volume 14 issue 4
december | 2012

The European Health Psychologist

www.ehps.net/ehp

Bulletin of the European Health Psychology Society

ISSN: 2225-6962

original article

Full disclosure: doing behavioural science necessitates sharing

Gjalt-Jorn Y. Peters
Open University of the Netherlands

Charles Abraham
University of Exeter

Rik Crutzen
Maastricht University

The way we currently organise and report research retards behavioural science. This brief article provides an explanation of how we are holding back scientific progress, how this situation developed, and how current practice is justified. We also recommend practical, low-cost solutions, which would facilitate scientific advances.

Reasons for full disclosure: how we hold back scientific progress

When we complete a study, we tend to publish only the results. Researchers rarely publish the questionnaires, computer tasks, intervention manuals, study protocols and other materials that were used to generate these results. Similarly, readers do not usually have access to complete data sets, all statistical analyses undertaken or the commands (e.g. 'syntax' in SPSS) needed to replicate these analyses. In addition, researchers do not always secure these supplemental materials sufficiently, and they frequently get lost over time. This has three consequences each of which retard the establishment of a science of behaviour.

1. It is not possible to critically scrutinize what researchers do not divulge in an article so the empirical data that reported results represent remain unknown and it is difficult to judge whether analyses and interpretation are optimal or even correct;

2. Accurate replication of research is compromised and often impossible;

3. Data syntheses (e.g. meta-analyses) are impeded, as they frequently have to exclude studies because they do not have access to the original data and procedures, and so may be based on incorrect assumptions.

A brief explanation of why each of these consequences is undesirable follows.

Maximum scrutiny

There are two reasons to desire maximum scrutiny of our research. First, we can make mistakes in analyses and interpretation (Bakker & Wicherts, 2011; Glantz, 1980; Wicherts, Bakker, & Molenaar, 2011) for example, when using an analytical method for the first time. This is especially likely when analyses are undertaken by those still learning analytic techniques (including both qualitative and statistical analyses). Since reviewers might also be limited in their competence, researchers' errors frequently end up in published papers (Glantz, 1980; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Allowing further scrutiny beyond the pre-publication peer review process can teach us what we are doing wrong, thereby making us better researchers and reviewers (see also Walther & van den Bosch, 2012). Second, as pointed out by Simmons, Nelson, and Simonsohn (2011), researchers make a lot of choices that are not disclosed in research reports. These choices frequently favour significant p-values (Masicampo & Lalande, 2012). This is likely a consequence of the Significant Outcome Bias in

our literature (Peters, Ruiter, & Kok, 2012): non-significant findings are rarely submitted, and if submitted, rarely published. Greater disclosure of materials and analyses would allow detection of these biases.

Accurate replication

As was recently argued (Abraham, 2012a), scientific progress requires replications. Successful replications strengthen our evidence that a given theory or model holds, whereas unsuccessful replications can falsify previously held theories or models (see for example Milton & Wiseman, 1999 or Ritchie, Wiseman, & French, 2012). Without the ability to replicate others' work, we are not accumulating knowledge, but rather false positives: once an article makes a claim, we thwart the possibility to replicate the study and thereby challenge the claim, especially if required materials are unavailable. Publication bias plays a role here (some journals have an official policy to not publish replications; French, 2012), and this is a problem that needs to be addressed in its own right. Nonetheless, at least publishing all relevant materials would permit *accurate* replications. At present, debates regularly arise as to whether or not failures to replicate should be attributed to a mistaken interpretation of empirical reality or differences in materials or study procedures; witness, for example recent replication debates in social psychology (Doyen, Klein, Pichon, & Cleeremans, 2012; Ritchie et al., 2012). This inability to replicate and acknowledge replication failures renders behavioural scientists unable to use data to choose between better and worse models of reality and better and worse approaches to intervening in the real world (van Dongen, Abraham, Ruiter, & Veldhuizen, in press).

Data syntheses

When trying to synthesize the findings of multiple studies, either quantitatively or qualitatively, it is crucial to understand the initial data, otherwise syntheses may result in counting apples and oranges as bananas to the determent of behavioural science. For example, in the case of meta-analysis it is necessary to convert study results into effect size measures that use the same metric. For simple designs where means in two groups are compared, Cohen's *d* is the obvious choice; when relating two continuous variables, Pearson's *r* is often used. Published studies frequently do not report effect sizes; and it is quite common that studies have to be excluded from meta-analyses because it is not possible to compute the required statistics on the basis of the reported results. Even when willing (which not all researchers are, see Wicherts et al., 2011), authors are often unable to send meta-analysts their datasets. Publishing dataset files, as well as the commands for the analyses that were used (e.g. syntax files in SPSS, scripts in R), along with the articles would considerably enhance the quality of evidence syntheses.

Given these clear advantages, one may wonder why we have resisted full disclosure. Surprisingly, there are very good reasons why this convention, to only publish results and not data, analyses and materials, developed.

How non-disclosure developed and current views

On 6 March 1665, the first purely scientific journal was published (Oldenburg, 1665), to enable (more or less) efficient communication between researchers. Since then, many such journals have been published. Journals initially faced serious limitations: production and distribution of physical journals was expensive, so contributions had to be brief and appendices were rare. Because of this, the address of at least one of the authors was always included, to enable researchers to request supplementary materials such as questionnaires that were used,

more detailed study protocols, and more recently, computer programs and algorithms.

Since establishment of the internet, most journals are published online (some exclusively), archiving articles and supplementary materials on servers. These materials can be accessed from anywhere, and the costs of storing material on a server are negligible[1]. Online-only journals are therefore relieved of the constraint of limiting the number of pages in an article, and all journals are relieved of the constraint of limiting the supplements. Therefore, it has been possible for a number of years to publish not only the report describing your methods and results, but also everything used in your study. However, despite the disappearance of practical and technological constraints, reporting practices in behavioural science remain largely unchanged. This could be remedied by the editors of leading journals; publishing only articles representing full online disclosure would change practice rapidly.

Because these constraints have been lifted, currently, there have been a number of pleas for full(er) disclosure. On the one hand, there were pleas for Open Access publication of articles reporting research funded with public money (Ghosh, 2012), but there have also been efforts to promote full disclosure of data and materials. For example, recently a consensus statement was developed by a group of health psychology researchers and journal editors urging editors to adopt a full disclosure policy in relation to behaviour change intervention development and design. The Workgroup for Intervention Development and Evaluation Research (WIDER) made four recommendations which can be summarized as: (1) provide detailed intervention descriptions, including (2) descriptions of control groups including usual care, (3) describe the intervention development process in detail and in relation to postulated change processes and (4) provide intervention delivery manuals that enable accurate replication (see e.g., Abraham, 2012b). Yet, despite the minimal costs, too few researchers routinely publish all their materials, data, analyses scripts and output, listing a variety of reasons.

Reasons for non-disclosure

In informal discussions, people have listed a variety of reasons for their reluctance to publish everything. The main ones are discussed below.

I may want to use my data again

The APA (American Psychological Association) requires that "psychologists do not withhold the data on which their conclusions are based" (APA, 2010, p. 12). However, when a researcher wants to publish several articles about one dataset it can be risky to publish the dataset before the work is completed. After all, somebody else might beat you to it—quickly publishing ideas arising from your own data. Luckily, there is a straightforward solution that negates the concern: do not publish the entire

[1]For example, a convenience sample of the Portable Document Format (PDF) files of 20 publications from 2012 shows that the average article has 21.50 pages (median = 13.50, sd = 26.86), is 594.80 kilobytes (median = 366.50, sd = 576.80), and that the average number of kilobytes for one page is 41.62 (median = 27.07, sd = 41.60). When saved as plain textfiles, a datafile with around 20 variables and 500 participants will be around 860 kilobytes if the variables are saved with 15 significant digits. Since most articles report far less data, it is safe to say that on average, one article plus supplemental materials can be easily stored in around 5 megabytes. As an example of how additional materials can be published, the datafile, R commands and output for these analyses are available online at http://sciencerep.org/1. Hosting prices vary, but plans exist where 1 gigabyte costs around €10 per year. This means that it is possible to host 1000 articles for around 50 euro per year.

dataset, but only those variables you described in your article. All statistical programs allow you to specify which variables should be 'kept' or 'dropped' when saving. This approach has two risks. First, it enables cherry picking of the data: researchers can omit variables that behave inconsistently with their hypotheses. This risk is addressed by the requirement to also publish the used study protocol, materials and questionnaires: these provide reviewers and other researchers with an overview of all studied variables. The second risk is that researchers conduct many trials, only publishing the data of those with results that fit their hypotheses, but not of the preceding trials with less fortunate results (the 'pre-replications', or preplications). This risk can be addressed by requiring that not only all data and materials relating to the published study are provided, but also all data and materials relating to previous preplications.

I am helping the competition if they have my materials

Sometimes, researchers are 'racing' each other to get results published first. At first glance, it seems as if you help the competition by publishing the materials (study protocol, computer tasks, questionnaires), because they no longer have to develop their materials themselves, which of course considerably speeds up their progress. However, materials would only be published when the article itself is published; so any race has already been won—or lost. This also resolves the potential problem where a researcher might be worried that a peer reviewer is a competitor and might abuse their 'preview access'. Of course, ideally, reviewers do have access to the materials, data, analysis scripts and output, as this would enhance the quality of the reviewing process. A solution could be to let reviewers sign an agreement to not use resources of reviewed articles until publication.

I want to sell my materials commercially

Researchers sometimes want to use their research to make money, for example by selling questionnaires or intervention manuals. Of course, when research is funded by public money, the results belong in the public domain, as taxpayers pay for the development of scientific materials and should therefore have access to the fruits of their initial investment. However, when research is conducted by commercial companies then they own their results and findings. We suggest that a clear line is drawn between scientific journals which share results and data for the advancement of science and so follow a full disclosure policy and commercial journals which do not require full disclosure because researchers publishing in them are seeking financial gain from work they own.

I don't want others to earn money with my hard work

Publishing your materials, protocols, data, analysis scripts, and output so that it is openly accessible, does not necessarily mean that everybody can use these resources however they please. A very simple way of determining which rights are provided is by using the Creative Commons licences. This not-for-profit organisation offers six licences, allowing researchers to determine whether they want others to be able to use a resource commercially or not, whether they want to be credited when the resource is used, and whether 'derivative works', works drawing on the original resource, have to be published under the same license ('share-alike'). Thus, it is easy to prohibit commercial use of research resources, securing citations, and 'paying it forward': making sure that others also share their work.

I am afraid errors are pointed out in my work

If a dataset and analysis protocol (preferably even the script file, e.g. an SPSS syntax file) is published, this makes a researcher vulnerable to others identifying errors in their methods and analyses. Few people like being told they were wrong. On the other hand, even less researchers would argue that science should be hampered to avoid researchers' or journals' embarrassment. A full disclosure behavioural science would facilitate collective acknowledgement that identification of mistakes in methodology and analyses are crucial to increasing accuracy in interpretation of data, thereby ensuring that mistakes do not retard scientific development; witness recent work in physics on faster than light neutrinos (Reich, 2011, 2012). Scientific data is not flawless. Falsification is central to scientific progress—the assumption that we will make mistakes is a basic assumption of our work. This is already reflected in letters to editors which allow researchers to react on papers (e.g. James & Smyth, 2012), and authors the possibility of reacting to such reactions (e.g. Crutzen, 2012).

