context of the replication crisis in psychology, meta-analysis has achieved even greater importance and visibility over the last 5-10 years (Open Science Collaboration, 2015). For example, it can help health psychologists identify more precise estimates of the magnitude of intervention effects, moderators of interventions effects, publication biases and indeed the absence of efficacy for some widely advocated approaches in the health psychology intervention literature (Hollands et al., 2016).

Indeed, in the wider literature evaluating complex interventions for health, standard pairwise meta-analysis is the data analytic mainstay of key evidence syntheses to inform healthcare practice. For example, provided that there are sufficient number of homogenous studies to synthesise, this approach is used in most Cochrane Reviews of RCT evaluations of healthcare interventions (Higgins & Green, 2011). One of the main limitations of pairwise meta-analysis, however, is that while it can tell whether an intervention works compared to something else e.g. 'treatment as usual' or a control condition, it cannot tell us which intervention is optimal out of all the available options for intervention. This is particularly problematic as many intervention approaches that may compete with each other for healthcare resources may not have been compared against each other within individual RCTs. Therefore, pairwise meta-analysis cannot address the critical research question of what intervention works best (Kanters et al., 2016).

A relatively recent data-analysis method where indirect comparisons can be made is known as **network meta-analysis or NMA** (Dias et al., 2018; Hutton et al., 2015). This approach has been developed over the last 10 to 15 years in the broader health literature and is gaining increasing prominence as a critical part of evidence synthesis, however the fundamentals are often unfamiliar to those working in health psychology and related fields (Molloy et al., 2018). In this paper, we will provide a short introduction to the key conceptual issues regarding NMA and a step-by-step tutorial, with accompanying annotated code, on the conduct of a NMA.

NMA can provide indirect comparison that allows assessment of comparative effectiveness of interventions that may not have been compared against each other within a single trial. This can be achieved when a number of conditions are met with the most fundamental being that studies have a control, treatment as usual or other intervention condition that is shared among the studies being compared – that is, we have a connected network of treatments. This allows for an indirect comparison to be made such as the one outlined in Figure 1 below. In this example, a number of studies have compared Intervention A with Intervention C, while others have compared Intervention B with Intervention C. NMA can be applied to estimate the indirect comparison between Intervention A and Intervention B. If direct comparisons between Intervention A and Intervention B exist, these can be synthesised with the indirect comparisons to produce a more accurate NMA estimate. Naci and Ioannidis (2013) produced an evidence network with a similar structure in one of their analyses where they synthesised direct comparisons between physical activity interventions and usual care, direct comparisons between antihypertensive drug interventions and usual care, and indirect comparisons between physical activity interventions and antihypertensive drug interventions.

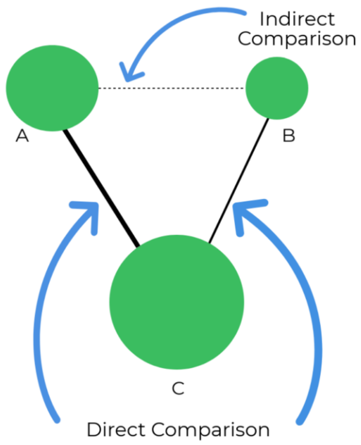


Figure 1. An evidence network including both direct and indirect comparisons.

The circles are referred to as nodes and represent each intervention. Their size usually represents the number of participants who received that intervention across all included studies. The lines connecting the nodes represent comparisons – solid lines indicate direct comparisons are present and dotted lines indicate that only indirect comparisons are possible. The thickness of the lines represents the number of studies which include that comparison.

In order to apply NMA validly, the assumption of *transitivity* must be met. When transitivity is present, it is assumed that any indirect comparison between two interventions in a network of evidence is a valid estimate of the direct comparison between these two interventions. When such direct comparisons do not exist, this assumption cannot be tested statistically. In these cases, transitivity can be qualitatively assessed by identifying potential effect modifiers (e.g. participant demographics, intensity of intervention, setting of intervention etc.) and assessing whether they are evenly distributed across the included studies (Salanti, 2012). When both direct and indirect comparisons exist, statistical tests of the *consistency* of the direct and indirect comparisons (i.e. their similarity) should be conducted (Dias et al., 2013).

Networks of Evidence in Health Psychology

There are specific considerations which need to be made when applying NMA to evidence from studies of behavioural interventions. This is because, in contrast to pharmacological interventions, on which the majority of studies applying NMA have focused so far, behavioural interventions are often made up of a number of different interacting components (Craig et al., 2013) and have much greater variation in the nature of their comparators (de Bruin et al., 2009).

This increases heterogeneity and affects the transitivity assumption.

The complex nature of behavioural interventions can affect transitivity because intervention components may be selected for specific groups or specific settings within the same patient population and this may introduce an uneven distribution of effect modifiers. Careful consideration of possible effect modifiers such as the setting, treatment intensity and participant characteristics is necessary. An extension of NMA – network meta-regression – can be applied to adjust for effect modifiers. Another important issue in considering evidence networks in health psychology is the content of control conditions. The control conditions to which behavioural interventions are compared are often complex too. Furthermore, they can vary significantly in their content which complicates the structure of the evidence network if several alternate interventions for a given behaviour and patient population are compared to several qualitatively different control conditions (de Bruin et al., 2009).The use of taxonomies developed within health psychology (e.g. Kok et al., 2016; Michie et al., 2013; Nudelman & Shiloh, 2015) can aid with the qualitative assessment of intervention components (including those within control conditions) and other effect modifiers.

When synthesising studies of complex interventions using traditional pairwise meta-analysis, we are forced to lump these interventions. This ignores the fact that these interventions are made up of a number of different components, the presence of which is likely to vary across the different interventions across the studies which are being synthesised. NMA allows us to represent different complex interventions as separate nodes in a network of evidence. Welton, Caldwell, Adamopoulos, & Vedhara (2009) explore four different modelling options for assessing the effects of components within complex interventions using NMA:

1. *Single Effect Model*: All behavioural treatments are grouped as one and compared to usual care.
2. *Additive Main Effects Model*: The effects of all components for each intervention are added together. This assumes that intervention components have independent treatment effects.
3. *Two-Way Interaction Model*: Allows for interactions between the components of each intervention. This assumes that the effect of one intervention component may enhance or diminish the effect of another intervention component.
4. *Full Interaction Model*: Each possible combination of components is treated as a different intervention.

The first model is a simply a traditional pairwise meta-analysis model. The fourth model is analogous to a standard NMA model, where each treatment is considered separately. In many NMAs we do not need to consider the middle models as the issue of multiple components does not arise. However, although models 2 and 3 are under-utilised at present, we recommend implementation of these models in health psychology to learn more about the nature of interactions between intervention components in complex interventions. These models should be tested against one another to determine which fits the data best (Caldwell & Welton, 2016).

A Tutorial on Applying Network Meta-analysis to Complex Health Interventions

The next section will describe an NMA of behavioural interventions for reducing systolic blood pressure (SBP) by increasing adherence to antihypertensive medication. Data and annotated code for the analyses presented are available at https://osf.io/6xp4s/. Use of this code requires the installation of R (R Core Team, 2016), which we recommend running through RStudio (RStudio Team, 2016), and WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). The following steps should be undertaken when carrying out an NMA:

1. Conduct a systematic review to identify relevant studies and code interventions.
2. Extract data from each study.
3. Select and run models.
4. Interpret and report the results.

It is necessary that the search, screening, intervention coding, data extraction and analysis are carried out according to a pre-specified protocol. We recommend using both the PRISMA-P (Moher et al., 2015) and PRISMA-NMA (Hutton et al., 2015) checklists to guide the development of the protocol.

Systematic Review

A systematic review and meta-analysis was conducted by Morrissey et al. (2016) to examine the effect of medication adherence interventions on blood pressure control in hypertension. While the review focused on a pairwise meta-analysis, subsequent work on the dataset has allowed a network meta-analysis to be conducted on the interventions focused on reducing SBP. SBP was chosen rather than DBP for illustrative purposes as it is considered to be the most clinically relevant biomarker of hypertension (Basile, 2009).

Interventions were coded according to the context of the delivery. The variation in delivery contexts were considered a priori to be the intervention components contributing most to the heterogeneity among the interventions and coding the interventions in this way allowed us to answer a substantive research question about the optimal mode of delivery of adherence interventions for people with hypertension. This coding was done by one reviewer and based on the intervention description provided in each paper. Details of the coding can be seen in Table 1. Among the 12 included studies, 6 unique interventions were identified. However, one of these interventions was composed of two separate components which meant that we needed to consider the complex intervention models as detailed by Welton and colleagues (2009). Therefore, the four models described earlier were tested against each other to clarify whether a single treatment effect underlies the difference between the behavioural interventions and usual care (model 1), whether independent treatment effects for each intervention component sum together to produce the treatment effect (model 2), whether independent treatment effects for each intervention component interact to produce the treatment effect (model 3) or whether each combination of intervention components produces a unique treatment effect (model 4). For the *Single Effect Model*, we could only use 11 studies as Svarstad (2013) did not have an arm for usual care.