My ethical committee requires me to use an informed consent where participants explicitly only provide permission for use of their data for my particular study, precluding re-analysis.

If a researcher is explicitly disallowed from publishing data because of a contract or because permission from an ethical committee is conditional upon non-publication of the data, little can be done. However, such arrangements are very rare; most legislators in fact encourage publication or research data. For example, the code of conduct for using personal data in scientific research that was developed by the Dutch universities based on the relevant Dutch legislation explicitly states "[...] in scientific research, the use of a previously created datafile is allowed, also if the file was created for another reason, unless the file contains identifying data [...]" (VSNU, 2005). Thus, the practice of removing all identifying information from a datafile, which is already common practice for most researchers, suffices to enable publication and re-use of the datafile in most situations. Note that in any case, restrictions regarding publication of datafiles do not extend to publication of materials, study protocols, analysis scripts, and output files.

Guidelines for a Full Disclosure Science

Publishers already facilitate publication of supplementary materials, so all that is required is a change in authors' motivation which could be brought about by changes in editorial policy. A number of guidelines to optimize the benefits from publishing such supplemental materials follow, phrased as journal policy suggestions.

1. Require researchers to supply:

a. Everything necessary for replication (e.g., questionnaires, source code of computer tasks, or at least compiled tasks, detailed protocols, manual etc.);

b. For quantitative research, a datafile containing all variables involved in the analyses that are reported, and for qualitative research, the coding tree, and ideally, the sources with their codes;

c. A document detailing the analyses, which, together with the datafile, must enable accurate reproduction of the reported results (ideally, a script with the commands used, such as a syntax file for SPSS);

d. A record of all data collected relevant to the reported analyses e.g., datasets from preliminary datasets not included in the analyses;

e. The output used by the researcher on which the reported results were based (although this output should be replicable perfectly using the datafile and the analysis script, not all researchers use scripts for their analyses; in addition, such replication requires access to the same software, and researchers often use different proprietary software packages such as SPSS, SAS and STATA).

2. Require that these files be supplied in non-proprietary formats (this is important because opening proprietary formats require the purchase of specific software, which other researchers may not have). This means that:

a. Resources that cannot be provided in the preferred format, are provided in one of the default non-proprietary formats, such as plain text, Open Document Format (ODF), Portable Document Format (PDF), or Hyper Text Markup Language files (HTML), or, for images, Portable Network Graphics (PNG) or Scalable Vector Graphics (SVG);

b. Questionnaires and computer tasks are preferably provided in a format that can be imported into free non-proprietary software, for example LimeSurvey for questionnaires (LimeSurvey Project Team/Carsten Schmitz, 2012) and OpenSesame for computer tasks (Mathôt, Schreij, & Theeuwes, 2012). Stimuli and intervention materials are provided in the default formats (see 2a);

c. Quantitative datafiles are provided in generic data formats such as Tab or Comma Separated Values, and qualitative data in one of the default formats (see 2a);

d. Files with commands for statistical programs (e.g. SPSS, SAS, R) are provided in their original format, because these generally already are in plain text format. When a program is used that does not store the analysis script in plain text, the researcher can copy-paste to one of the default formats (see 2a). Note that of course, using free non-proprietary programs such as R (R Development Core Team, 2012) is preferred, as all other researchers have access to this software;

e. The output can be provided in one of the default formats (see 2a).

These guidelines entail minimal efforts (and virtually no costs) from both journal editors and authors. However, the benefits are substantial: our evidence will become more accurate through correction of errors, replication, and much higher quality meta-analyses, and in addition, developing oneself as researcher will be much easier, which will benefit ourselves and our students once we involve these supplementary materials in our courses. Finally, it is likely that the mere requirement of publication of materials, data and analyses will already have a beneficial effect on the quality of our evidence base, as it has been shown that willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results (Wicherts et al., 2011). It is up to all of us to change and share to accelerate science, starting with our own sub-discipline; health psychology. Reactions from the editors of *Psychology & Health* and *Health Psychology Review* are more than welcome!

Acknowledgements

We would like to thank Gerjo Kok and Emily de Vet for feedback on earlier drafts of this text, and Catherine Bolman for advice on ethical aspects of data sharing.

References

Abraham, C. (2012a). Creating change that counts: Evidence-led co-creation as a pathway to impact. *Psychology & Health, 27(S1)*, 1.

Abraham, C. (2012b). Designing and evaluating interventions to change health-related behaviour patterns. *Randomized Clinical Trials of Nonpharmacologic Treatments* (pp. 357–368). London: Chapman & Hall/CRC.

American Psychological Association. (2010). Ethical principles of psychologists and code of conduct. *The American psychologist* (Vol. 57). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20642307

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*, 666–78. doi:10.3758/s13428-011-0089-5

Crutzen, R. (2012). Author response: associations and mixed findings. *Mental Health and Substance Use, 5*, 271.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral Priming: It's All in the Mind, but Whose Mind? (J. Lauwereyns, Ed.) *PLoS ONE, 7*, e29081. doi:10.1371/journal.pone.0029081

French, C. (2012, March 15). Precognition studies and the curse of the failed replications. *The Guardian*. London. Retrieved from http://www.guardian.co.uk/science/2012/mar/15/precognition-studies-curse-failed-replications

Ghosh, P. (2012). Report calls on government to back open access science. *BBC News: Science & Environment*. Retrieved November 7, 2012, from http://www.bbc.co.uk/news/science-environment-18497083

Glantz, S. A. (1980). Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation, 61*, 1–7. doi:10.1161/01.CIR.61.1.1

James, P., & Smyth, B. P. (2012). Commentary on Crutzen et al. (2012), Negative associations between primary school children's perception of being allowed to drink at home and alcohol use, *Mental Health and Substance Use, 5*, 64–69.

LimeSurvey Project Team/Carsten Schmitz. (2012). *LimeSurvey: An Open Source survey tool*. Hamburg, Germany. Retrieved from http://limesurvey.org

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly journal of experimental psychology*, 37–41. doi:10.1080/17470218.2012.711335

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*, 314–24. doi:10.3758/s13428-011-0168-7

Milton, J., & Wiseman, R. (1999). Does Psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*, 387–91.

Oldenburg, H. (1665). Epistle Dedicatory. *Philosophical Transactions of the Royal Society of London, 1*(1-22). doi:10.1098/rstl.1665.0001

Peters, G.-J. Y., Ruiter, R. A. C., & Kok, G. (2012). Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychology Review*. doi:10.1080/17437199.2012.703527

R Development Core Team. (2012). *R: A language and environment for statistical computing. Vienna, Austria*. Retrieved from http://www.r-project.org/

Reich, E. S. (2011). Faster-than-light neutrinos face time trial. *Nature.com news*. Retrieved October 12, 2012, from http://www.nature.com/news/2011/111005/full/news.2011.575.html

Reich, E. S. (2012). Leaders of faster-than-light neutrino team resign. *Nature.com news*. Retrieved October 12, 2012, from http://blogs.nature.com/news/2012/03/leaders-of-faster-than-light-neutrino-team-resign.html

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: three unsuccessful attempts to replicate Bem's "retroactive facilitation of recall" effect. *PloS one, 7*, e33423. doi:10.1371/journal.pone.0033423

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–66. doi:10.1177/0956797611417632

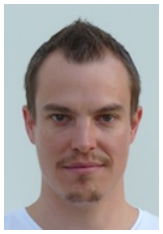
VSNU. (2005). *Gedragscode voor het gebruik van persoongegevens in wetenschappelijk onderzoek* [Code of conduct for the use of personal data in scientific research]. Den Haag, Netherlands.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology, 100*, 426–32. doi:10.1037/a0022790

Walther, A., & van den Bosch, J. J. F. (2012). FOSE: a framework for open science evaluation. *Frontiers in Computational Neuroscience, 6*, 1–8. doi:10.3389/fncom.2012.00032

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS one, 6*, e26828. doi:10.1371/journal.pone.0026828

van Dongen, A., Abraham, C., Ruiter, R. A., & Veldhuizen, I. (in press). Does questionnaire distribution promote blood donation? An investigation of question-behavior effects. *Annals of Behavioural Medicine.*



Gjalt-Jorn Peters

is Assistant Professor at the Faculty of Psychology, Open University, the Netherlands

gjalt-jorn@behaviorchange.eu

Charles Abraham

is Professor of Behaviour Change at the Peninsula College of Medicine & Dentistry, Universities of Exeter & Plymouth, UK.

Rik Crutzen

is a post-doctoral researcher at the Department of Health Promotion, Maastricht University /CAPHRI, the Netherlands

original article

Exploiting mistakes as learning opportunities to improve patient safety

Florina Spanu
Babeș-Bolyai University

Mistakes are a reality in healthcare systems as in any other industry. In the 2000, the Institute of Medicine report "To Err is Human" estimated that 44,000-98,000 US patients die every year in hospitals as a result of medical errors (Kohn, Corrigan & Donaldson, 2000). Recent statistics from Latin America countries indicate that 10% of patients receiving medical care suffer some kind of unwanted consequences related directly to the medical act; the numbers go up to 20% for inpatients (WHO, 2011). Although capturing the magnitude of healthcare casualties is difficult because they are spread temporally and geographically (Leape, 1994), it is estimated that more people die annually as a result of medical care than in car accidents and plane crashes, or from breast cancer and AIDS (Kohn et al., 2000; Berwick & Leape, 1999).

Adverse events that happen in hospitals are preventable (NHS, An Organisation with a Memory, 2000), provided that hospitals develop their capacity to exploit past experiences as learning opportunities (Aspden, Corrigan, Wolcott & Erikson, 2004; Aspden, Wolcott, Bootman & Cronenwett, 2006; Edwards, 2012). To date, healthcare systems in countries such as the USA, Australia and the UK have launched initiatives for developing hospitals' capacity to exploit their and other organizations' experience in order to deliver safer and more reliable medical care. Several priorities were set, such as implementing error reporting systems, designing organizational structures capable of sustaining change, empowering patients, developing teamwork abilities, or developing a blame-free culture in which one feels safe to acknowledge and discuss medical errors and mishaps (Committee on Quality of Health Care in America, IOM, 2001; NHS, 2000; Australian Commission on Safety and Quality in Healthcare, 2010; WHO, 2012). Although important advances have been made, especially in terms of detecting errors, the success of these initiatives has been mixed and pace of change is still far from matching the initial objectives (Leape & Berwick, 2005).

System-wise instruments and local mechanisms for exploiting mistakes as learning opportunities

Health care systems have drawn on the experience of high risk industries and adopted centralized error reporting systems. Anonymous error reporting systems facilitate error detection in order to analyze their underlying causes and prevent them from happening in the future (Carroll & Edmondson, 2002; Leape, 1994; Hudson, 2003; Mahajan, 2010; Cohen, 2000). Implementing this kind of formal collective learning mechanisms was an important breakthrough in the efforts to improve patients' safety and increase quality of care, as it allowed a better estimation of the magnitude of the medical error phenomenon (Brennan & Safran, 2004), and it led to the development of standardized protocols and procedures (Australian Commission on Safety and Quality in Healthcare, 2010). However, implementing error reporting systems as learning instruments is not always easy, as medical professional culture can represent a significant hindrance to

implementation. For example, qualitative research has found that health professionals are rather reluctant to adopt these kind of systems (Iedema, Allen, Sorensen & Gallagher, 2011; Waring, 2005). Physicians have been reluctant to embrace such initiatives due to the fear of being blamed, the lack of trust in the utility of it, and the belief that it is an extra administrative burden in their already busy agenda (Waring, 2005).

Although the role of error reporting systems cannot be underestimated, empirical data suggest that error reporting depends on actual error rates, but more importantly on the willingness to report them—which is highly dependent on the work interpersonal climate (Edmondson, 1996). Conducting a mixed methods study investigating factors influencing error reporting rates in nurses units, Edmondson (1996) found that high error rates were reported in units in which nurses felt that they trusted and respected each other, and that if they were to admit making a mistake, they would not be judged or rejected by colleagues. Using interviews and observation, the author found that in units in which nurses did not share such a high quality interpersonal climate, they were more likely to not report errors when they happened. These results support the hypothesis that error reporting is actually a function of actual error rates and the willingness to engage in error reporting. Organizational behavior research suggests that capitalizing on past experiences in order to improve future performance is a rather local and team/unit phenomenon, and not an organizational-wide one per se (Edmondson, 1999; 2002; Lipshitz & Popper, 2000). Health professionals reflect on their activity, and use it as a source for improving future performance, but lessons learned tend not to cross the boundaries of the department (Lipshiptz & Popper, 2000); and error rates were found to be smaller in nurses units in which all the members of the unit were involved in all stages of error reporting, error analyzing, identifying solutions to avoid them in the future and implementing the solutions, as opposed to when different stages were the responsibility of different members of the organization (Drach-Zahary & Pud, 2010). These findings stress the role unit-level practices and team climate play in exploiting errors as learning opportunities.

Hospitals as organizations have often been described as having a culture dominated by blame, fear and defensiveness, that blocks open communication and has a negative impact on health professionals' willingness to engage in error acknowledgement and analysis (Kohn et al., 2000; Berwick & Leape, 1999; Collins, Block, Arnold & Christakis, 2009; Catino, 2009; Iedema, Jorm, Braithwaite, Travaglia & Lum, 2006; Iedema et al., 2011; Waring, 2005). Admitting a mistake is not easy in any industry, but it is particularly difficult when the smallest error can have catastrophic implications for patients' life and health professionals' careers. Admitting one's own mistakes can easily be interpreted as incompetence or professional insecurity (Edmondson, 2004), while bringing up for discussion someone else's mistakes can be taken as lack of collegiality (Leape, 2006). The result is a professional environment with low tolerance for errors, or at least for open discussion of errors (Waring, 2005). Congruently, medical schools train highly independent health professionals, who are capable of making decisions on their own under time pressure and in emotionally demanding situations (Hoff, Pohl & Bartfield, 2006). Doctors in particular are educated in a rather individualistic spirit and are taught to rely only on themselves (Waring, Harrison & McDonald, 2007). This favors a culture of mistrust and blame, and leads to a working environment that lacks transparency and the capacity for collaboration (Leape et al.,

2009). A culture of fear and blame adds to a highly hierarchical organizational structure in which nurses and residents often report feeling uncomfortable to openly address physicians (Reeves et al., 2009), even in matters that are directly related to patients' well-being (Edmondson, 2003). Establishing an open, defensive-free communication environment is key point in developing an organization's capacity to capitalize on its failures as assets for future improvement (Argyris, 2000; Senge, 1994; Carroll & Edmondson, 2002). Field research suggests that developing these kinds of working environments is more likely to be successful when addressed at the department level, as opposed to when they address the entire organization (Edmondson, 1999; 2002).

Facilitating learning from failures as a social, interpersonal phenomenon

Several characteristics have been identified as being particularly relevant for developing medical departments' capacity to exploit their past experiences as learning opportunities such as a coaching oriented leadership style, a psychological safe unit climate, and empowering low status health professionals (Edmondson, 1996; 2003; Lipshitz & Popper, 2000; Edmondson, Bohmer & Pisano, 2001, Nembhard & Edmondson, 2006; Hirak, Peng, Carmeli, & Schaubroeck, 2012; Tucker, 2007; Waring et al., 2007). Leaders can either facilitate or block collective learning from failures. The relationship was found to be consistent in nurse units (Edmondson, 1996), mixed surgical units (Edmondson, 2003), and medical departments as a whole (Lipshitz & Popper, 2000; Nembhard & Edmondson, 2006; Hirak et al., 2012). Leaders who are open to admitting their own mistakes, who encourage members to speak openly, who are problem-focused, as opposed to blame-focused, when they come to know about an error, who acknowledge the contribution of all team members irrespective of their organizational position, and model feedback asking and feedback giving, were found to increase health professionals' willingness to engaging in open communication about their work and the problem they encounter on a day to day basis. One of the mechanisms through which they manage to do this is by creating a psychological safe climate, in which unit members feel that they can bring up for discussion sensitive issues. Openly admitting medical errors and failures exposes to criticism (Edmondson, 1999; 2004). In order to speak honestly about it, people have to trust that they will not be judged, and that they will be helped to manage the problem and its implications. Research has shown that team-leaders are key facilitators of such an interpersonal climate in medical organizations (Edmondson, 1996; Nembhard & Edmondson, 2006; Hirak et al., 2012). A key feature that distinguishes medical organizations is the pronounced power imbalance between different professional categories. Nurses, for example, perceived themselves as being less entitled to address doctors, but not the other way around (Reeves et al., 2009). Nurses were found to admit that they censure themselves, and do not share valuable information, or do not confront doctors even when they might think that a mistake is being made (Edmondson, 2003; Edmondson et al., 2001; Waring et al., 2007). People tend to be very accurate evaluators of their status within a group, and they rarely engage on their own in status self-enhancement (Anderson, Srivastava, Beer, Spataro & Chatman, 2006). For this reason, those that hold power within the organization can facilitate all members' participation in voicing problems in order to improve future performance (Bunderson & Reagans, 2010).

Conclusions

Improving patients' safety and quality of care is a priority for healthcare systems around the world. Important progress has been made by

implementing error reporting systems, but the availability of these instruments does not guarantee that health professionals will use it, or that lessons learned from it will be implemented. System level, centralized solutions are not always easily embraced by medical personnel, and they sometimes find ways to work around them (Iedema et al., 2006; Iedema et al., 2011). Research on collective learning found that exploiting past experiences as learning opportunities is a rather local, interpersonal phenomenon, and not an organization-wide one. This suggests that capitalizing on errors in the medical system might benefit if it is conceptualized as an informal, department-level process.

References

Anderson, C., Srivastava, S., Beer, J., Spataro, S.E., and Chatman, J.A. (2006). Knowing your place: Self-perceptions of status in social groups. *Journal of Personality and Social Psychology, 9*, 1094-1110. doi: 10.1037/0022-3514.91.6.1094

Argyris, C. (2000). On Organizational Learning (2nd Edition). Malden, MA: Blackwell.Aspden, P., Corrigan, J.M., Wolcott, J., and Erikson, S.M. (2004). *Patient Safety: Achieving a New Standard for Care*. Washington, DC: The National Academies Press.

Aspden, P., Wolcott, J., Bootman, J.L., & Cronenwett, L.R. (2006). *Preventing Medication Errors: Quality Chasm Series* (1st edition). Washington, DC: The National Academies Press.

Australian Committee on Safety and Quality in Healthcare (2010). *Australian Safety and Quality Framework for Health Care*. Available at: http://www.safetyandquality. gov.au/wp-content/uploads/2012/04/Australian-SandQ-Framework1.pdf

Berwick, D.M., & Leape, L.L. (1999). Reducing errors in medicine. *Quality in Health Care, 8*, 145-146. doi: 10.1136/qshc.8.3.145

Brennan, P.F. & Safran, C. (2004). Patient safety: remember who it's really for! *International Journal of Medical Informatics, 73*, 547-550. doi: 10.1016/j.ijmedinf. 2004.05.005

Bunderson, J.S., & Reagans, R.E. (2010). Power, status, and learning in organizations. *Organization Science, 22*, 1182-1194. doi: 10.1287/orsc.1100.0590

Carroll, J.S., & Edmondson, A.C. (2002). Leading organizational learning in health care. *Quality in Health Care, 11*, 51-56. doi: 10.1136/qhc.11.1.51

Catino, M. (2009). Blame culture and defensive medicine. *Cognition, Technology & Work, 11*, 245-253. doi: 10.1007/s10111-009-0130-y

Cohen, M.R. (2000). Why error reporting systems should be voluntary. *British Medical Journal, 320*, 728-729. doi: 10.1136/bmj.320.7237.728

Collins, M.E., Block, S.D., Arnold, R.M., & Christakis, N.A. (2009). On the prospects for a blame-free medical culture. *Social Science and Medicine. 69*, 1287-1290. doi: 10.1016/j.socscimed.2009.08.033

Committee on the Quality of Health Care in America (2001). *Crossing the quality chasm: a new health system for the 21st century*. National Academy Press, Washington DC. Available at http://www.nap.edu/ catalog/10027.html.

Department of Health (2000). *An organization with a memory. Report of an expert group on learning from adverse events in the National Health Service*. Available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4065086.pdf.

Drach-Zahavy, A., Pud, D. (2010). Learning mechanisms to limit medication administration errors. *Journal of Advanced Nursing, 66*, 794-805. doi: 10.1111/j.1365-2648.2010.05294.x

Edmondson, A.C. (1996). Learning from mistakes is easier said than done: group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science, 32*, 5-28. doi: 10.1177/ 0021886396321001

Edmondson, A.C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350-383. doi: 10.2307/2666999

Edmondson, A. C. (2002). The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science, 13*, 128-146. doi: 10.1287/orsc.13.2.128.530

Edmondson, A.C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies, 40*, 1419-1452. doi: 10.1111/1467-6486.00386

Edmondson, A. C. (2004). Psychological safety, trust and learning: a group-level lens. In Kramer, R. & Cook, K. (Eds.) *Trust and Distrust in Organizations: Dilemmas and Approaches*, (pp. 239-272). New York: Russell Sage.

Edmondson, A.C., Bohmer, R.M., & Pisano, G.P. (2001). Disrupted

routines: team learning and new technology implementation in hospitals. *Administrative Science Quarterly, 46*, 685-716. doi: 10.2307/3094828

Edwards, N (2012). *Improving Hospitals and Healthcare Delivery Systems. A report on the priorities for strengthening the hospital and health care delivery systems in the WHO European Region*. World Health Organization.

Hirak, R. Peng, A. C., Carmeli, A., & Schaubroeck, J. M. (2012). Linking leader inclusiveness to work unit performance: The importance of psychological safety and learning from failures. *The Leadership Quarterly, 23*, 107-117. doi: 10.1016/j.leaqua.2011.11.009

Hoff, T.J., Pohl, H., Bartfield, J. (2006). Teaching but not learning: how medical residency programs handle errors. *Organizational Behaviour, 27*, 869-896. doi: 10.1002/job.395

Hudson, P.T.W. (2003). Applying the lessons of high-risk industries to medicine. *Quality and Safety in Health Care, 12*, 7-12. doi: 10.1136/qhc.12.suppl_1.i7

Iedema, R., Allen, S., Sorensen, R., & Gallagher, T.H. (2011). What prevents incident disclosure, and what can be done to promote it? *Joint Commission Journal on Quality and Patient Safety, 37*, 409-417.