Data from RCTs

When modelling a continuous outcome the mean in each group at the start of the study (mean at baseline), the mean change in each group, and the standard deviation (SD) of the change in each group are required. All studies reported the mean at baseline and the mean change in each group, (or we were able to compute the mean change using mean of each group at follow-up). However, most studies reported the SD at baseline and follow-up as opposed to the SD of the change. Using Higgins & Green (2011), it is possible to compute a correlation coefficient from studies which report all three SDs (baseline, follow-up, and change), and then use this coefficient to impute the SD of the change. Two studies in our analysis (Marquez Contreras, 2005; Marquez Contreras, 2006) reported all three SDs. We therefore computed a correlation coefficient from these studies. However, the five arms from these studies had very different

Table 1. *Description of included interventions.*

Name	Intervention Type	Context
Amado 2011	Educational intervention	Primary care and home (materials)
Amado 2011	Usual care	Usual care
Dusing 2009	Supportive measures	Primary care and home (materials)
Dusing 2009	Usual care	Usual care
Friedberg 2014	Stage matched intervention	Home (telephone)
Friedberg 2014	Educational intervention	Home (telephone)
Friedberg 2014	Usual care	Usual care
Hosseininasab 2014	Self-monitoring of BP	Home (materials)
Hosseininasab 2014	Usual care	Usual care
Ma 2013	Motivational interviewing	Primary care and home (materials)
Ma 2013	Usual care	Usual care
Marquez Contreras 2005	Mail intervention	Home (materials)
Marquez Contreras 2005	Telephone intervention	Home (telephone)
Marquez Contreras 2005	Usual care	Usual care
Marquez Contreras 2006	Self-monitoring of BP	Home (materials)
Marquez Contreras 2006	Usual care	Usual care
Morgado 2011	Pharmacist intervention	Secondary care
Morgado 2011	Usual care	Usual care
Ogedegbe 2012	Postive affect and educational intervention	Primary care and home (telephone)
Ogedegbe 2012	Educational intervention	Primary care and home (telephone)
Rudd 2004	Nurse management	Primary care and home (telephone)
Rudd 2004	Usual care	Usual care
Schroeder 2005	Nurse management	Primary care
Schroeder 2005	Usual care	Usual care
Stewart 2014	Pharmacist intervention	Primary care and home (materials)
Stewart 2014	Usual care	Usual care
Svarstad 2013	Pharmacist intervention	Primary care and home (materials)
Svarstad 2013	Patient information	Home (materials)
Tinsel 2013	Shared decision making	Primary care
Tinsel 2013	Usual care	Usual care
Wong 2013	Pharmacist intervention	Primary care and home (materials)
Wong 2013	Patient information	Usual care

correlation coefficients (0.22-0.71), with a weighted average (using the square root of the number of people in each arm) of 0.41. Two previous NMAs on blood pressure use a correlation coefficient of 0.5, which is close to the mean we obtained, so we also use 0.5 to impute the SD of the change for other studies (Follmann, Elliott, Suh, & Cutler, 1992; Welton et al., 2009). The choice of the correlation coefficient could potentially alter the results of the NMA, therefore we could test other values of the coefficient in what is known as a sensitivity analysis.

Models and Software

The NMA was carried out in WinBUGS (Lunn et al., 2000) using the R2WinBUGS package in R (Sturtz, Ligges, & Gelman, 2005). This is Bayesian software, based on Markov Chain Monte Carlo (MCMC), which uses an iterative process. When using MCMC we need to check for convergence by checking that the Brooks-Gelman-Rubin statistic is close to 1 (Gelman & Rubin, 1992, Brooks & Gelman, 1998). An acceptable threshold is generally 1.1. This is given by "Rhat" in the R2WinBUGS output.

We modelled an improvement in the SBP based on Schmitz and colleagues (2012) and Schmitz, Adams and Walsh (2013), with adjustments for multiple components based on work by Welton and colleagues (2009). These models use a random effects assumption which assumes that the true underlying effect can vary from study to study. These models are included in the appendix. As we have no treatment with three components the Two-Way Interaction Model and the Full Interaction Model simplify to be the same model, which we will refer to as the Interaction Model. We used the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) to distinguish between the three different models. Differences greater than three are usually deemed to mean that the model with the lower DIC has a better fit (Welton et al., 2009).

We checked for inconsistency by comparing our standard consistency model to an inconsistency model. The standard NMA (consistency) model assumes transitivity, i.e. it assumes that the estimate of the effect of treatment A relative to B must be equal to the sum of the estimate of C relative to A and the estimate of C relative to B. The inconsistency model, however, does not force this assumption, and instead estimates all relative treatment effects separately. For our analysis we used the interaction model specified by Dias and colleagues (2011) to check this assumption. We compared the deviance computed from both models, the DIC from both models, and the results of each treatment relative to usual care. Models are provided online. We expected a deviance contribution of approximately 1 from each datapoint, with higher deviances indicating a worse fit (Speigelhalter et al., 2002; Dias et al., 2011).

Methods for Summarising Results from the NMA

We calculated the difference in percentage reduction in all treatments versus usual care, taking the baseline value into account. As WinBUGS uses an iterative process we could store the rank of each intervention at each iteration of the MCMC chain, and use these values to estimate the probability of each intervention being in each position. We can then sum these probabilities to find the probability of each intervention being in each position or better, and plot these on a rankogram. Calculating the SUCRA (SUrface under the Cumulative RAnking curve; Salanti, Ades, & Ioannidis, 2011) gives us a one number summary for each intervention. Possible SUCRA scores range from 0 to 1. A treatment with a value of 1 means that it is the best intervention with no uncertainty, and a value of 0 mean that it is the

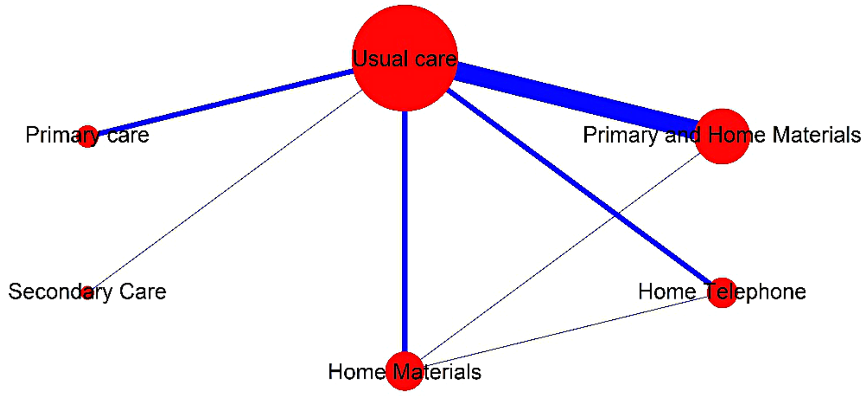


Figure 2. Network diagram (generated through pcnetmeta). Nodes and edges are proportional to the number of direct comparisons.

worst intervention with no uncertainty.

Results

Table 2. *Percentage reduction in SBP for each treatment versus usual care.*

Model	Intervention	Mean	SD	Lower CrI	Upper CrI
Single Main Effects Model	All Interventions	0.02	0.01	0.00	0.04
Additive Main Effects Model	Primary Care	0.00	0.02	-0.03	0.04
	Secondary Care	0.05	0.04	-0.02	0.12
	Home Telephone	0.00	0.02	-0.04	0.05
	Home Materials	0.02	0.02	-0.01	0.05
Interaction Model	Primary Care	0.01	0.03	-0.04	0.06
	Secondary Care	0.05	0.04	-0.03	0.13
	Home Telephone	0.01	0.02	-0.04	0.05
	Home Materials	0.02	0.02	-0.02	0.06
	Primary Care & Home Materials	0.02	0.02	-0.01	0.06

The network diagram can be seen in Figure 2. To compare the DIC across the three model we omitted the intervention study from Svarstad (2013) to ensure that we were comparing like with like. We found no difference between the three models. This is most likely due to the limited number of studies and, in particular, the fact that we only had one treatment node which involved more than one delivery context. We also compared the DIC using all 12 studies for the *Additive Main Effects Model* and the *Interaction Model*, and once again we found no difference. We therefore present the results of all three models.

The difference in percentage reduction in all treatments versus usual care is shown in Table 2. *The Single Effect Model* shows that the behavioural interventions grouped as one are superior to usual care at reducing SBP, with a Credible Interval (CrI), which does not span zero. However, for all other models all comparisons cross zero, which indicates that although the mean of each intervention is superior to usual care, we cannot be certain that these interventions have an effect on SBP compared to usual care.

The rankograms for the *Additive Main Effects Model* and the *Interaction Model* are shown in Figure 3. The SUCRA scores are shown in Table 3. We see that usual care is the lowest ranked intervention in each model. Secondary care is the

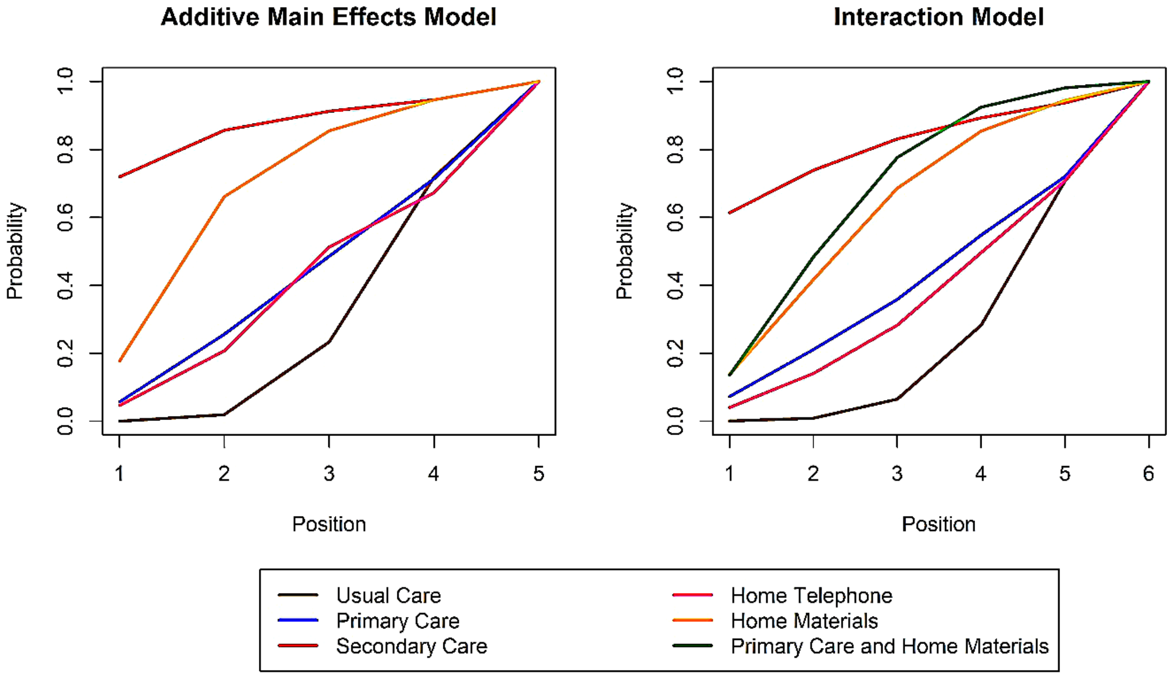


Figure 3. Rankogram for each treatment. At each point on the x-axis we see the probability of being in the nth position or better.

highest ranked intervention. It's worth noting that secondary care is the only intervention in our network that was included in one study only, so it may be that the intervention was applied particularly well in that study.