Iedema, R., Jorm, C.M., Braithwaite, J., Travaglia, J., & Lum, M. (2006). A root cause analysis of clinical error: Confronting the disjunction between formal rules and situated practice. *Social Science & Medicine, 63*, 1201-1212. doi: 10.1016/j.socscimed.2006.03.035

Kohn, L.T., Corrigan, J.M., & Donaldson, M.S. (2000). *To Err Is Human: Building a Safer Health System*. Washington: National Academy Press.

Leape, L.L. (1994). Error in medicine. *Journal of the American Medical Association, 272*, 1851-1857. doi: jama.1994.03520230061039

Leape, L.L. (2006). Full disclosure and apology—an idea whose time has come. *Physician Executive Journal, 32*, 16-18.

Leape, L.L., & Berwick, D.M. (2005). Five years after to err is human: what have we learned? *Journal of the American Medical Association, 293*, 2384-2390. doi: 10.1001/jama.293.19.2384

Leape, L.L., Berwick, D., Clancy, C., Conway, J., Gluck, P., Guest, J., et al. (2009). Transforming healthcare: A safety imperative. *Quality and Safety in Health Care, 18*, 424-428. doi: 10.1136/qshc.2009.036954

Lipshitz, R., & Popper, M. (2000). Organizational learning in a hospital. *Journal of Applied Behavioral Science, 36*, 345-361. doi: 10.1177/0021886300363005

Mahajan, R.P. (2010). Critical incident reporting and learning. *British Journal of Anaesthesia, 105*, 69-75. doi: 10.1093/bja/aeq133

Nembhard, I.M. & Edmondson, A.C. (2006). Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior, 27*, 941–966. doi: 10.1002/job.413

Reeves, S., Rice, K., Gotlib Conn, L., Miller, K. L., Kenaszchuk, C., & Zwarenstein, M. (2009) Interprofessional interaction, negotiation and non-negotiation on general internal medicine wards. *Journal of Interprofessional Care, 23*, 633-645. doi: 10.3109/13561820902886295

Senge, P.M. (1994). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Knopf Doubleday Publishing Group.

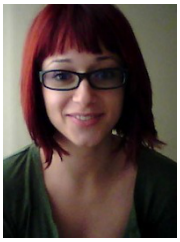
Tucker, A.L. (2007). An empirical study of system improvement by frontline employees in hospital units. *Manufacturing and Service Operations Management, 9*, 492-505. doi: 10.1287/msom.1060.0156

Waring, J.J., Harrison, S.R., & McDonald, E.R. (2007). A culture of safety or coping? Ritualistic behaviours in the operating theatre. *Journal of Health Services Research and Policy, 12*, 3-9. doi: 10.1258/135581907780318347

Waring, J.J. (2005). Patient safety: new directions in the management of health service quality. *Policy & Politics, 33*, 675-692. doi: 10.1332/030557305774329145

World Health Organization (2011). *IBEAS: a pioneer study on patient safety in Latin America. Towards safer hospital care.* WHO Document Production Services, Geneva, Switzerland.

World Health Organization (2012). World Alliance for Patient Safety. Available at http://www.who.int/patientsafety/worldalliance/en/ - retrieved on 20 September 2012.



Florina Spanu
is a PhD student at the Department of Psychology at the Babeș-Bolyai University in Cluj-Napoca, Romania
florinaspanu@psychology.ro

EHPS conference 2012

Delegate Feedback on the 2012 EHPS Conference, Prague, Czech Republic

Paul Norman
Conference Officer, Past President

The 2012 EHPS Conference took place in Prague (21-25 August 2012) and was attended by 814 delegates. An online conference evaluation survey was sent to all delegates, of whom 373 (46%) completed the survey. Of the delegates who completed the survey, the highest numbers were from the UK (n=74), The Netherlands (n=39) and Germany (n=29), which broadly reflects the profile of EHPS members and conference delegates. For 39% of respondents, this was their first conference, although a similar number of respondents had attended at least 3 EHPS conferences in the past 5 years (38%).

As shown in Table 1, respondents' overall ratings of the scientific programme were high, with the exception of the rating for the poster presentations which was lower, but still positive. These ratings were reflected in respondents' comments on the conference.

Table 1. *Scientific Programme: Overall Ratings (1=Poor to 5=Excellent)*

	M	*SD*
Overall quality	4.20	0.64
Keynotes	4.34	0.74
Symposia	4.26	0.65
Oral presentations	4.05	0.64
Roundtables/debates	4.11	0.79
Chairing	4.36	0.70
Pre-conference workshops	4.44	0.91
Poster presentations	3.82	0.91

"The keynotes were really inspiring and the quality of the talks and posters was outstanding"

"Really good conference, but poster sessions not working"

Table 2. *Balance of Sessions in the Scientific Programme*

	Too Few	*Fine*	*Too many*
Symposia	7.90%	88.00%	4.10%
Workshops	10.20%	88.40%	1.40%
Keynotes	11.10%	84.60%	4.30%

In terms of the balance of sessions in the scientific programme, the vast majority of respondents (>80%) were happy with the numbers of symposia, workshops, keynotes and oral presentations. However, 39.0% of respondents felt that there were too many poster presentations, which may reflect the experience of many delegates when trying to view the posters and/or listen to the short presentations. This was reflected in many delegates' comments on the poster sessions.

"There were too many poster presentations at the same time—it is usually very difficult to hear what the presenter is saying"

"The poster area was too noisy and crowded for this to work properly"

When asked whether they would prefer to have poster sessions with or without short presentations, just over half (50.5%) indicated that they would prefer poster sessions with short presentations, with 32.3% indicating that they would prefer poster sessions without short presentations and 17.3% undecided. Delegates' comments indicated that many like the interactive poster sessions, especially as it gives new researchers an opportunity to present their work, but also that the sessions are too crowded to work well.

"The interactive format provides a chance for everyone, including early career researchers, to present their work"

"I like the format of the poster presentations but felt the venue was too small for it to be done properly"

Table 3. *Specific Aspects of the Conference Programme (1=Poor to 5=Excellent)*

	M	*SD*
Good quality research	4.14	0.69
Range of theoretical approaches	4.08	0.80
Theory-based interventions	4.09	0.77
New/yet to be published research	4.04	0.81
Range of methods	3.84	0.87
Relevant to clinical practice	3.80	0.79
Relevant a health psychologist's work	3.91	0.80

Respondents' ratings of aspects of the conference were generally positive (Table 3), although respondents felt that the programme was less successful as regards to including papers that had a range of methods, were relevant to clinical practice and address issues relevant to all aspects of a health psychologist's work.

Respondents' ratings of other aspects of the conference were generally positive (Table 4). Overall, respondents gave very high ratings for the overall time schedule of the conference as well opportunities to meet and talk with colleagues. Other aspects of the conference such as the social events received slightly lower, but still positive, ratings. These ratings were also reflected in delegates' comments.

Table 4. *Various Aspects of the Conference (1=Poor to 5=Excellent)*

	M	*SD*
Overall time schedule	4.26	0.71
Venue	3.95	1.00
Social programme	3.64	1.02
Opportunities to meet colleagues	4.25	0.78
Value for money	3.88	0.88

"Good venue; good, interesting keynotes; good time management; easy to meet colleagues"

"Overall very good. Only regret is that because of too many people, the opening ceremony and the conference dinner were too crowded"

The majority of respondents (69.8%) reported that they had accessed the online Abstract Book before the conference. Most of the comments on the abstract book were positive although a minority of delegates indicated that they would prefer to have a paper version at the conference.

"Excellent that you have gone digital only. It is also much more useful because you can do searches"

"Whilst it would be nice to have a printed copy of the supplement booklet, it is fully understandable"

"The key thing is to have wireless access all over the conference site to facilitate access to the abstract book"

Looking forward, there are a few issues that the EC will need to consider when planning future conferences. First, as with previous conferences, the poster sessions attracted the most negative comments. These centred around the physical space devoted to the posters which often made it difficult to hear presenters or to move around the poster space. Many respondents suggested that the sessions would be improved if there were fewer posters; future Scientific Committees may decide to look at ways to limit the number of posters. Second, the online Abstract Book attracted some negative comments with some respondents indicating that they would prefer a printed version. However, other respondents agreed with the EC's decision to have an online-only Abstract Book on environmental grounds.

In conclusion, respondents' ratings of, and comments on, the conference were very positive. It was clear that respondents enjoyed the conference and thought that the scientific quality was high. We are indebted to the hard work of the Conference President, Vladimir Kezba, and the Chair of the Scientific Committee, Aleksandra Luszczynska, for ensuring the success of the conference.

"Excellent conference—congratulations to the local organisers and scientific committee for their hard work"

"I enjoyed it very much and am looking forward to attending the one next year!"

Thank you to all delegates who completed the conference evaluation survey—your comments and suggestions are very helpful and will help to shape the structure of future EHPS conferences.

Paul Norman
EC Conference Officer

EHPS conference 2012

Opportunities and challenges in real time data capture: Methods in Health Psychology Symposium II

Martyn Jones*
University of Dundee

Daryl O'Connor
Leeds University

Joseph Schwartz
Stony Brook University

Derek Johnston
University of Aberdeen

David French
University of Manchester

Marijn de Bruin*
University of Amsterdam

Following the successful first symposium on "Current issues in Randomized Controlled trials" at the EHPS conference in Crete (2011), the decision was made to organize a yearly state-of-the-art and thought-provoking symposium on methods in health psychology (a collaborative initiative by Marie Johnston and Marijn de Bruin). This year's symposium was on a topic that has the potential to radically alter the way in which we collect our data, to enhance the validity and reliability of the data collected, to advance our statistical approaches and theories, and to allow the design of individually-tailored interventions in real-time with exciting technological advances: Ecological Momentary Assessment (EMA). Fairly straightforward applications of this approach in other domains, like the electronic monitoring of medication intake behaviours in drug trials, have revolutionized models for understanding complex processes and opened up opportunities for intervention as problems occur in real-time within people's normal, everyday context (e.g., Blaschke, Osterberg, Vrijens, & Urquhart, 2012; de Bruin, Hospers, van Breukelen, Kok, Koevoets, & Prins, 2010; Haberer, Robbins, Ybarra, et al., 2012).

The many advantages of using real time data capture techniques are best summed up in the words of Affleck and colleagues (1999), who argued that these approaches allow researchers "(a) to capture as closely as possible the "real-time" occurrences or moments of change (in study variables); (b) to reduce recall bias; (c) to mitigate some forms of confounding by using participants as their own controls, and (d) to establish temporal precedence to strengthen causal inferences" (p. 747). Moreover, techniques such as EMA can be used not just to record on-going daily processes but also to examine how the co-variation between important behavioural processes (e.g., effects of daily stressors on food intake; see O'Connor et al., 2008) varies as a function of psychological interventions and different personality traits.