Checks for inconsistency

We can see from table 4 that the difference in DIC between the consistency and the inconsistency model is less than three so we find no meaningful difference in DIC. This indicates that it is correct to use the standard consistency model, which assumes transitivity. While there are some differences in the point estimates of some treatments versus usual care each mean is contained in the CrI of the other model. Figure 4 shows the deviance from the consistency model versus the deviance from the inconsistency model. Although there are some deviations from the line of equality, in absolute terms the differences are quite small. Overall, we find no concerning evidence of inconsistency between the models and therefore it is likely that the transitivity assumption holds. Therefore, the set of studies that we have included are likely to be suitable to analyse in an NMA.

Further Learning for Applying Network Meta-Analysis in Health Psychology

The effective application of NMA to networks of evidence in health psychology will require knowledge and skill in describing components of behaviour change interventions, managing and modelling data from RCTs and using statistical software packages that are infrequently employed by health psychologists. We recommend that readers stay up-to-date with the statistical courses and workshops such as those offered by the

Table 3. *SUCRA (SUrface under the Cumulative RAnking curve) score for each treatment. Higher values indicate better treatments. A treatment with a value of 1 means that it is the best intervention with no uncertainty, and a value of 0 mean that it is the worst intervention with no uncertainty.*

Single Effect Model	Intervention	Usual Care				
	0.99	0.02				
Additive Main Effects Model	Secondary	Home Materials	Primary	Home Tele	Usual Care	
	0.86	0.66	0.38	0.36	0.24	
Interaction Model	Secondary	PC & Home Materials	Home Materials	Primary	Home Tele	Usual Care
	0.80	0.66	0.61	0.38	0.33	0.21

Table 4. *Comparison of results from the consistency and the inconsistency model for the full interaction model.*

	Consistency Model				Inconsistency Model			
	Mean	SD	Lower CrI	Upper CrI	Mean	SD	Lower CrI	Upper CrI
Primary Care	0.01	0.03	-0.04	0.06	0.01	0.02	-0.03	0.05
Secondary Care	0.05	0.04	-0.03	0.12	0.05	0.03	-0.01	0.11
Home Telephone	0.00	0.02	-0.04	0.05	0.01	0.02	-0.02	0.05
Home Materials	0.02	0.02	-0.02	0.07	0.04	0.02	0.00	0.08
Primary Care & Home Materials	0.03	0.02	-0.01	0.06	0.01	0.01	-0.01	0.04
DIC	82.99				83.74			

University of Bristol, Oxford University, the Swiss Epidemiology Winter School and the Medical Research Council in the UK in order to avail of training in the application of NMA. A comprehensive treatment of NMA can be found in "Network Meta-analysis for Decision-making" by Dias and colleagues (2018). For a conceptual primer on the use of NMA in health psychology and behavioural medicine, see the work of Molloy and colleagues (2018).

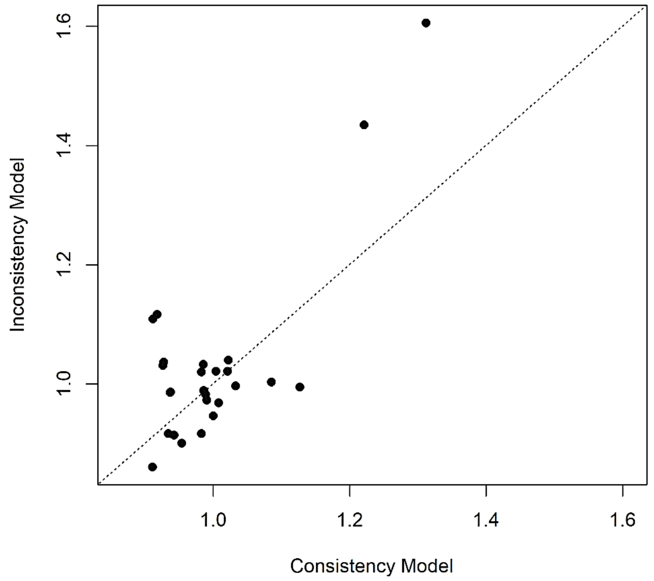


Figure 4. Deviance contribution from each study arm for the consistency model and the inconsistency model.

Conclusion

In this tutorial, we have discussed some basic concepts of NMA and demonstrated the application of NMA to a set of studies which examined the use of behavioural interventions to increase medication adherence in people with hypertension. By applying NMA to this network we have not only been able to address the question of whether these behavioural interventions work in terms of reducing blood pressure, but the more complex question of which intervention does this best by providing a ranking of behavioural interventions in terms of efficacy. However, due to the small number of studies, some uncertainty remains in these rankings.

Applying NMA in this manner is likely to have increasing importance for evidence synthesis in health psychology in the coming years. Appropriate application of the method requires adequate support from a multi-disciplinary team including biostatisticians to ensure that the synthesis of the evidence is reliable and valid. When used appropriately the method has the potential to influence the role of the health psychology in the delivery of healthcare, as it can help reveal important insights into the comparative effectiveness of behavioural interventions in health.

References

Amado Guirado, E., Pujol Ribera, E., Pacheco Huergo, V., & Borras, J. M. (2011). Knowledge and adherence to antihypertensive therapy in primary care: Results of a randomized trial. *Gaceta Sanitaria, 25*, 62–67. doi:10.1016/ j.gaceta.2010.09.015

Basile, J. (2009). Blood pressure responder rates versus goal rates: which metric matters?. *Therapeutic advances in cardiovascular disease, 3*(2), 157-174.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics, 7*(4), 434-455.

Caldwell, D. M., & Welton, N. J. (2016). Approaches for synthesising complex mental health interventions in meta-analysis. *Evidence-based mental health, 19*(1), 16-21.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: the new Medical Research Council guidance. *International journal of nursing studies, 5*0(5), 587-592.

de Bruin, M., Viechtbauer, W., Hospers, H. J., Schaalma, H. P., & Kok, G. (2009). Standard care quality determines treatment outcomes in control groups of HAART-adherence intervention studies: implications for the interpretation and comparison of intervention effects. *Health Psychology, 28*(6), 668.

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P., & Sutton, A. J. (2018). *Network Meta-Analysis for*

Decision-Making. John Wiley & Sons.
Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., & Ades, A. E. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making, 33*(5), 641-656.
Düsing, R., Handrock, R., Klebs, S., Tousset, E., & Vrijens, B. (2009). Impact of supportive measures on drug adherence in patients with essential hypertension treated with valsartan: The randomized, open-label, parallel group study validate. *Journal of Hypertension, 27*, 894–901. doi:10.1097/HJH.0b013e328323f9be
Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694.
Follmann, D., Elliott, P., Suh, I. L., & Cutler, J. (1992). Variance imputation for overviews of clinical trials with continuous response. *Journal of clinical epidemiology, 45*(7), 769-773.
Friedberg, J. P., Rodriguez, M. A., Watsula, M. E., Lin, I., Wylie-Rosett, J., Allegrante, J. P., …Natarajan, S. (2015). Effectiveness of a tailored behavioral intervention to improve hypertension control: Primary outcomes of a randomized controlled trial. *Hypertension,65*, 440–446. doi:10.1161/hypertensionaha.114.03483
Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science,* 457-472.
Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*(7695), 175.
Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions (Vol. 4)*. John Wiley & Sons.
Hollands, G. J., French, D. P., Griffin, S. J., Prevost, A. T., Sutton, S., King, S., & Marteau, T. M. (2016). The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. *BMJ, 352*, i1102.
Hosseininasab, M., Jahangard-Zafsanjani, Z., Mohagheghi, A., Sarayani, A., Rashidian, A.,Javadi, M., … Gholami, K. (2014). Self-monitoring of blood pressure for improving adherence to antihypertensive medicines and blood pressure control: A Randomized controlled trial. *American Journal of Hypertension, 27*, 1339–1345. doi:10.1093/ajh/hpu062
Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., ... & Mulrow, C. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of internal medicine, 162*(11), 777-784.
Kanters, S., Ford, N., Druyts, E., Thorlund, K., Mills, E. J., & Bansback, N. (2016). Use of network meta-analysis in clinical guidelines. *Bulletin of the World Health Organization, 94*(10), 782.
Kok, G., Gottlieb, N. H., Peters, G. J. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A., ... & Bartholomew, L. K. (2016). A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health psychology review, 10*(3), 297-312.
Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing, 10*(4), 325-337.
Ma, C., Zhou, Y., Zhou, W., & Huang, C. (2014). Evaluation of the effect of motivational interviewing counselling on hypertension care. *Patient Education and Counseling, 95*(2),231–237. doi:10.1016/j.pec.2014.01.011
Márquez-Contreras, E., Martell-Claros, N., Gil-Guillén, V., de la Figuera-Von Wichmann, M.,Casado-Martínez, J. J., Martin-de Pablos, J. L., … Serra, A. (2006). Efficacy of a home blood pressure monitoring programme on therapeutic compliance in hypertension: The EAPACUM-HTA study. *Journal of Hypertension, 24*, 169–175.10.1097/01.hjh.0000198023.53859.a2

Marquez Contreras, E., GarcÍa, O., Martel Claros, N., Gil Guillen, V., de la Figuera von Wichmann, M., Casado Martinez, J. J., & Fernández, R. (2005). Efficacy of telephone and mail intervention in patient compliance with antihypertensive drugs in hypertension. ETECUM-HTA study. *Blood Pressure, 14*, 151–158. doi: 10.1080/08037050510008977

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine, 46*(1), 81-95.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews, 4*(1), 1.

Molloy, G. J., Noone, C., Caldwell, D., Welton, N. J., & Newell, J. (2018). Network meta-analysis in health psychology and behavioural medicine: a primer. *Health psychology review*, 1-17.

Morgado, M., Rolo, S., & Castelo-Branco, M. (2011). Pharmacist intervention program to enhance hypertension control: A randomised controlled trial. *International Journal of Clinical Pharmacy, 33*, 132–140. doi:10.1007/s11096-010-9474-x

Morrissey, E. C., Durand, H., Nieuwlaat, R., Navarro, T., Haynes, R. B., Walsh, J. C., & Molloy, G. J. (2017). Effectiveness and content analysis of interventions to enhance medication adherence and blood pressure control in hypertension: a systematic review and meta-analysis. *Psychology & health, 32*(10), 1195-1232.

Naci, H., & Ioannidis, J. P. (2013). Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. *Bmj, 347*, f5577.

Nudelman, G., & Shiloh, S. (2015). Mapping health behaviors: Constructing and validating a common-sense taxonomy of health behaviors. *Social Science & Medicine, 146*, 1-10.

Ogedegbe, G., Tobin, J. N., Fernandez, S., Cassells, A., Diaz-Gloster, M., Khalida, C., …Schwartz, J. E. (2014). Counseling African Americans to control hypertension: Cluster-randomized clinical trial main effects. *Circulation, 129*, 2044–2051. doi:10.1161/circulationaha.113.006650

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston. Retrieved from http://www.rstudio.com/.

Rudd, P., Miller, N. H., Kaufman, J., Kraemer, H. C., Bandura, A., Greenwald, G., & Debusk,R. F. (2004). Nurse management for hypertension. A systems approach. *The American Journal of Hypertension, 17*, 921–927. doi:10.1016/j.amjhyper.2004.06.006

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research synthesis methods, 3*(2), 80-97.

Salanti, G., Ades, A. E., & Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of clinical epidemiology, 64*(2), 163-171.

Schmitz, S., Adams, R., Walsh, C. D., Barry, M., & FitzGerald, O. (2012). A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. *Annals of the rheumatic diseases, 71*(2), 225-230.

Schmitz, S., Adams, R., & Walsh, C. (2013). Incorporating data from various trial designs

into a mixed treatment comparison model. *Statistics in medicine, 32*(17), 2935-2949.
Schroeder, K., Fahey, T., Hollinghurst, S., & Peters, T. J. (2005). Nurse-led adherence support in hypertension: A randomized controlled trial. *Family Practice, 22*, 144–151. doi:10.1093/fampra/cmh717
Stewart, K., George, J., Mc Namara, K. P., Jackson, S. L., Peterson, G. M., Bereznicki, L. R., …Lau, R. (2014). A multifaceted pharmacist intervention to improve antihypertensive adherence: A cluster-randomized, controlled trial (HAPPy trial). *Journal of Clinical Pharmacy and Therapeutics, 39*, 527–534. doi:10.1111/jcpt.12185
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.
Sturtz, S., Ligges, U., & Gelman, A. (2010). R2OpenBUGS: a package for running OpenBUGS from R. URL http://cran. rproject. org/web/packages/R2OpenBUGS/vignettes/R2OpenBUGS.pdf.
Svarstad, B. L., Kotchen, J. M., Shireman, T. I., Brown, R. L., Crawford, S. Y., Mount, J. K., …Wilson, D. A. (2013). Improving refill adherence and hypertension control in black patients: Wisconsin TEAM trial. *Journal of the American Pharmacists Association, 53*, 520–529. doi:10.1331/JAPhA.2013.12246
Tinsel, I., Buchholz, A., Vach, W., Siegel, A., Dürk, T., Buchholz, A., … Fischer, K. G. (2013).Shared decision-making in antihypertensive therapy: A cluster randomised controlled trial. *BMC Family Practice, 14*, 1. doi:10.1186/1471-2296-14-135
Welton, N. J., Caldwell, D. M., Adamopoulos, E., & Vedhara, K. (2009). Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *American Journal of Epidemiology, 169*(9), 1158-1165.
Wong, M. C., Liu, K. Q., Wang, H. H., Lee, C. L., Kwan, M. W., Lee, K. W., … Griffiths, S. M. (2013). Effectiveness of a pharmacist-led drug counseling on enhancing antihypertensive adherence and blood pressure control: A randomized controlled trial. *The Journal of Clinical Pharmacology, 53*, 753–761. doi:10.1002/jcph.101

Chris Noone
School of Psychology, National University of Ireland, Galway
chris.noone@nuigalway.ie

Joy Leahy
School of Psychology, National University of Ireland, Galway
School of Computer Science and Statistics, Trinity College Dublin
joleahy@tcd.ie

Eimear Morrissey
School of Psychology, National University of Ireland, Galway
eimear.morrissey@nuigalway.ie

Laure Ngouanfo
Health Research Institute and MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland
Laure.Ngouanfo@ul.ie

Tamara Navarro
Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada
navarro@mcmaster.ca

Robby Nieuwlaat
Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada
nieuwlr@mcmaster.ca

Gerry Molloy
School of Psychology, National University of Ireland, Galway
gerry.molloy@nuigalway.ie

original article

Establishing determinant importance using CIBER: an introduction and tutorial

Gjalt-Jorn Y. Peters
Open University of the Netherlands
Rik Crutzen
Maastricht University

When developing behavior change interventions, it is important to target the most important determinants of behavior (i.e. psychological constructs that predict behavior). This is challenging for two reasons. First, determinant selection requires integrating multiple information sources: determinants' associations with either behavior or with determinant that mediate their effect on behavior (i.e. effect sizes), as well as how much room for improvement there is in the population (i.e. means and spread). Second, only information from samples is normally available, and point estimates obtained from samples vary from sample to sample, and therefore cannot be interpreted without information about how much they can be expected to vary over samples. In practice, determinant studies often present multivariate regression analyses, but this is problematic because by default, shared covariance is removed from the equation (literally), compromising operationalisations' validity and affecting effect sizes (i.e. the results of such analyses cannot be used as a first source of information regarding each determinant's association to behavior).

In the present contribution, we will briefly explain these points in more detail, after which we will introduce a solution: confidence interval based estimation of relevance (CIBER). We will then present a brief tutorial as to how to generate CIBER plots and how to interpret them. This is a more detailed explanation and introduction; originally CIBER was published in Crutzen, Peters & Noijen (2017).

Why determinant importance is important

Public health interventions have to potential to be cost-effective means to improve health and well-being (Masters, Anwar, Collins, Cookson & Capewell, 2017). They often do this by targeting human behavior. All overt human behavior is controlled from neurons in the motor cortex, activation of which occurs through activation of other networks of neurons (for more background, see Peters & Crutzen, 2017, and Crutzen & Peters, 2018). The networks of neurons that form a human brain can be considered the neural substrate of the entirety of human psychology. Therefore, while on a neuronal level, any successful behavior change intervention necessarily achieves this success by changing neural networks that ultimately activate motor cortex neurons, on a psychological level, any successful behavior change intervention can be said to necessarily achieve this success by changing aspects of the human psychology that are important for the target behavior.

Successfully changing aspects of human psychology requires learning in the target individuals (Crutzen & Peters, 2018). Humans have evolved several learning processes which, if properly leveraged, may realise this learning. These evolutionary learning principles correspond to different types of memory, and therefore, different evolutionary learning principles may be used to target different types of aspects of the human psychology (e.g. based on emotional memory, procedural memory, or declarative memory; see e.g. Aunger & Curtis, 2015). These evolutionary learning principles operate at a very fundamental level of human psychology, but psychologists studying behavior change have usually studied behavior change principles on higher levels of abstraction. For example, behavior change principles such as goal setting or planning coping responses represent packages of instructions that, when implemented properly, reliably engage one or more evolutionary learning processes. Two prominent lists of behavior change principles are the behavior change technique (BCT) taxonomy (Abraham & Michie, 2008) and the taxonomy of methods of behavior change (Kok et al., 2016), based on the Intervention Mapping framework for intervention development (Bartholomew, Parcel & Kok, 1998; Cullen, Bartholomew, Parcel & Kok, 1998; Bartholomew Eldredge, Markham, Ruiter, Fernàndez, Kok, & Parcel, 2016).

Similarly, psychologists have studied the aspects of human psychology that determine whether an individual performs a behavior on relatively high levels of abstraction. Many theories of behavior change propose constructs that predict behavior called determinants. These determinants, like other psychological constructs, have a definition and instructions for operationalisation. Psychological constructs can be operationalised in two ways: they can be measured and they can be manipulated. If a psychological construct is a determinant, its operationalisation into a manipulation is by definition a behavior change principle: to the degree that the determinant is important for the target behavior, changing the determinant also changes that target behavior.

Given the richness of human psychology, it is no surprise that there exist no 'magic bullet' behavior change principles that can always be relied on. Instead, which behavior change principles are most likely to be effective depends on which types of memories must be targeted (Crutzen & Peters, 2018). This link manifests as a pairing of determinants and behavior change principles, such that the likelihood of engaging the underlying evolutionary learning principles is optimal. Note that this is also true for efforts to change behavior that are based on an ecological approach. Aspects of individuals' environments (contextual factors, environmental conditions, et cetera) cannot have any influence on the behavior of those individuals without changing aspects of their psychology. An individual's behavior, after all, is exclusively controlled by activation patterns in their motor cortex; and those activation patterns cannot be changed directly, but only through changes in other aspects of the individual's psychology (Crutzen & Peters, 2018).