This symposium overview offers a summary of what EMA does best in health psychology (by Martyn Jones), and considers the challenges of linking individual difference and trait data which may be captured from different devices or gathered at different intervals (by Joseph Schwartz). This is further illustrated with studies that detail the value of examining health psychology theory in within-person as opposed to between-person designs (by Derek Johnston) and relate laboratory and field measures (by Daryl O'Connor). The symposium ends with a critical overview of the main challenges to conducting well-designed EMA studies, and our take on the future of EMA

Note: All authors contributed equally to this paper.

within health psychology (by David French and Marijn de Bruin, with contributions from the other authors).

What does EMA do best?

This symposium is timely given the focus of health psychology on dynamic processes that underlie behaviours which are often studied using methods and frequencies of data collection that cannot capture such complexity. Retrospective approaches fail to capture such within-person variation using data that are aggregated and collapsed over time (Jones & Johnston, 2011). One might question the relevance and validity of findings using such measures.

Stone & Shiffman (1994) have advocated ecological momentary assessment (EMA) as an alternative. EMA allows collection of longitudinal data from a representative part of the participant's daily experience, in real time and in the participant's natural environment. For example, behavioural diaries can capture data provided repeatedly over time using paper (Takarangi, Garry, & Loftus, 2006), PDA handheld computers, or Smartphones (Johnston, Beedie, & Jones, 2006). Behavioural diaries can capture within-person data on cognition, affect, behaviour and even performance in the social world (Bolger, Davis, & Rafeli, 2003). EMA accounts are gathered more closely in time to the event and are less biased by heuristic, autobiographical memory strategies. Experiential Sampling Methodology (ESM) is a closely related approach, first developed in the Netherlands (e.g. Csikszentmihalyi & Larson, 1987).

Diary-based EMA approaches allow the provision of reminders for diary completion, called signal-contingent recording. Data may be gathered at a set interval or times of day (interval-contingent), or following particular incidents of interest (event-contingent). Diary-based EMA methods generally have good or excellent compliance (Takarangi, Garry, & Loftus, 2006). Real-time longitudinal data may be combined from a range of devices like self-reports linked with physiological data. EMA allows testing of within-person variation in variables in a way that is difficult, often impossible, to achieve using retrospective measures and between-person (group level) designs.

The key benefit of this approach lies in the examination of events as they occur in their natural setting and allows the time course of the behaviour of interest to be modelled. Data collection can be scheduled to fit the respondent's day to explore the antecedents, correlates and consequences of daily experiences. Repeated, real-time EMA approaches are thought to improve the reliability and validity of data collection and to improve the quality of collected data (Piasecki, Hufford, Solhan, & Trull, 2007; Burton, Weller, & Sharpe, 2007). Data are time-stamped and entered into a spread-sheet automatically, with no error (Bolger, Davis, & Rafeli, 2003). Real time data collection can also be programmed to request information following state or physiological changes in the respondent (Picard & Liu, 2007). This approach has huge potential to provide accurate, real time evidence to assist in the therapeutic decisions of practitioners.

In sum, EMA approaches provide more data, probably better data and certainly different data than previously, allowing the application of more powerful analytic techniques to critical, real life questions than ever before.

Can intensive 1-day EMA monitoring be used to assess traits?

EMA is advocated as a strategy for generating ecologically valid assessments of individuals' emotions, cognitions, behaviors, and physical

states. The extent to which these assessments provide reliable, reproducible and valid measures of individual differences is unknown. Their relationship to traditional trait questionnaires is also not known. In the Masked Hypertension Study, we collected electronic diary assessments of affect (e.g., anxiety, depression) every 30 minutes during waking hours for two 24h periods (i.e., the EMA measures), several months apart, and a variety of traditional questionnaires including the Spielberger Trait Anxiety Inventory (STAI, 1970) from 157 employed individuals. In a multilevel model (PROC Mixed in SAS; Schwartz & Stone 2007) we treated 24h subject-level means of EMA anxiety as a latent variable measure of trait anxiety and estimated 'EMA trait stability' (correlation of Time 1 means with Time 2 means), and the correlations of these EMA means with the STAI scores. The results showed that the 24h average of EMA anxiety is very stable (r = 0.91), strongly suggesting that one day of intensive EMA monitoring is adequate for capturing individual differences. EMA mean anxiety, however, correlated only modestly with the STAI assessment of the same construct (correlations ranging from .21-.24). The question for future studies is now whether this implies that the EMA assessment (100-pt VAS rating of a single item, "anxious/tense") fails to capture important aspects of the STAI, or that the STAI suffers (more) compared with EMA from recall bias, the difficulty of mentally aggregating over time, reliance on semantic memory (self-image), and/or social desirability response bias. Hence, using EMA within this context was feasible, provided reliable and reproducible results, and raised interesting questions regarding the validity of a widely used traditional questionnaire measure.

Testing theories within individuals

One of the powerful features of EMA is that it enables, indeed encourages, the repeated measurement of the behaviour of individuals. This allows the testing of theory within individuals as well as the between-person tests that are more common in psychology. The importance of testing theory within individuals has been pointed out very vigorously by Molenaar (2004) who argues that most psychology theories should explain the behaviour of individuals, and variations in that behaviour. He has clearly described the danger of accepting the fallacy that a theory that explains differences between individuals will necessarily explain variations within an individual. We tested two of the main theories of work related stress, namely Karasek's (1979) Demand Control (DC) model and Siegrist's (1996) Effort Reward Imbalance (ERI) model, in a large sample of nurses measured every 90 minutes over 3 working shifts. We used PDA-based EMA that we have previously described (Johnston, Beedie & Jones, 2006) and multilevel modelling to conduct within-person tests. We showed that, as predicted, Negative Affect (NA) was greatest when Demand or Effort was high and this was moderated by Control and Reward. This indicates that the same determinants of work related stress operate within people as between. However EMA studies and multilevel modelling enables one to take this further and examine whether one's models apply to the individuals under study. We can show that while the DC model applied to virtually all nurses, the ERI Model appeared to be inappropriate for approximately 30%. This could not be established by traditional between-subject methods and illustrates how EMA studies can increase our understanding in unique ways. EMA studies lead us to ask new questions; in this case the challenge becomes to establish what other factors (environmental or personal) determine this difference between people. This is an important theoretical question which has practical implications since it suggests that some interventions, such as increasing reward,

may not be effective or might even be harmful for some individuals.

Testing the efficacy of interventions in real time

In the current study, we used real time data capture techniques to explore whether a brief, easy to administer intervention, known as the written emotional disclosure paradigm (Pennebaker & Beall, 1986), was able to buffer against the effects of maladaptive rumination (i.e., brooding) on daily cardiovascular outcomes. Most previous research in this area has traditionally investigated the impact of written emotional disclosure on one-off, single assessment outcome measures (e.g., number of visits to general practitioners, frequency of cold symptoms, antibody concentrations). However, using innovative techniques allowed us to investigate whether the intervention was effective at lowering blood pressure and/or momentary levels of psychological stress on multiple occasions throughout normal working days. To this end, EMA was applied to ambulatory blood pressure assessments taken every 30 minutes for 12 hours on two weekdays following the intervention yielding 1339 observations from 55 participants. In addition, we were able to explore whether the efficacy of the intervention was moderated by important between-participant factors such as personality, whilst controlling for baseline levels of blood pressure and other potential confounders (e.g., physical activity level, body mass index). To our mind, these approaches are important as they generate a large number of observations in real-world contexts, which in turn increase ecological validity and confidence that any observed intervention effects are real and meaningful for health.

Research priorities and future directions

The studies presented at this conference contrast key opportunities for advancing health psychology research with a range of methodological challenges and questions. These opportunities and challenges are discussed in terms of the main features of EMA, namely those that relate to Ecological aspects, Momentary aspects, and Assessment aspects of such research.

First, studying measures taken within the context where behaviours, emotions, and cognitions actually occur may seem like such an obviously good thing as to not require stating. However, in literatures relating to common social cognition models, such as the Theory of Planned Behaviour (TPB, Ajzen, 1991), measurement within context is the exception rather than the norm. People are usually asked to complete questionnaires about behaviours such as alcohol consumption, exercise, and screening attendance either at home or in laboratories/classrooms. This lack of context has been shown to be misleading, at least in relation to alcohol consumption. Most TPB studies indicate that normative factors are not important in predicting drinking intentions and behaviour, whereas when people are asked to complete questionnaires about drinking behaviour in the context in which it occurs, i.e. in bars, then normative factors become very strongly predictive (Cooke & French, 2011; Cooke & French, 2012). There is a need for more consideration of context generally, and more examination of where completing measures in an inappropriate context may produce misleading results.

The momentary aspects of EMA may be potentially both a strength and a weakness. First, as already noted, intensive repeated measurement allows more appropriate tests of theory within people, rather than between people. A second potential advantage is that designs using such intensive measurement may make more sense to participants: the focus of

the research is on variation within themselves. This may partly explain the good levels of retention which EMA studies show, despite high respondent burden (Burton, Weller, & Sharpe, 2007). However, a potential downside is that the prospect of high respondent burden may lead to higher levels of selection bias in recruitment, relative to less intensive measurement. This issue warrants closer attention, especially for "convenience" samples not drawn from a clearly defined sampling frame.

The final aspect of EMA relates to Assessment, which is an enormous topic in its own right (Meier, 1986). Repeated measurements may make EMA liable to reactivity of measurement effects (for an overview see French & Sutton, 2010). Mean levels of variables such as reported pain do not alter across repeated measurement (Aaron et al, 2005), suggesting a lack of reactivity. However, it may be worth considering the framework of Golembiewski, Billingsley and Yeager (1976), who propose three kinds of change in measurement, namely alpha, beta and gamma change. Gamma change indicates a reconceptualization of the domain of interest, e.g. a person may initially consider "stress" to be synonymous with anxiety but later understand "stress" to be more composed of excess demand. Beta change indicates a recalibration in scaling, where e.g. a rating of "5" on the first occasion of measurement is not the same as a rating of "5" on the hundredth occasion. Alpha change indicates "true" change, which is what researchers are usually interested in. It is possible that the observed lack of change in mean scores across time may be taken at face value as due to a genuine lack of alpha, beta or gamma change, or it may indicate that such changes are occurring, but are not resulting in mean score changes. It would be worth examining such potential changes in response shift, given that repeated measurement is a core characteristic of EMA studies. It is certainly the case that questionnaire measurement can involve the creation of new cognitions, as well as their assessment, as indicated by the use of "think aloud" methods (e.g. Darker & French, 2009).

Apart from reactive effects, other measurement challenges include making sure that any measures used possess good sensitivity to change. This psychometric criterion is even more important in within-person designs, and tends to be neglected relative to the criteria of reliability and validity. There is still a need to establish that single item measures are reliable and valid, which may be a challenge given that single item measures are often used to reduce respondent burden. In addition, asking questions within context should result in more valid self-reports. It does not require people to recall their past behaviour nor does it require people to mentally aggregate their experiences to produce an overall summary score, as is the case with more traditional approaches. However, a comparison of where there are differences between more traditional summary measures and EMA measures may shed light on the mental processes that people use to produce such summary scores. More generally, these advantages of EMA should not deflect attention away from the fact that people still need to interpret questions, retrieve appropriate information from memory and format their responses (Jobe et al, 1991). Although EMA may ameliorate some of these issues for self-report measurement, it is still important to examine how people approach the task of completing self-report measures.