As a consequence, a crucial step in the development of behavior change interventions is the selection of the most important determinants. Colloquially, these determinants can be seen as the buttons one needs to push to establish behavior change.

When a determinant is important

Determinant importance depends on two things. The first is the determinant's association to behavior, or, as is often the case, to a theoretical mediator of the determinant's effect on behavior. For example, when an interventon developer develops an intervention for a reasoned behavior, a suitable theory may be the Reaoned Action Approach (RAA; Fishbein & Ajzen, 2010). This theory holds that behavior is predicted by a determinant called intention (i.e. a person's intention to engage in the behavior), which in turn is predicted by three other determinants: attitude (a person's evaluation of the behavior's consequences), perceived norms (a person's perception of the approval and behavior of relevant social referents), and perceived behavioral control (a person's perception of their ability and control over the behavior). If a determinant study is conducted and the correlation of attitude to intention and behavior is zero, it seems unlikely that changes in attitude will result in behavior change. However, even if a determinant is strongly associated to behavior or a theoretical mediator, it may still not be a relevant intervention target.

This is because of the second thing that determinant importance depends on: the distribution of the determinants' scores in the population (as estimated by inspecting the distribution of sample scores). A determinant that is strongly associated to behavior may still be a bad choice as intervention target if its distribution is very skewed. For example, most ecstasy users are aware that using a high dose of ecstasy is bad for their health. Even if this variable is strongly associated to their behavior, this association is caused by only a few people who deny these health effects. When developing an intervention, investing resources in targeting this small group will yield less total effects on behavior than when targeting a determinant with a weaker association but with more room for improvement.

Note that this reasoning does not only hold when selecting determinants (such as attitude), but also when selecting subdeterminants. Subdeterminants are here defined as determinants at a lower level of psychological generality that are theoretically assumed to predict or be a part of overarching determinants. This definition means that whether a determinant is called a 'determinant' or a 'subdeterminant' is somewhat arbitrary. For example, within the RAA, attitudinal beliefs such as expectancies or risk perceptions can be called subdeterminants, because they are theoretically assumed to predict, or be a part of, their 'overarching' determinant attitude. At the same time, attitude, perceived norms, and perceived behavioral control can be called subdeterminants because they are theoretically assumed to predict, or be a part of, their 'overarching' determinant intention (note that perceived behavioral control is also assumed to influence behavior directly, so the case could be made that labeling it a subdeterminant would be inaccurate).

So, to summarize, successful behavior change requires successful change of one or more aspects of human psychology. These aspects are defined in, and can be operationalised using, psychological theory, and are called (sub-)determinants. Once operationalised, their importance can be established to identify the best intervention targets. Establishing this (sub-)determinant importance requires simultaneous inspection of the determinant's association to theoretical mediators of its effects on behavior, potentially to behavior directly, and of the determinant's distribution. Most researchers do this by computing point estimates (e.g. correlation coefficients), but unfortunately, these are virtually uninformative on their own.

Why point estimates cannot be used to estimate determinant importance

When inspecting association and distribution estimates, the population values are always unknown. The only way to learn about a population is by taking a random sample and inspecting that sample. This instrument, however, is somewhat of a mixed blessing. On the one hand, sampling provides the researcher with a way to 'look at' the population. On the other hand, sampling, by its random nature, necessarily introduces random variation. This means that whatever is observed in the sample may not reflect the population.

This creates the somewhat frustrating situation that the only means available to observe a population also inevitably distort that observation. Any value computed from a sample will have a different value if the sampling is repeated. Therefore, the specific estimate arrived at on the basis of any particular sample has next to no value. It is also necessary to know how accurate the estimate is: how much it can be expected to differ between samples. Fortunately, there is a way to estimate this.

This estimation of accuracy is based on the concept of the sampling distribution: the theoretical distribution containing all potential values for any sample estimate, given its (unknown) population value and the sample size. Because the population value is always unknown (otherwise one wouldn't have to sample in the first place), the true sampling distribution is necessarily also known. However, for many parameters that can be estimated from a sample, the shape and spread of the sampling distribution are known. This means that the sampling distribution can be constructed for any hypothetical population value.

The best known example is perhaps the sampling distribution of the mean, which is approximately normally distributed (except for extremely small samples) with a standard deviation equal to the population standard deviation divided by the square root of the sample size. Knowing the sampling distribution's distribution shape and spread allow computation of intervals that contain, in infinite repetitions of the sampling procedure, the population value in a given percentage of the samples: the confidence interval. A wide confidence interval means that the point estimate is very unreliable and can have a substantially different value in a new sample, whereas a tight confidence interval means that a substantially different value in a new sample is less likely. These properties, in combination with the fact that health psychologists are generally familiar with confidence intervals, make them well suited for estimation of population values from sample data.

Therefore, whenever using sample data to draw conclusions for intervention development (or anything, really), point estimates should not be used. Instead, also considering estimate accuracy, for example by computing confidence intervals, allows taking the inevitable sampling and error variation into account. However, this also means that inspecting determinant importance becomes almost an inhuman task: one has to simultaneously compare three times as much information (e.g. means and correlations coefficients, as well as the confidence intervals regarding both point estimates). Visualisation can help, and this is what confidence interval based estimation of relevance (CIBER) is based on. CIBER plots simultaneously visualise (sub-)determinant distributions, confidence intervals for the mean, and confidence intervals for bivariate correlations to one or more theoretical mediators and/or behavior. Before explaining how to order and read a CIBER plot, we will explain why CIBER plots use correlations instead of regression coefficients.

Why regression coefficients cannot be used to estimate determinant importance

Determinant studies often contain regression analyses where a theoretical mediator of determinants' effects on behavior (e.g. intention) or behavior itself, is regressed on the measured determinants (or subdeterminants). Such regression analyses are useful, because they yield a multiple correlation coefficient: the correlation of the criterion (dependent variable) with the best prediction of the criterion as computed from the predictors in the model. Squaring this multiple correlation coefficient yields R^2, the proportion of the variance in the criterion that can be explained by the predictors in this sample. Because the distribution of R^2 is known, a confidence interval can be constructed, allowing tentative conclusions as to likely population R^2 values, which is indicative of the maximum effect that can be expected of an intervention that successfully changes all determinants in the model.

A convenient feature of regression analysis is that overlap between predictors in their explanation of the criterion is removed from the equation (quite literally, in the case of regression). Squaring a correlation coefficient always yields the proportion of explained variance: if attitude and intention have a bivariate (i.e. zero-order) correlation of r = .32, that means that they each explain .1 (i.e., .32 x .32) of each other's variance in the sample. The 95% confidence interval runs from [0.03; 0.19], which gives some idea of how far the explained variance in the population can be expected to deviate from that sample estimate. Another determinant, self-identity, has a correlation of r=.47 with intention, and so this determinant explains .22 of intention.

However, attitude and self-identity correlate with each other (r = .32). It is therefore likely that they also share explained variance in intention. In that case, simply adding together the proportion of intention's variance they each explain (.1 + .22 = .32) would yield an overestimate of how much intention these determinants explain together (which is in fact .25 in this sample, with a 95% confidence interval of [0.15; 0.36]).

This correction of overlap in explained variance is very useful, and enables better estimation of the variance explained by all predictors together. However, this overlap between predictors is in itself highly problematic when dealing with the separate regression coefficients of all psychological constructs used as predictors (Azen & Budescu, 2003; Budescu, 1993; Elwert & Winship, 2014). This problem is in part the consequence of potential overlap in the operationalisations of these psychological constructs.

Assuming the applications of the used operationalisations in the relevant sample have high validity (after all, if they have low validity, it makes no sense to analyse the resulting data), correlation between the corresponding data series represents relevant information about human psychology. For example, the two constructs may cover the same aspects of human psychology according to their definition. In that case their operationalisations will also measure the same aspects of human psychology, and therefore, the data series generated by these operationalisations will correlate. Or alternatively, the constructs may be independent but causally related, either because they influence each other (directly or through one or more mediators) or are both influenced by the same third variable. As we argued before, it is hard to empirically distinguish between constructs that influence or consist of each other (Peters & Crutzen, 2017), and the distinction is irrelevant with respect to the problem that surfaces in multivariate analyses.

In this case, removing the variance representing this overlap from the data series corresponding to a construct's operationalisation means removing variance that corresponds to aspects of human psychology that fall within the definition of the construct. In other words, removing this shared variance from a determinant and only considering variance that is not shared with other determinants means that the resulting data series no longer represents the determinant as originally operationalised, and therefore, as defined, but an unknown alteration of this determinant.

This is a necessary consequence of observational research: if two dataseries share explained variance in a third dataseries, it is impossible to know to which dataseries the shared explained variance 'belongs'. In fact, it is likely it belongs to both: if the correlation between the dataseries is indicative of overlap in the definitions of the two constructs that correspond to the data series, these two constructs explain the same aspects of the criterion. Therefore, removing this shared explained variance when estimating the regression coefficients means that these regression coefficients no longer represent the association of each predictor to the criterion. Instead, they represent the association of some unknown part of each predictor with some unknown part of the criterion.

Another way to think about this is by using the formulation often invoked when explaining regression analyses: the regression coefficient expresses the association of a predictor to the criterion *holding all other predictors constant*. If two predictors overlap in their definition, or, in other words, if the definitions of the constructs represented by the two predictors contain the same aspects of human psychology, then 'holding all other predictors constant' means 'neglecting a part of human psychology'. This means the resulting situation is unrealistic and can never occur. Given that the operationalisations of both constructs was valid, this also means that the omitted aspects of human psychology are in fact important to predicting the relevant behavior. Therefore, a predictor that represents an important determinant of behavior may nonetheless have a small regression coefficient, because an important part of the human psychology as defined in the constructs definition was omitted from the coefficient.