Conclusions

Repeated real-time EMA offers the possibility of providing better, more reliable, more context-specific data that are relatively unaffected by a range of recall and other biases. EMA is a flexible methodology that combines self-report

of behaviours, cognition and emotion with other forms of real-time data, e.g. physiological measurement. It allows the assessment of both trait-like individual difference variables and within-person changes allowing the possibility of providing innovative within-person tests of Health Psychology theory. EMA offers the possibility of testing mechanisms of change within complex interventions set in a real-world context, in a manner not previously possible. An exciting prospect is that it also allows us to intervene with unhealthy behaviours or cognitions as they occur in real-time. These aspects of EMA methodology are ripe for further development. EMA is not without its challenges, however. Further research is needed to detail the precise effects of perceived burden and selection bias, to establish the psychometric properties of short scale EMA measures, and to further elaborate the context specific effects of repeated measurement on reactivity, complexity and entrainment for EMA-derived outcomes.

References

Aaron, L., Turner, J., LLoyd, M., Brister, H., & Sawchuk, C. (2005). Electronic Diary Assessment of Pain Related Variables: Is Reactivity a Problem? *Journal of Pain, 6*, 107-115. doi: 10.1016/j.jpain.2004.11.003

Affleck, G., Zautra, A., Tennen, H., & Armeli, S. (1999). Multi-level daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746-754. doi: 10.1037/0022-006X.67.5.746

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes, 50*, 179-211. doi: 10.1016/0749-5978(91)90020-T

Blaschke, T.F., Osterberg, L., Vrijens, B., & Urquhart, J. (2012). Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. *Annual Review of Pharmacology & Toxicology, 10*, 275-30. doi: 10.1146/annurev-pharmtox-011711-113247

de Bruin, M., Hospers, H.J., van Breukelen, G.J. P., Kok, G., Koevoets, W.M., & Prins J.M. (2010). Electronic monitoring-based counselling to enhance adherence among HIV-infected patients: A randomized controlled trial. *Health Psychology, 29*, 421-428. doi: 10.1037/a0020335.

Bolger, N., Davis, A., & Rafeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology, 54*, 579-616. doi: 10.1146/annurev.psych.54.101601.145030

Burton, C., Weller, D., & Sharpe, M. (2007). Are Electronic Diaries Useful for Symptoms Research? A Systematic Review. *Journal of Psychosomatic Research, 62*, 553-61. doi: 10.1016/j.jpsychores.2006.12.022

Cooke, R., & French, D.P. (2011). The role of context and timeframe in shaping relationships within the theory of planned behaviour. *Psychology and Health, 26*, 1225-1240. doi: 10.1080/08870446.2011.572260

Csikszentmihalyi, M., & Larson R. (1987). Validity and Reliability of the Experience Sampling Method. *Journal of Nervous and Mental Disease, 175*, 526-37. doi: 10.1097/00005053-198709000-00004

Darker, C.D., & French, D.P. (2009). What sense do people make of a Theory of Planned Behaviour questionnaire? A think-aloud study. *Journal of Health Psychology, 14*, 861-871. doi: 10.1177/1359105309340983

French, D.P., & Hevey, D. (2008). What do people think about when answering questionnaires to assess Unrealistic Optimism about skin cancer? A think aloud study. *Psychology, Health and Medicine, 13*, 63-74. doi: 10.1080/13548500701243959

French, D.P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology, 15*, 453-468. doi: 10.1348/135910710X492341

French, D.P., & Cooke, R. (2012). Using the Theory of Planned Behaviour to understand binge drinking: the importance of beliefs for developing interventions. *British Journal of Health Psychology, 17*, 1-17. doi: 10.1111/j.2044-8287.2010.02010.x

Golembiewski, R.T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science, 12*, 133-157. doi: 10.1177/002188637601200201

Haberer, J.E., Robbins, G.K., Ybarra, M., Monk, A., Ragland, K., Weiser, S.D., Johnson, M.O., & Bangsberg, D.R. (2012). Real-time electronic adherence monitoring is feasible, comparable to unannounced pill counts, and acceptable. *AIDS & Behavior, 16*, 375-82. doi: 10.1007/s10461-011-9933-y

Jobe, J.B., & Mingay, D.J. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology, 5*, 175-192. doi: 10.1002/acp.2350050303

Johnston, D., Beedie, A., & Jones, M. (2006). Using computerized ambulatory diaries for the assessment of job characteristics and work-related stress in nurses. *Work and Stress, 20*, 163-72. doi: 10.1080/02678370600902872

Jones, M.C., & Johnston, D. (2011). Understanding phenomena in the real world: The case for real time data collection. *Journal of Health Services Research & Policy, 16*, 172-176. doi: 10.1258/jhsrp.2010.010016

Karasek, R.A. (1979). Job demands, job decision latitude, and mental strain: implications for job redesign. *Administrative Science Quarterly, 24*, 285-307. doi: 10.2307/2392498

Meier, S.T. (1994). *The chronic crisis in psychological measurement and assessment: A historical survey*. San Diego, CA: Academic Press.

Molenaar, P.C.M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*, 201-218. doi: 10.1207/s15366359mea0204_1

Piasecki, T., Hufford, M., Solhan, M., & Trull, R. (2007). Assessing clients in their natural environments with electronic diaries: rationale, benefits, limitations and barriers. *Psychological Assessment, 19*, 25-43. doi: 10.1037/1040-3590.19.1.25

Picard, R., & Liu, K. (2007). Relative subjective count and assessment of interruptive technologies applied to the remote monitoring of stress. *International Journal of Human-Computer Studies, 65*, 361-75. doi: 10.1016/j.ijhcs.2006.11.019

O'Connor, D.B., Jones, F.A., Conner, M., McMillan, B. & Ferguson, E. (2008). Effects of daily hassles and eating style on eating behaviour. *Health Psychology, 27*, 20-31. doi: 10.1037/0278-6133.27.1.S20

van Oort, L., Schröder, C., & French, D.P. (2011). What do people think about when they answer the Brief Illness Perception Questionnaire? A 'think aloud' study. *British Journal of Health Psychology, 16*, 231-245. doi: 10.1348/135910710X500819

Pennebaker, J.W., & Beall, S.K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*, 274-281. doi: 10.1037/0021-843X.95.3.274

Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In A. Stone, S. Shiffman, A. A. Atienza & L. Nebeling (Eds.), *The Science of Real-Time Data Capture: Self-Report in Health Research* (pp. 76-113). Oxford: Oxford University Press.

Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal Occupational Health Psychology, 1*, 27-41. doi: 10.1037/1076-8998.1.1.27

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA.: Consulting Psychologists Press.

Stone, A., Shiffman, S. (1994) Ecological Momentary Assessment (EMA) in Behavioral Medicine. *Annals of Behavioral Medicine, 16*, 199-202.

Takarangi, M., Garry, M., & Loftus, E. (2006) Dear diary, is plastic better than paper? I can't remember: Comment on Green, Rafaeli, Bolger, Shrout and Reis. *Psychological Methods, 11*, 119-22. doi: 10.1037/1082-989X.11.1.119

**Corresponding authors:*

Martyn Jones

m.c.jones@dundee.ac.uk

Marijn de Bruin

m.debruin@uva.nl

synergy workshop 2012

Mixed methodology in Health Psychology: using Pragmatism to overcome the 'irreconcilable epistemological differences' between quantitative and qualitative methods

What was it all about?

We started with a theoretical input on pragmatism and on how to take into account theory and data to generate and specify research questions. This was the basis for discussing qualitative and quantitative methods: Each method can explore different aspects of a phenomenon and using them together can increase the confidence in findings. We developed research proposals in small groups and argued about at which stages of a specific design, qualitative and quantitative methods are most appropriate. Our discussion stopped at a point that might be a research question in the future. Most mixed methods studies profit from quantitative and qualitative approaches but not of their integration. To sum it up: To conduct good research, you have to have a real world research question that changes health services and people's health. You have to use various methods to reach these goals. Qualitative and quantitative methods can equally contribute to improving research and ideally, combining both is more than the sum of its parts.

What was it like?

The SYNERGY workshop created a dialogue between quantitative and qualitative researchers. The contrasting expertise was applied to clarify the meaning of mixed methods methodology and to develop mixed-method designs. Actually, it was not only the topic of mixed methodology which generated more knowledge; it was great fun to meet researchers from all over the world and exchange research experiences. Thanks to Paul (Flowers) and Rachel (Shaw) for facilitating, SYNERGY for organizing and all participants for discussing.

What did you gain?

Discussions about pros and cons of qualitative and quantitative methods in research designs opened my mind for different approaches and encouraged me to more specifically state mixed methods in a research proposal. It is up to the future to find more ways to integrate quantitative and qualitative methods and results. The special interest group will keep us connected and up-to-date with upcoming proposals, projects and papers.

Anna Levke Butt

synergy workshop 2012

Mixed methodology in Health Psychology: using Pragmatism to overcome the 'irreconcilable epistemological differences' between quantitative and qualitative methods

synergy

This year's SYNERGY workshop on mixed method methodology aimed to advance our understanding of applying qualitative and quantitative research methodologies in an integrated and synergistic fashion to health psychology research. The 3 day workshop was led by Dr Rachel Shaw & Professor Paul Flowers, two experts in the field of mixed methodologies. The workshop covered a wide range of topics relevant to mixed methodology including a particular focus on epistemological differences between qualitative and quantitative methods, and ways of overcoming these within mixed methods research studies. In addition to theoretical and methodological issues surrounding mixed methods research, the workshop introduced the UK Medical Research Council's (MRC) complex interventions development framework to explore and guide the development of mixed methods approaches to research projects. This was attained through small group work which provided a context for applying the covered mixed methodology approaches to practically relevant health psychology research. Moreover, participants were able to bring their own research projects, challenges and experiences into the workshop environment to centre discussion on issues most relevant to participants.

The workshop was not only well organised and thought provoking, but also an extremely useful exercise to reflect on one's own overall approach to health psychology research, which for many participants included using either qualitative or quantitative methods. By considering what could be gained by embracing other methodologies in an integrative fashion the potential for improving health psychology research became apparent, and many different ways of achieving integrative mixed methods research were discussed. Additionally, as mixed methods research has only relatively recently become established as a research approach, the challenges in obtaining sufficient expertise in various different research methodologies with a view to truly integrating approaches became apparent. More high quality and collaborative research is needed to advance this new development and establish it as a mainstream approach within health psychology methodology.

Overall, the relaxed and productive atmosphere combined with the flexible delivery of the workshop content made the 3 days enjoyable and relevant for participants. The overall level of expertise of the workshop facilitators as well as the participants ensured stimulating debates, which persisted despite the increasingly tropical temperatures in Prague that week (there were rumours that the last workshop day was the hottest day in the history of Prague ever recorded). What the workshop allowed me personally was, in addition to getting to know a great bunch of people, to question my habitual approaches to designing research and to embrace other research methods with a view of integrating these in order to improve the overall quality and ultimately the impact of health psychological science.

Stephan Dombrowski

synergy workshop 2012

Overcoming our 'irreconcilable epistemological differences' and moving beyond the dichotomy of qualitative versus quantitative research in health psychology

Rachel Shaw & Paul Flowers
Synergy workshop facilitators

Mixed methods research is becoming more popular with European Health Psychology and beyond. Such approaches are complex and challenging yet potentially offer the best approach to addressing complex applied real world research questions. In Charles University on the hottest days on record in Prague, a very broad range of European scholars dedicated themselves to reflect, learn and play with a variety of mixed methods approaches addressing a range of research questions. The intensive workshop covered a vast terrain, from the philosophy of Pragmatism and the logics of inquiry (Hiles, 2012), to developing research designs which addressed current calls for commissioned health research. The ethos focussed on group learning, career development and critical discussion throughout the two and a half days.

The workshop began with a discussion about the opportunities Pragmatism can offer Health Psychologists as a philosophical touchstone for thinking about mixed methods research. Pragmatism will enable Health Psychology to develop across Europe and deliver impact at multiple levels (transforming health and well-being) as its validity is judged by whether it reaches its goal, i.e., whether the research question has been answered satisfactorily (Yardley & Bishop, 2008) rather than any rigid commitment to any single epistemology, sometimes referred to as methodolatry (Chamberlain, 2000).

The MRC Complex Intervention Framework (2008) was presented and discussed as one way of implementing mixed methods in applied health research. Through a number of small group exercises which mixed those with qualitative and quantitative expertise, we worked on quickly developing research designs based around the insights of pragmatism. Each design also followed the iterative cycle of the stages of feasibility/piloting, development, implementation and evaluation to address a principle research question.

In our small group discussions, we attempted to address particular research questions set by commissioned calls for research proposals. In these discussions, we applied our learning concerning Pragmatism and began to think carefully about the appropriateness of method(s) in relation to particular objectives rather than blindly favouring qualitative or quantitative methods. A key challenge remained though: how and when do we mix methods responsively instead of formulaically. Simple sequential mixed methods research designs might represent an emerging norm but perhaps are not always the most appropriate.

Synergy presented an opportunity for people from across Europe with diverse skills and expertise to focus upon a common language. Our differences were set aside and replaced by a common, shared concern with answering particular research questions to the best of our collective ability. It would be fair to say that the days in Prague have only started the attendees on a journey and for all of us the workshop raised more questions than it provided answers.

References

Chamberlain, E. (2000). Methodolatry and qualitative health research. *Journal of Health Psychology,5*, 285-296.

Hiles, D.R. (2012). Mixed methods and the problem of the logic of inquiry. Paper presented at the QMiP symposium on Mixed methodology in Psychology. *Annual Conference of the British Psychological Society*, April 18-20th, London.

MRC (2008). Developing and evaluating complex interventions: new guidance. Medical Research Council. Available at: www.mrc.ac.uk/complexinterventionsguidance (accessed 17/10/12).

Yardley, L. & Bishop, F. (2008). Mixing qualitative and quantitative methods: a Pragmatic Approach. In C. Willig & W. Stainton Rogers (Eds.) *The Sage handbook of qualitative research in Psychology*. London: Sage, pp. 352-369.

create workshop 2012

"Qualitative research in perspective: Achieving excellence in qualitative research practice"

"Be informed, critical and reflexive" was the message that resonated in the minds of 32 researchers who came from 12 different countries to attend the CREATE workshop. Held in Prague, Czech Republic, and facilitated by Professor Kerry Chamberlain from Massey University, this year's workshop provided many invaluable insights into the vital components of qualitative research, an approach that remains less common among health psychologists at large.

The 3-day workshop focussed on the complete process of conducting qualitative research from its theoretical assumptions and design, through data collection methods, interpretation of results and finally publishing. The first part of the workshop focussed on the differences in the views and understanding of the world between qualitative and quantitative approaches. Much of this session looked at their complementary as opposed to conflicting aspects. As in any research process, it is essential to select beforehand what pathway one will take to answering one's research questions. Comprehension of the delineation and interaction between epistemology, theoretical frameworks, methodologies and methods outlined during the workshop does not only help 'wrap' a research idea into a coherent concept; it also serves as guide, providing theoretical support. By asking questions about what it is that we want to know and how we can attain the required knowledge, we need to choose or define an epistemological standpoint from which to start. Epistemology is a theory of knowledge that "deals with the nature, sources, and processes of knowledge" (Baptiste, 2001). The CREATE participants developed understanding of it through exposure to three examples of the many epistemological stances; objectivism, social constructivism and subjectivism. Epistemology affirms theoretical perspectives, or a set of assumptions about the world that we are asking questions about, which provides a primary framework for data interpretation. Theoretical perspectives (e.g. hermeneutics, feminism,…) are further linked to methodologies (e.g. discourse analysis, ethnography,…) which offer more concrete plans of verifying those assumptions by applying specific techniques or procedures to data collecting methods (e.g. interview, document analysis,…) (Crotty, 1998).

Although one could argue that the process of defining of the abovementioned elements of research design could be more deductive than inductive, it is meaningful to set these elements before the actual research takes place, regardless of the direction of the thinking that is employed. Creativity and cross-combination of various elements to best accommodate a research idea is encouraged rather than "getting stuck" with existing theoretical choices. As long as the approach chosen is justified via an informed and thorough manner, there is no reason for it not to be undertaken. Such an approach enables us to justify a particular research design and explain a research rationale, while developing a critical mind at the same time.

The second part of the workshop delved more deeply into qualitative data analysis. Some of the qualitative research methodologies, such as grounded theory, phenomenology, ethnography and thematic analysis were explored. Starting out with an admittedly limited view of qualitative research, our perspectives were broadened as the facilitator provided us with insights into his work by sharing his knowledge and experience of several data collection methods. These ranged from individually-focussed to group-based approaches; which included written, audio and visual means of data collection. Furthermore, various issues were highlighted that a qualitative researcher should consider during the process of data analysis, including the distinction between description and interpretation and thinking beyond themes when interpreting information. We soon appreciated the need to challenge our pre-conceived notions of qualitative research. The key factors above all else are that one should be informed, critical and reflective. By being reflective, we become aware of our own values and assumptions and how they shape and direct our research and we learn to question assumptions underlying research paradigms.

The writing session was defined by the self-explanatory motto "*Invent, compose, revise*". Qualitative studies typically use quotes, transcripts or pictures to pass their message on to their audience and this session, filled with such documentation, was nothing less than an excellent representation of what dimensions of a story are possible to capture via qualitative techniques. Therefore, because of the inherent complexity of many of the findings in qualitative studies, it is essential to make the message as clear as possible, while preserving its richness and meaning. It is also important to be sensitive to any inconsistencies in our data rather than being satisfied by confirmed hypotheses, as it is these incongruities that lead to a better understanding of the "real world".

Undertaking qualitative research can be highly rewarding and the capabilities of a critical mind should be harnessed, refined and confronted so that we are inspired to push the boundaries of qualitative research. Throughout the course the CREATErs were consistently encouraged to be inquisitive and critical, yet sensitive and perceptive.

The 2012 CREATE session on 'Qualitative Research in Perspective' resulted with its delegates walking away with a newly acquired outlook on conducting qualitative research. In addition to its highly informative nature, the workshop also inspired many of us to re-think our ways of approaching qualitative studies and how they could be pursued in a more systematic and informed manner. The organisation of this workshop has provided a highly conducive environment for intellectual exchange between

researchers desiring to conduct potent qualitative work within the field of health psychology, and the rich social programme offered ample opportunities for participants to forge international relationships. We believe that participation in this workshop was an inspiring first step into the waters of qualitative research, and that the discourse on conducting effective qualitative research in health psychology will continue at future group meetings.

Lucia Rehackova & Alden Lai

References

Baptiste, I. (2001). Qualitative data analysis: common phases, strategic differences. *Forum: Qualitative Social Research, 2*. Retrieved November 12, 2012 from http://www.qualitative-research.net/

Crotty, M. (1998). *The foundations of social research: meaning and perspective in the research process.* London: Sage.

EHPS report

Methods in health psychology: how do we know what we really know?

Marijn de Bruin & Marie Johnston

University of Amsterdam & University of Aberdeen

Part of the charm of health psychology is the focus on direct applicability of our research results. There is also a risk in there, however, which is that our research can be focused more on obtaining applied results (e.g., the key determinants of achieving sufficient physical exercise within a particular population within a particular context) rather than contributing to a cumulative science through methodological and theoretical progress. Yet, the fact that our results might be applied immediately and influence behaviour of at-risk groups, prevention/health care workers, and policy makers, makes it even more important that the methods are sound and the conclusions valid. We need to take time to consolidate the methods we have established over the short life of our discipline and to identify the areas where we need to critique and improve our methods. We are therefore excited that EHPS has decided to run a **Methods in health psychology track**, starting in 2013 in beautiful Bordeaux.

In this article we want to discuss, first, an approach that we think could—if more widely adopted—contribute to a field with a richer, firmer set of research methods suitable for attacking the questions arising from both theory and practice; that generates conclusions that are valid; offers insights that are interesting and valuable for health psychology as a *science* rather than for the selective group of people working on the same health topic in a similar context; and that this is seen as an ongoing process where every investigation has the potential to contribute to better research methods and to advance theory while generating findings of practical use for promoting people's health and well-being. In other words, an approach that lets us find out what we *really* know. After that we discuss how this approach could translate to health psychology, and we end with suggestions of topics that could be covered in the new health psychology methods track.

In the Methods track, we expect to highlight and consolidate not only the key advances in commonly used research methods, but also to identify opportunities for (more rapid) progress that may not have been exploited to date due to researchers having paid more attention to applied outcomes than to the methodological and theoretical implications, i.e. the 'missed opportunities'. Reviews of 25 years of health psychology (Johnston, Weinman and Chater, 2011) and comments we have received suggest this is timely and that many health psychologists are in fact concerned about these issues. So we offer our reflections as a starting point for debate within our research community, which will hopefully translate into thought-provoking symposia in the newly established Methods track.

What approach produces rapid scientific progress?

As a young discipline, it has been important to demonstrate that we can make relevant contributions. For example, our work on developing an intervention to reduce disability following stroke has been implemented by the

Scottish Government (Johnston et al 2007), psychometric approaches are now part of the expected standards in the measurement of health outcomes (Bowling, 2001), and we are currently developing national guidelines and an implementation strategy for the delivery of 'current best practice' in HIV adherence care based on the results of meta-analyses (de Bruin et al, 2009; 2010). Without achieving this kind of base, health psychologists would not attract the funding to continue their work. However, the drive to produce results that are directly applicable can result in a confirming and narrow mind set, and can lead to repetitive work that does not present a relevant scientific or theoretical advance (e.g., the umpteenth observational cross-sectional study showing that the Theory of Planned Behavior (TPB) constructs explain a health behaviour). So what mind set could help us rise above these applicable outcomes?

In 1964, Platt published a paper in Science in which he reflected on fields where scientific progress was more rapid than in others. According to Platt, the usual explanations like "the tractability of the subject, or the quality of men drawn into it, of the size of the research contracts are important but inadequate". He suggested that in the most prolific fields there is a culture of applying 'Strong inference methods', of which the separate elements are no different from "the old-fashioned method of inductive inference that goes back to Francis Bacon" (Platt, 1964). What Platt refers to as Strong Inference is the systematic, formal and explicit application of the following steps following the initial proposition of an hypothesis, theory or explanation: (1) Formulate alternative explanations that could explain observed results; (2) Devise a crucial test (or several) that will exclude one or more of these explanations; (3) Carry out the research; and repeat this cycle by making sub hypotheses and sequential hypotheses to refine the options that remain. In this way an initial invention moves along the branches of a logical tree, with several options (i.e., hypotheses) at each fork that are then refuted, leaving one option open which is then pursued leading to the next fork, and so forth, until a particular conclusion has been reached. Platt observed that in the most productive fields Strong Inference was integrated in all thinking, publications, conference presentations, and so forth. So after one group published their results and conclusions, including possible alternative explanations and tests, other groups also pursued these ideas, confirmed or refuted the hypotheses, presented the alternative explanations and experiments for their findings, and so forth; progressing efficiently up the same logic tree.