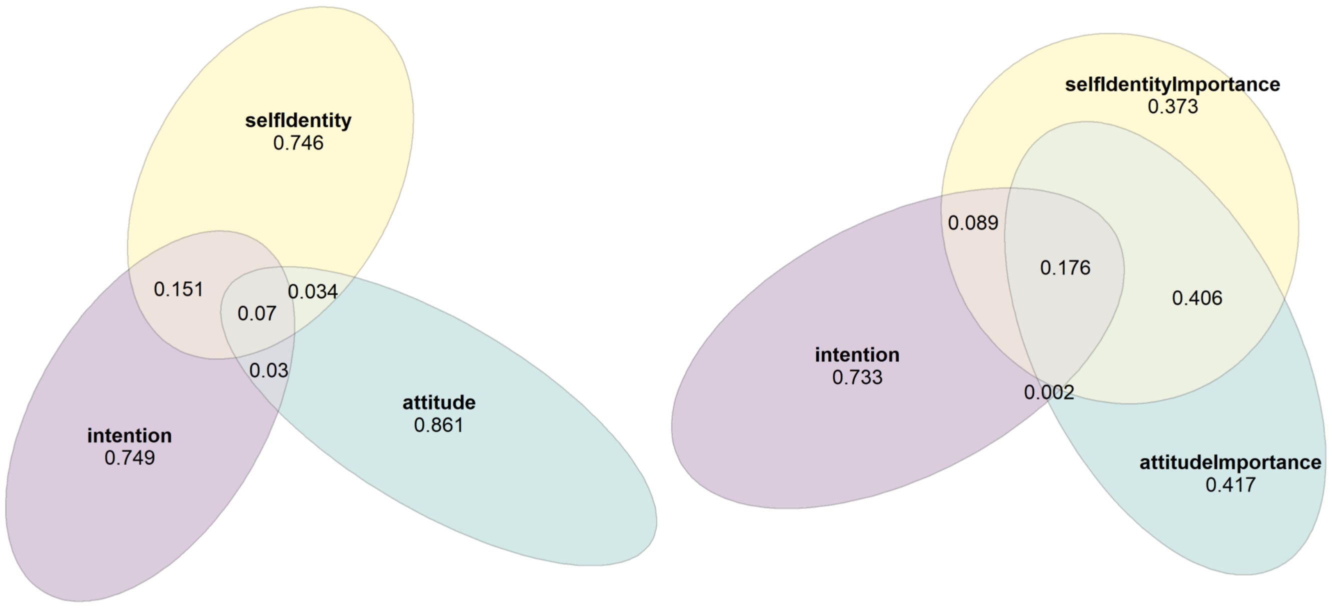
Because this can be hard to grasp, we include an example. Imagine we do a small-scale determinant study. We measure two determinants of intention: attitude and self-identity. Self-identity is one of the variables explicitly covered by Fishbein and Ajzen (2010) in their discussion of potential fourth variables that could be added to the Theory of Planned Behavior (or, by implication, its successor, the Reasoned Action Approach). They argued that the concept was ill-defined, and that common operationalisations actually covered the perceived 'importance' of a behavior. They argued that this can be considered part of the attitude construct, and therefore, including 'importance scale' in attitude's measurement would eliminate any additional explained variance by self-identity: "[...] if importance scales were included in the semantic differential measure of attitude, obtaining a separate measure of self-identity by means of importance items would be of little value." (Fishbein & Ajzen, p. 292). Given that importance can clearly be considered both a part of attitude and self-identity, this lends it well to an illustration of our point.

In this hypothetical determinant study, therefore, we include the importance scale in addition to the determinants (attitude and self-identity) and the criterion (intention). We have included the items used in this hypothetical study in the R Markdown file in the supplementary materials (see the Open Science Framework at https://osf.io/hg4ks/). The correlations used in the earlier illustrations were in fact derived from the dataset we simulated for this hypothetical determinant study. Figure 1 shows two Venn Euler diagrams that use the eulerr package (Larsson, 2018) to show the proportional areas of overlap in explained variance between the three variables in this determinant study.

The left diagram shows the situation where attitude and self-identity are operationalised without including the importance scale. Therefore, these variables represent a more limited definition of the attitude and self-identity constructs. In this sample, the correlation coefficients with intention are .32 for attitude and .47 for self-identity, they together explain .25 of the variance in intention, and their regression coefficients are respectively .18 and .41 (all variables are standardized). As the left diagram shows, the squared correlation between attitude and intention is $r^2 = .03 + .07 = .10$, and the squared correlation between self-identify and intention is $r^2 = .151 + .07 = .221$. In this situation, .07 or seven percent of the covariance between the variables is omitted from the equation when the regression coefficients are estimated.

The right diagram shows the situation where the

Figure 1: Venn Euler diagrams showing the overlap in explained variance between attitude, self-identity, and intention. In the diagram on the left, the importance scale is left out of the operationalisation op attitude and self-identity; in the diagram on the right, it is included in both operationalisations.



importance scale is included in the operationalisation of both constructs. The definitions of both constructs are therefore more broad than in the left diagram; but note that these broader definitions can be argued to be correct, and can conceivable be used in the same study. In this sample, the correlation coefficients with intention are .42 for attitude and .51 for self-identity, they together explain .27 of the variance in intention, and their regression coefficients are respectively .07 and .46. This diagram shows the large overlap between the variables: .176 of the variance is shared between attitude, self-identity, and intention. This .176 represents almost twenty percent of the variance in intention that cannot be designated to one of the predictors (and therefore, is not reflected in their regression coefficients).

In the left situation, the correlations indicate that both attitude and self-identity seem feasible intervention targets. When removing their overlap, the apparent feasibility of attitude drops a bit, and although this paints a slightly misleading picture by exaggerating the differences in importance between attitude and self-identity, the effect is quite subtle.

However, the right graph paints a different picture. When both predictors represent determinants that are defined, and operationalised, as partly covering the same aspects of human psychology, the difference between correlations and regression coefficients becomes substantial. Whereas the correlation coefficients would again imply that both determinants are feasible intervention targets, based on the regression coefficients, attitude seems irrelevant to predicting intention.

That conclusion, however, would be wrong. It would be valid only if one would redefine attitude such that all overlap with self-identity, in the prediction of intention, is removed from attitude's variance. That would mean the resulting data series (i.e. the residuals) no longer represent the determinant attitude. After all, that construct's definition did include importance.

It is unclear what exactly the two remaining data series do represent. Psychological constructs often covary, and this covariance represents not bias or measurement error, but real aspects of human psychology. Removing such covariance from estimations of a construct's importance means that it is no longer clear what is being inspected.

This becomes problematic when engaging in behavior change. For example, in this case, the intervention developer may mistakenly decide to only try and target self-identity. While for attitude, a wealth of behavior change methods exists (see Kok et al., 2016), for self-identity, no effective methods have been identified in the available lists of behavior change principles, and while some may exist, it seems likely that successfully changing self-identity is much harder than successfully changing attitude.

Thus, because estimates from multivariate analyses are problematic when establishing determinant relevance, it is better to base such decisions on the bivariate correlations, or more accurately, on the confidence intervals for these correlation coefficients, together with the information about the (sub-)determinants' distributions and means. We will now illustrate a method for efficiently inspecting all this information simultaneously: confidence interval based estimation of relevance.

Confidence interval based estimation of relevance

To illustrate confidence interval based estimation of relevance (CIBER), we will use four subdeterminants of attitude as these allow a more complete demonstration. The resulting CIBER plot is shown in Figure 2 (we refer readers who are interested in the CIBER plots obtained from the determinants' association with intention to the OSF repository of this article at https://osf.io/hg4ks/).

A CIBER plot contains a large amount of information. First, the left-hand panel shows the questions used to measure these subdeterminants, the left and right anchors of the answer scales, each participants' score, and a 99.99% confidence interval for the mean. This allows easy spotting of skewed distributions or other deviations from normality which are important to take into account when selecting determinants for intervention (with these simulated data, these distributions are approximately normal; for a real-life example, see Crutzen, Peters & Noijen, 2017).

The right-hand panel shows each subdeterminant's association to both attitude and intention. Each correlation coefficient is represented by a diamond showing the point estimate as well as the lower and upper bounds. Because attitude is the mean of the scors on these four items, the correlations of the subdeterminants with attitude are very high, while the correlations with intention are considerably lower.

Finally, the CIBER plot's title shows the proportions of explained variance. This title simultaneously functions as a legend to identfty which diamonds correspond to which determinant. As can be seen here, the proportion of explained variance of attitude could not be estimated; this makes sense, because it is necessarily 1 (after all, attitude is the mean of the four subdeterminants).

To generate a CIBER plot, a function is available in the free R package userfriendlyscience (Peters, 2017). To install the package in R, use the following command;

```
install.packages('userfriendlyscience';
```

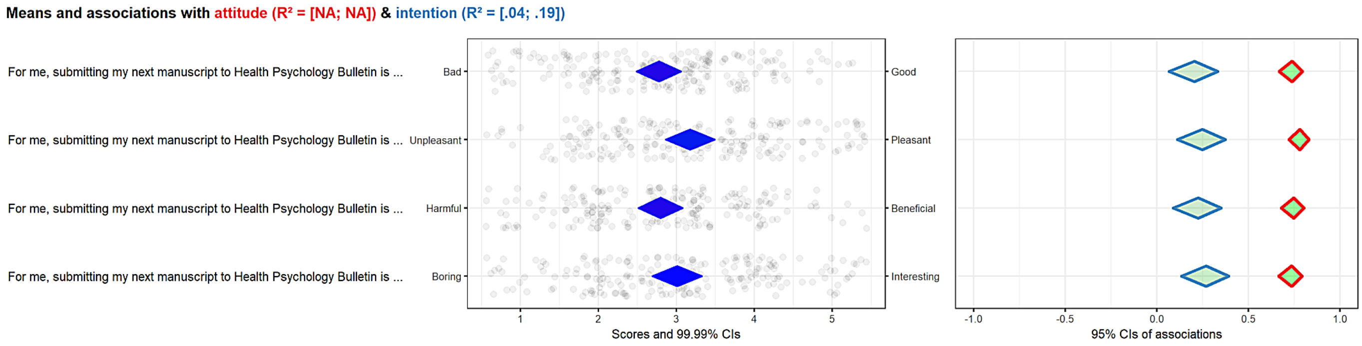
This command is necessary only once; once the package has been installed, it will remain available. After having installed the package, it can be loaded in an R session by using the following command:

require('userfriendlyscience');

This needs to be repeated in every R session (because R has thousands of packages available, these are not all automatically loaded every time; users can indicate which packages they need in a session).

Then, the CIBER plot can be generated using the CIBER command. For example, a simple version of

Figure 2: A CIBER plot showing hypothetical subdeterminants of attitude (i.e. attitudinal beliefs), their distributions and means (left panel) and their association to attitude and intention (right panel).



the CIBER plot shown in Figure 2 can be obtained with this command:

```
CIBER(data=dat,
determinants=c('attitude_good',
'attitude_pleasant',
'attitude_beneficial',
'attitude_interesting'),
               targets=c('attitude',
'intention'));
```

In this command, the first argument ('data') specifies the data to use. This is a dataset that can be loaded into R using, for example, the getData command:

```
dat<-getData();
```

This will open a popup dialog where a datafile can be selected. The selected datafile is then read into memory and named dat (in R, multiple datasets can always be open, and therefore, naming a dataset when loading it is mandatory). The other two arguments, 'determinants', and 'targets' specific the variable names of the determinants (the rows of the CIBER plot) and the the higher level determinants or behavior variables with which to show associations in the right-hand panel. Thus, in the simplest case, it is possible to simply load one's dataset into R using the getData command and then use the CIBER command to specify which determinants and targets to plot.

It is also possible to customize the plot by specifying, as was done in Figure 2, the questions used for each (sub-)determinant by using the 'subQuestions' argument; the left and right anchors by using the 'leftAnchor' and 'rightAnchor' arguments, and it is also possible to change the colors and set other options. An overview of all available options is available by using the following command, which will load the manual page for the CIBER command:

```
?CIBER
```

Conclusion

Establishing the relative importance of a set of (sub-)determinants, to then select the best intervention targets and be able to select the most fitting behavior change principles (e.g. methods for behavior change or behavior change techniques), is no straightforward affair. There are a number of potential pitfalls. In this article, we aimed to describe these pitfalls, explain why they are problematic, and we present an easy-to-use solution that is freely available. CIBER plots allow researchers and intervention developers to simultaneously evaluate the large amounts of information that need to be evaluated to select the determinants to target in an intervention to optimize the probability of successful behavior change. We hope this can contribute to more informed determinant selection and ultimately, more effective behavior change interventions.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27*(3), 379–87. dio:10.1037/0278-6133.27.3.379

Aunger, R., & Curtis, V. (2015). *Gaining control: how human behavior evolved*. Oxford: Oxford University Press.

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*, 129–148.

Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention Mapping: A process for developing theory- and evidence-based health education programs. *Health Education and Behavior, 25*(5), 545–563. doi:10.1177/109019819802500502

Bartholomew Eldrigde, L. K., Markham, C. M., Ruiter, R. A. C., Fernàndez, M. E., Kok, G., & Parcel, G. S. (2016). *Planning health promotion programs: An Intervention Mapping approach*. San Francisco: Jossey-Bass.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114,* 542–551. doi: 10.1037/0033-2909.114.3.542

Cullen, K. W., Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention mapping: use of theory and data in the development of a fruit and vegetable nutrition program for girl scouts. *Journal of Nutrition Education, 30*(4), 188–195.

Crutzen, R., & Peters, G.-J. Y. (2018). Evolutionary learning processes as the foundation for behaviour change. *Health Psychology Review, 12*(1), 43–57. doi: 10.1080/17437199.2017.1362569

Crutzen, R., Peters, G.-J. Y., & Noijen, J. (2017). Using Confidence Interval-Based Estimation of Relevance to select social-cognitive determinants for behaviour change interventions. *Frontiers in Public Health*. doi: 10.3389/fpubh.2017.00165

Elwert, F., & Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annual Review of Sociology, 40*(1), 31–53. doi:10.1146/annurev-soc-071913-043455

Fishbein, M., & Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York: Taylor & Francis Group.

Larsson, J. (2018). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. *R package version 4.1.0*, https://cran.r-project.org/package=eulerr

Kok, G., Gottlieb, N. H., Peters, G. Y., Mullen, P. D., Parcel, G. S., Ruiter, R. A. C., … Bartholomew, L. K. (2016). A taxonomy of behavior change methods: an Intervention Mapping approach. *Health Psychology Review, 10*(3), 297–312. doi: 10.1080/17437199.2015.1077155

Masters, R., Anwar, E., Collins, B., Cookson, R., & Capewell, S. (2017). Return on investment of public health interventions: a systematic review. *Journal of Epidemiology & Community Health, 71*(8), 827–834. doi:10.1136/jech-2016-208141

Peters, G.-J. Y. (2017). Diamond Plots: a tutorial to introduce a visualisation tool that facilitates interpretation and comparison of multiple sample estimates while respecting their inaccuracy. *Health Psychology Bulletin* (under Review). Retrieved from http://diamondplots.com

Peters, G.-J. Y., & Crutzen, R. (2017). Pragmatic nihilism: how a Theory of Nothing can help health psychology progress. *Health Psychology Review, 11*(2). doi: 10.1080/17437199.2017.1284015

Gjalt-Jorn Ygram Peters
Department of Methodology and Statistics, Faculty of Psychology and Education Science, Open University of the Netherlands
gjalt-jorn@behaviorchange.eu

Rik Crutzen
Department of Health Promotion, Maastricht University/CAPHRI
rik.crutzen@maastrichtuniversity.nl

original article

Combining explicit and implicit measures when measuring health dilemmas

Marleen Gillebaart
Utrecht University

Dilemmas are a core aspect of health behavior. Many people hold intentions and goals with respect to various aspects of their health and the concurrent behavior, including diet, exercise, and sleep. However, people are also subjected to several dilemmas concerning these long-term goals in daily life. These dilemmas often include short-term temptations (e.g., sugary snacks, canceling a gym class, staying up late on a work night) that are not in line with long-term health goals (e.g., a healthy BMI, being in good shape, feeling fit at work), and that cannot both occur at the same time. Because of these dilemmas, people's health behaviors are sometimes suboptimal and not in line with their long-term goals. This phenomenon has been labeled the intention-behavior gap (Sheeran, 2002). The notion that people's health behavior does not always align with their intentions has implications for studying health behavior.

Research into the determinants of behavior has early on pegged a number of factors that influence our intentions to behave in certain ways. A prime example is the theory of planned behavior, proposing perceived control, social norms, and attitude as primary influences on behavior (Ajzen, 1991). However, much of this research has focused on intentions rather than behavior, and understandably so, since the determinants of actual behavior may me much more complex and difficult to oversee. For instance, there is a wide array of nonconscious processes like biases and heuristics, as well as environmental cues that trigger behavior apart from intentions (e.g., Gigerenzer & Gaissmaier, 2011; Kahneman, 2011; Tversky & Kahneman, 1974). These influences often remain obscured from introspection and other forms of explicit measurement (Nisbett & Wilson, 1977; Wilson & Schooler, 1991). As such, the emergence of implicit measures in psychology has great potential, and has already significantly benefitted the field of health behavior. In this paper, we will discuss how explicit (in this case, self-reports that rely on introspection) and implicit measurement (measurements that are designed to tap into otherwise unaccessible aspects of behavior or its underlying processes) of health behavior dilemmas has developed recently, and what implications as well as complications that may hold for the field.

A health dilemma, or response conflict, emerges when people are confronted with different behavioral tendencies that cannot be combined into one behavior. Oftentimes, these dilemmas include a short-term goal and a long-term goal. For example, for someone with a dieting goal, temptations are everywhere during the day, and dilemmas ensue when one is offered a biscuit with their tea, birthday cake from a colleague, or a good-looking dessert at a restaurant. For someone with an exercise goal, there are the ever-lurking temptations of Netflix and napping on the couch. To handle these dilemmas, people have to use their self-control: the capacity they have to inhibit impulses and initiate behavior into the direction of their long-term goal (Carver, 2005; Friese, Hofmann, & Wiers, 2011; Myrseth & Fishbach, 2009).

Explicit measures

When one wants to study health dilemmas, explicit self-reports have proven to be a valuable mode of measurement. For example, in the field of ambivalence (i.e., attitudinal dilemmas) research, people have been asked to provide information on their subjective affective and cognitive experiences of ambivalence, and how uncertain they felt about the attitude object. Some types of measurement have tried to somewhat surpass the highly subjective nature of these types of self-report by asking people to separately rate positivity and negativity of an attitude object, subsequently calculating an ambivalence index that may be relatively more objective than the subjective, or 'felt' ambivalence self-report (Breckler, 1994; Kaplan, 1972). These measures have been translated to health behavior dilemmas, for example in research exploring the underlying processes of self-control. In a paper by Gillebaart, Schneider, and De Ridder (2016), a first attempt at investigating how self-control affects the health dilemma that people experience when being confronted with tasty, yet unhealthy snacks was made by simply asking people how conflicted, mixed, and indecisive they felt about the food items. People were also asked to provide a positivity rating about the food item, thereby ignoring the negative aspects that item may also hold, and vice versa a negativity rating that did not take any positive aspects into account. These ratings result in a polarity index that is thought to indicate how big the dilemma actually is (Kaplan, 1972; Priester & Petty, 1996). Interestingly, results demonstrated that people with a higher level of trait self-control showed lower ratings of conflictedness and a lower 'objective' index of conflict compared to people with a lower level of self-control. These results were in line with findings from a current study into threat and challenge appraisals that repeatedly demonstrated that people with higher levels of trait self-control considered self-control dilemmas more challenging and less threatening than people with lower levels of self-control (Gillebaart, Bogaers, & De Ridder, 2018). Although this line of research provided some insight into why people with higher levels of self-control are better able to handle health dilemmas (i.e., they report feeling less conflicted and less threatened), information about the process that led to the conscious self-report of feelings of conflict and challenge appraisals was lacking. The self-reports from these studies are a reflection of the outcome of a process in which the dilemma is noticed, identified, and resolved one way or the other. This entire process however is not reported on when people are asked about their feelings of conflict.