An interesting side-effect of this approach of framing multiple alternative hypotheses for one's research findings is that people do not become too attached to a single hypothesis; in fact, researchers can take pride in formulating alternative hypotheses and clever experiments that can then be pursued and confirmed or refuted by others, thereby contributing to scientific progress beyond the results of their own experiments. There are more interesting ideas in this paper, but we would like to highlight this Strong Inference approach since we could probably use a bit of this ourselves.

What approach might produce more rapid progress in health psychology?

So how does this rationale translate to health psychology (Platt reflected on fields like molecular biology)? We think, first, that many of our applied studies permit us to ask more questions and pose more hypotheses that go beyond the applied questions, and thereby contribute to resolving ongoing theoretical or methodological debates. Second, after studying

these initial hypotheses (regardless of whether these are confirmed or not), we could generate alternative explanations for the results obtained and propose methods required to test these alternative hypotheses. These steps could be adopted in observational and intervention/experimental studies with diverse research designs and methods. By focusing our applied studies more on hypotheses of wider scientific interest, by generating alternative hypotheses, and through critiquing the methods we use, we may be able to reach conclusions that are relevant not only for the application, but which inform theoretical and methodological development in the field as a whole.

To give an example for observational studies, let's return to the example of the TPB (the umpteenth…). We can see that the results of such a study may be relevant for the applied context, but they can additionally test whether, for example, attitudes are more predictive of intention than subjective norms, which kind of attitudes and which kind of subjective norm is most relevant, or how subjective norms can best be measured; they may also allow simultaneous testing of different hypotheses about the intention-behaviour gap brought forward previously by others (e.g. explanations based on intention stability, planning ability, or self-regulatory skills; e.g. de Bruin et al., 2012; DiBonaventura & Chapmann, 2005; Sniehotta, Nagy, Scholz & Schwarzer, 2006), or studying the impact of past behaviour in the model (does it capture habit or does it mainly control for confounding?); or examine what is left of the theory when it is tested using a within- and between-subject repeated-measures model that captures change over time; or the difference in results when a subjective versus objective behavioural measure is being used as the dependent variable.

To give an illustration for intervention studies, consider the example of an intervention directed at increasing the uptake of an effective treatment. Whereas for that study the key outcome is whether it does indeed result in an improved uptake, one might also test theoretical or methodological hypotheses of wider interest, such as whether the delivery of particular behaviour change techniques indeed produce the intended change in determinants and behaviour, and if so for whom and under what conditions; evaluate the role of demand characteristics (McCambridge, de Bruin, & Witton, 2012) and other potential sources of bias; comparing different methods for assessing and controlling for variability in care provided to control groups; or comparing the feasibility and accuracy of different measures for assessing the quality of intervention delivery. If we identify the 'hot topics' in our field, and in our applied studies consistently pit hypotheses directed towards such theoretical and methodological questions against each other, as a field we could become much more efficient in understanding the processes involved and the conditions under which each of these hypotheses may hold.

So how is this linked with the starting point of this paper, namely a new EHPS conference track on research methods? We think that by putting more emphasis on methods, we will be able to ask the questions and present the evidence compatible with adopting a Strong Inference approach. Upon hearing an explanation for a result, we will have the opportunity to ask the question 'But what investigation could disprove your hypothesis?' (cf. Platt), and to reflect at a higher level on the scientific nature of our research (e.g. formulation of initial and alternative hypotheses beyond the applied question at hand) and on how our designs, measures and analytical/statistical models could be challenged. Our hypothesis (or hope) is that by

increasing the emphasis on improving our methods, we will encourage the research practices compatible with a Strong Inference approach, which could lead to an—if not exponential, than a substantial—increase in the advancement of our theory, the quality of research methods, and the impact of our research.

What methodological issues do we need to address?

There are numerous methodological challenges in the field of health psychology. We invite you to propose your ideas and symposia for the methods track for the 2013 conference and for subsequent conferences. In order to illustrate the breadth of topics we might consider, the

Table 1. *Possible methodological questions that could be central to a symposium*

Process of research	Methodological topics	Possible methodological questions
Background	Systematic reviews	What are the basic standards for a systematic review? When is meta-analysis appropriate?
	Feasibility & acceptability of interventions & research studies	How should feasibility and acceptability be assessed? e.g. how should quantitative and qualitative methods be used and integrated?
	Piloting methods	What pilot work is necessary before trialling an intervention? – e.g. for power calculations
	Existing data sets	How can we use existing data sets to investigate new research questions, without collecting new data?
Research question	Confirmation or testing mindset	How can we move from 'confirmatory' research questions to scientific 'testing' questions? How can one test competing hypotheses?
	Questions about methods, theory and/or application	Does the test of the intervention also allow test of theory, e.g. in process evaluation or in (fractional) factorial designs? Or to develop better methods of reporting intervention content e.g. using BCTs?
	Replication	When is a replication study needed?
Research design	Choice of research design	Have we approached the research question in the right way? Are there better research designs to answer the question?
	Cross-sectional studies	When are cross-sectional designs appropriate?
	RCTs	How many control groups are necessary? How can one characterize the active ingredients of a control group? How can one test theories or hypotheses within an RCT design? What sources of bias need to be controlled?
	N of 1 studies and within person studies	What can these research designs contribute? How do they complement between person studies?
	Qualitative and mixed method studies	What are the basic standards required for different types of studies using qualitative data? How can qualitative and quantitative data be integrated?
Participants	Who?	When is it appropriate to include e.g. students, clinical populations? How can one ensure representative populations?
	How many?	How does one justify the number of participants included? Are CONSORT methods of reporting adequate?

Process of research	Methodological topics	Possible methodological questions
Measures	Selection	How do we choose an appropriate measure: one which measures the intended construct? What evidence of validity is required before using a measure? When is it appropriate to use validated tests with norms and when should new measures be developed?
	Methods of measurement	What are the biases associated with subjective and objective measures? Are there systematic differences in the results of self-report, observational and automatic measures? When is it appropriate to have real-time (EMA) or retrospective assessments? How should physiological and clinical measures be used?
	Psychometrics	What are minimum reporting standards? Is internal consistency enough?
	Timing	When should measures be taken? How does one decide on the appropriate follow-up period?
Interventions	Reporting intervention content	How can we improve methods of reporting intervention content? How far does the BCT approach get in doing this? How far does the Intervention Mapping approach get in doing this?
	Reporting intervention delivery	What is the minimum set of information required to report the methods of delivery of an intervention? How do we ascertain whether the intervention was delivered a) by someone trained and competent (competence) and b) as reported (fidelity, adherence)?
	Clarifying distinction of intervention from control conditions	What is the minimum set of information needed to distinguish the key differences between the intervention and control groups?
Analysis	EMA data	How should real-time data be analysed to control for time-series effects? How do real-time data relate to retrospective reports?
	Longitudinal data	How can we use more advanced methods for handling longitudinal data?
	Prediction of change	How is change measured? How should one control for past behaviour?
	Large data sets	Are there improved methods of handling and analysing large data sets, especially those from large existing data sets?
Interpretation	Inference from findings	What *can* we infer from our results with some certainty? What are alternative explanations for our findings?
	Causality	What are the minimum findings necessary to either confirm or test a causal hypothesis?
	Prevalence of a problem	What evidence is required to assert that there is a problem which needs to be addressed by health psychology intervention e.g. prevalence, comparison with norms?
	Bias	What biases need to be considered in interpreting findings?
	Identifying strengths	Can we agree on what makes a study 'strong'?
	Identifying limitations	Can we agree on what makes a particular study 'weak'?
	Conclusions	Should conclusions refer to all three of theory, methods and application?

table identifies potential methodological topics and questions as they may occur throughout the process of research reported in a typical journal article or research grant application. Topics for future methods symposia might be derived from these or other topics—but no doubt you will have ideas that are more original than ours. We hope that with this paper, a yearly symposium on Methods in health psychology of which an overview will be presented in this Bulletin, and the Methodology track, we will see an increased rigor and impact of our discipline.

References

Johnston, M., Weinman, J., & Chater, A. (2011). 25 years of Health Psychology in the BPS. *The Psychologist, 24*, 890-892.

Johnston, M., Bonetti, D., Joice, S., Morrison, V., & Francis, J.J., & MacWalter, R. (2007). Recovery from disability after stroke as a target for a behavioural intervention: Results of a randomised controlled trial. *Disability & Rehabilitation, 29*, 1117-27. doi: 10.1080/03323310600950411

Bowling, A. (2001). *Measuring disease: a review of disease-specific quality of life measurement scales*. Open University Press, Buckingham.

de Bruin, M., Viechtbauer, W., Hospers, H.J., Schaalma, H.P., & Kok, G. (2009). Standard care impact on effects of highly active antiretroviral therapy adherence interventions: A meta-analysis of randomized controlled trials. *Health Psychology, 28*, 668-74. doi: 10.1037/a0015989

de Bruin, M., Viechtbauer, W., Schaalma, H.P., Kok, G., Abraham, C., & Hospers, H.J. (2010). Standard care impact on effects of highly active antiretroviral therapy adherence interventions: A meta-analysis of randomized controlled trials. *Archives of Internal Medicine, 170*, 240-50. doi: 10.1001/archinternmed.2009.536

Platt, J.R. (1964). Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science, 146*, 347-53. doi: 10.1126/science.146.3642.347

de Bruin, M., Sheeran, P., Kok, G., Hiemstra, A., Prins, J.M., Hospers, H.J., & van Breukelen, G.J.P. (2012). Self-regulatory processes mediate the intention-behavior relation for adherence and exercise behaviors. *Health Psychology, 31*, 695-703. doi: 10.1037/a0027425.

DiBonaventura, M.D., & Chapman, G.B. (2005). Moderators of the intention-behavior relationship in influenza vaccinations: Intention stability and unforeseen barriers. *Psychology and Health, 20*, 761-774. doi:10.1080/14768320500183368

Sniehotta, F.F., Nagy, G., Scholz, U., & Schwarzer, R. (2006). The role of action control in implementing intentions during the first weeks of behaviour change. *British Journal of Social Psychology, 45*, 87-106. doi:10.1348/014466605X62460.

McCambridge, J., de Bruin, M., & Witton, J. (2012). The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. *PLoS One, 7*, e39116. doi:10.1371/journal.pone.0039116.

Marijn de Bruin

m.debruin@uva.nl

Marie Johnston

m.johnston@abdn.ac.uk

EHP Editorial Board

EHPS Executive Committee (2012-14)

Disclaimer: The views expressed within the European Health Psychologist are those of the authors and do not necessarily represent those of the European Health Psychology Society (EHPS) or the European Health Psychologist's (EHP) editorial board.