Integrating implicit measures

To get a better hold on the processes that take place outside of conscious awareness, implicit measures need to be incorporated into study designs. For example, in Gillebaart et al. (2016), an implicit measure was added to the design, by applying a 'mousetracking' paradigm. With mousetracking, people's hand movements are measured while they perform a choice or categorization task on a screen (Freeman & Ambady, 2010). These movements serve as a proxy for the processes that take place during the categorization or choice, and that are rarely tapped into by simply measuring the outcome or asking people about it. Mousetracking has been on the rise as a valuable tool for implicitly assessing all kinds of conflict, from attitudinal ambivalence (Buttlar & Whalther, 2018; Schneider & Schwarz, 2017) to self-control and self-regulation contexts (Lim, Penrod, Ha, Bruce, & Bruce, 2018; Lopez, Stillman, Healtherton, & Freeman, 2018), and to social and affective settings (Brambilla, Biella, & Freeman, 2018; Lazerus, Ingbretsen, Stolier, Freeman, Yamauchi & Xiao, 2017). In the case of Gillebaart et al. (2016), the mousetracking data showed a different pattern from the feelings of conflictedness explicitly reported by participants. Of course, the rich data from the mousetracking provided additional information about timing (i.e., response time, time of peak conflict), but also about the magnitude of the conflict. Interestingly, although people with a higher level of trait self-control reported feeling less conflicted on the explicit level, this pattern did not show up in the mousetracking data at all: no differences were found in conflict magnitude variables (i.e., 'area under the curve', 'maximum deviation') between people with higher and lower levels of self-control. The explicit and implicit measure thus diverged rather than converged. A similar divergent pattern of results was obtained when Gillebaart et al. (2018) conducted a study that measured the psychophysiological underpinnings of threat and challenge appraisals (i.e., cardiac output). Whereas self-reports showed clear differences in appraisals as a function of trait self-control, this pattern was absent from the implicit, psychophysiological measure.

Potential and pitfalls of combining explicit and implicit measures

These recent studies represent of course a small selection of an array of studies that combine explicit and implicit measures in the field of health behavior research. However, they do highlight how adding implicit measures can enrich our understanding of how people deal with health dilemmas. Specifically, they provide insight beyond self-reports, into the processes that take place before or while people are making a choice or decision. In the mousetracking example, the authors were able to demonstrate that the dilemma emerged similar in size for all participants, but that those with high self-control were able to resolve the dilemma faster, which may have translated into the differences observed in their self-reports (Gillebaart et al., 2016). Similarly, the fact that self-control did not predict any differences in psychophysiological preparation for conflict, while people with higher self-control did report to feel more challenged and less threatened, is informative with regards to the underpinnings of successful self-control. It may for instance mean that at the most basic level, dilemmas are experienced similarly for people with high and low self-control. However, in the process that take place from the emergence of the dilemma to resolution and explictly reporting on it, differences ensue between people with high and low self-control. These differences could be due to the ability to identify a conflict earlier on (as in the Gillebaart et al., 2016) study, or in the (pro-active) coping mechanisms (Aspinwall & Taylor, 1997) or situational strategies (Duckworth, Gendler, & Gross, 2016) that allow for appraisals of challenge over threat, which may be subject to individual differences. Adding the implicit psychophysiological measure to this study allowed for a more focused perspective on the underlying process, and adds to the understanding of the whole dilemma and how it is solved, instead of focusing only on the outcome.

A limitation of combining explicit and implicit measures is that whereas convergence between these measures is interpreted rather unambiguously, divergence between these measures is meaningful, but can also be a sign that either measure's validity is compromised. There has been extensive debate on whether and how implicit measures predict behavior. For example, the Implicit Association Test (Greenwald et al., 1998), arguably the most used implicit measure for assessing attitudes, stereotypes, and self-esteem, has been heavily criticized. It has been suggested that the measure is able to tap into nonconscious processes that are not accessible for explicit self-reports (Greenwald & Banaji, 1995), but there is also accumulating consensus on the idea that in fact, these nonconscious processes are accessible to people's introspection, but are suppressed in explicit self-reports due to factors like social desirability and cognitive elaboration (Fazio & Olsen, 2003; Hofmann, Gawronski, Le, & Schmitt, 2005). There is some meta-analytic evidence demonstrating that the IAT is able to predict behavior with a moderate effect size, and to a bigger extent as explicit self-reports, especially when it comes to topics sensitive to social desirability, suggesting to combine the two types of measures when seeking to predict behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), as explicit and implicit attitudes seem to be different yet related constructs (Nosek & Smyth, 2007). However, other meta-analytic evidence has indicated that these associations between the IAT and behavior were significantly overestimated and identified a number of methodological issues with how these associations are interpreted (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Importantly, this debate has led to agreement on the need for more research before the IAT can actually be used to predict people's behavior (Greenwald, Banaji, & Nosek, 2015).

Conclusion and future directions

The debate on the IAT illustrates the complexity of using implicit measures in psychological research. When it comes to measuring how people deal with health dilemmas, some similar issues will arise and will need to be addressed. Research on self-control measures has already identified these different measures may actually tap into different dimensions of the same construct, which affects con- or divergence between different measures (Duckworth & Kern, 2011). Furthermore, there are some indications that the time available for deliberation affects impulsive choices (e.g, when solving a health dilemma; Veling et al., 2017), which shows that tracking the process is of utmost importance. As such, there is promise in measurements like mousetracking, eyetracking, and similar measures that assess an ongoing, online process instead of simply an outcome. When selecting an implicit measure, it is thus advisable to think about the process that you are trying to tap into, and select a paradigm that able to provide you with this insight. As behavior, as well as dilemmas and choices, do not exist in a vacuum, measuring a process may be more useful in addition to an explicit measure of the outcome compared to measuring the outcome on another level. Of course, caution is needed when designing or adopting implicit measures into your design. Integrating theory and study results with investigations into the validity and robustness of the measures that are used in the field is one of the cornerstones of psychological research (Mischel, 2009). Moreover, we should not be discouraged by the theoretical and methodological intricacies of including implicit measures but rather experience this as a challenge that comes with the job. After all, as psychologists we are already well aware of how complex and opaque human behavior can be.

Statement of competing interests

The author states no competing interests.

Acknowledgements

The author would like to thank Rob Holland for his comments on an earlier draft of this manuscript.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Aspinwall, L. G., & Taylor, S. E. (1997). A stitch in time: Self-regulation and proactive coping. *Psychological Bulletin, 121*, 417–436.

Breckler, S. J. (1994). A comparison of numerical indexes for measuring attitude ambivalence. *Educational and Psychological Measurement, 54*, 350-365.

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthine. *Journal of Experimental Social Psychology, 78*, 34-42.

Buttlar, B., & Walther, E. (2018). Measuring the meat paradox: How ambivalence towards meat influences moral disengagement. *Appetite, 128*, 152-158.

Carver, C. S. (2005). Impulse and constraint: Perspectives from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review, 9*, 312–333.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*, 259-268.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science, 11*, 35-55.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Freeman, J.B. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226-241.

Friese, M., Hofmann, W., & Wiers, R. W. (2011). On taming horses and strengthening riders: Recent developments in research on interventions to improve self-control in health behaviors. *Self and Identity, 10*, 336–351.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.

Gillebaart, M., Bogaers, M., & De Ridder D.T.D. (2018). Self-control and appraisals of threat and challenge. *Unpublished manuscript.*

Gillebaart, M., Schneider, I. K., & De Ridder, D. T. D. (2016). Effects of trait self-control on response conflict about healthy and unhealthy food. *Journal of Personality, 84*, 789-798.

Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow (Vol. 1)*. New York: Farrar, Straus and Giroux.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553-561.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin, 77*, 361–372.

Lazerus, T., Ingbretsen, Z. A., Stolier, R. M., Freeman, J. B., Cikara, M. (2016). Positivity bias in judging ingroup members' emotional expressions. *Emotion, 16*, 1117-1125.

Lim, S. L., Penrod, M. T., Ha, O. R., Bruce, J. M., & Bruce, A. S. (2018). Calorie labeling promotes dietary self-control by shifting the temporal dynamics of health- and taste-attribution integration in overweight individuals. *Psychological Science, 29*, 447-462.

Lopez, R. B., Stillman, P. E., Heatherton, T. F., & Freeman, J. B. (2018). Minding one's reach (to eat): The promise of computer mouse-tracking to study self-regulation of eating. *Frontiers in*

nutrition, 5, 43.
Mischel, W. (2009). Becoming a cumulative science. *APS Observer, 22*.
Myrseth, K. O. R., & Fishbach, A. (2009). Self-control: A function of knowing when and how to exercise restraint. *Current Directions in Psychological Science, 18,* 247–252.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.
Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. *Experimental Psychology, 54*, 14-29.
Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171-192.
Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology, 71,* 431–449.
Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: the case of ambivalence. *Current Opinion in Behavioral Sciences, 15*, 39-45.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.
Veling, H., Chen, Z., Tombrock, M. C., Verpaalen, I. A., Schmitz, L., Dijksterhuis, A., & Holland, R. W. (2017). Training impulsive choices for healthy and sustainable foods. *Journal of Experimental Psychology: Applied, 23*, 204-215.
Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60,* 181-192.

Marleen Gillebaart
Social, Health, and Organizational Psychology, Utrecht University
M.Gillebaart@uu.nl