

Matteo M. Galizzi and Daniel Wiesen
**Behavioral experiments in health
economics**

Book section

Original citation:

Originally published in Hamilton, Jonathan H. and Dixit, Avinash and Edwards, Sebastian and Judd, Kenneth, (eds.) *Oxford research encyclopedia of economics and finance*. Oxford research encyclopedias. Oxford: [Oxford University Press](#), 2018. (In Press)

DOI: [10.1093/acrefore/9780190625979.013.244](https://doi.org/10.1093/acrefore/9780190625979.013.244)

© 2018 Oxford University Press

This version available at: <http://eprints.lse.ac.uk/88155/>

Available in LSE Research Online: June 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's submitted version of the book section. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Behavioral Experiments in Health Economics

Matteo M Galizzi^a, Daniel Wiesen^b

a London School of Economics, Department of Psychological and Behavioural Science, Queens House, 55/56 Lincoln's Inn Fields, WC2A 3LJ London. Email: m.m.galizzi@lse.ac.uk.

b University of Cologne, Department of Business Administration and Health Care Management.

(Accepted for publication in Oxford Research Encyclopedia of Economics and Finance, Oxford University Press, Oxford, UK)

Abstract. The state-of-the-art literature at the interface between experimental and behavioral economics and health economics is reviewed by identifying and discussing ten areas of potential debate about behavioral experiments in health. By doing so, the different streams and areas of applications of the growing field of behavioral experiments in health are reviewed, by discussing which significant questions remain to be discussed, and by highlighting the rationale and the scope for the further development of behavioral experiments in health in the years to come.

Keywords: behavioral experiments in health, behavioral health economics, experimental health economics, randomized controlled experiments, behavioral data linking.

JEL codes: C90, D03, I10, I18.

Introduction

In the last decades, experiments have been successfully introduced in many fields of economics, such as industrial organization (e.g. Chamberlain, 1948; Sauermann and Selten, 1959; Plott, 1982), labor economics (e.g. Kagel, Battalio, Rachlin, and Green, 1981; Fehr, Kirchsteiger, and Riedl, 1993), or public economics (e.g. Marwell and Ames, 1981; Bohm, 1984; Andreoni, 1988). Despite the fact that the use of experiments was first advocated by leading health economists a long ago (Fuchs, 2000; Frank, 2007; Newhouse, Manning, Morris, Orr, Duan, Keeler, Leibowitz, Marquis, Marquis, Phelps, and Brook, 1981), their introduction and employment has been relatively slow to be widely accepted in health economics, policy, and management.

Recently, however, two special issues in the *Journal of Economic Behavior & Organization* (Cox, Green, and Hennig-Schmidt, 2016) and in *Health Economics* (Galizzi and Wiesen, forthcoming), and a number of dedicated special sessions in major field conferences (e.g. EuHEA, iHEA) indicate the increasing acceptance of experiments by the health economics, policy, and management communities.

The rise in interest in using experiments in health economics has coincided with the parallel growing interest in applying behavioral economics to health, as witnessed by an increasing number of books and articles on the topic (Roberto and Kavachi, 2015; Bickel, Moody, and Higgins, 2016; Hanoch, Barnes, and Rice, 2017; Loewenstein, Schwartz, Ericson, Kessler,

Bhargava, Hagmann, Blumenthal-Barby, D'Aunno, Handel, Kolstad, Nussbaum, Shaffer, Skinner, Ubel, and Zikmund-Fisher, 2017).

Among health policy-makers and practitioners, the use of insights from behavioral economics, and, in particular, of “nudges” has recently led many governments around the world to set up behavioral or “nudge units” within their civil services, starting from the *Behavioural Insights Team* in the UK Cabinet Office, to the analogous initiatives within the UK Department of Health, the NHS, and Public Health England, in the governments of Australia, Canada, Denmark, Finland, France, Israel, Italy, the Netherlands, New Zealand, Norway, Singapore, the US, and in the European Commission (Sunstein, 2011; Dolan and Galizzi, 2014a; Oliver, 2017).

We start with a simple operational definition of “*behavioral experiments in health*”, arguably the first such a definition. We then identify ten key areas of potential debate about behavioral experiments in health, which we think deserve explicit discussion. In what follows, we address one by one each of these ten areas of possible debate and controversy by answering ten corresponding questions. By doing so, we review the state of the art of the different streams and areas of applications of the growing field of behavioral experiments in health; we discuss which significant questions remain to be discussed; and we highlight the rationale and the scope for the further development of behavioral experiments in health in the years to come.

In a nutshell, “*behavioral experiments in health*” make use of a broad range of experimental methods typical of experimental and behavioral economics to investigate individual and organizational behaviors and decisions related to health and healthcare.

The behaviors and decisions considered in behavioral experiments in health therefore usually take place, or are framed, in a health, healthcare, or medical setting or context.

The term “behavioral” in “behavioral experiments in health” requires a first clarification. Common to experimental economics and behavioral science, in fact, the outcomes of behavioral experiments in health are “behavioral” in that they consist of directly observable and measurable behavioral responses or directly revealed preferences, rather than self-reported statements. For example, subjects in behavioral experiments in health are typically observed in real health or healthcare field situations, or, if not, they face real consequences for their choices or behaviors through aligned monetary and non-monetary incentives. Behaviors and decisions of participants to a behavioral experiment in health are thus typically “natural” – that is they take place in naturalistic situations - or “incentive-compatible” in the usual experimental economics sense that participants bear some real behavioral consequences for their choices in the experiment (e.g., Smith, 1976, 1982; Friedman and Sunder, 1994; Cassar and Friedman, 2004). This defining feature makes behavioral experiments in health distinct from “stated preference experiments”, such as contingent valuation studies, or “discrete choice experiments” (DCEs), which have since long been used in health economics, and which do not typically consider real behavior or incentive-compatible choice situations (e.g., Ryan, McIntosh, and Shackley, 1998; Ryan and Farrar, 2000; de Bekker-Grob, Ryan, and Gerard, 2012).

Furthermore, methodologically, behavioral experiments in health purportedly cover the entire continuum spectrum of experimental methods spanning the lab to the field, passing through online and mobile experiments, and experiments pursuing “behavioral data linking” (for more, see questions 3 and 10).

Finally, and following the usual methodological convention in experimental economics, behavioral experiments in health do not deceive subjects. Some behavioral experiments in health can, nonetheless, entail some degree of “obfuscation” when, in the attempt to minimize possible “experimenter demand effects” (Zizzo, 2010), subjects are not told about the exact purpose and research question of the experiment. This is in line with the spirit of those experiments that intend to minimize the alteration of, and interference with, naturally occurring behavior by not telling subjects that they are part of an experiment (that is, in the spirit of “natural field experiments” according to the taxonomy by Harrison and List, 2004, discussed in question 3; or of “lab-field” experiments as in Dolan and Galizzi, 2014a).

To sum up, five characterizing features of behavioral experiments in health are therefore: (i) the fact that the decisions and behaviors are health-related; (ii) the fact that, whenever possible, the outcomes of the decisions in the experiment are “behavioral” in the sense of consisting of directly observable and measurable behavioral responses, or of bearing real consequences for the decision-makers; (iii) the open-minded consideration of principles and insights from both behavioral economics and conventional economics, as well as their combination and integration; (iv) the use of a broad range of experiments spanning the lab to the field, passing through online and mobile experiments, and “behavioral data linking” experiments; and (v) the tendency to avoid deception which, however, does not prevent the use of obfuscation, natural field experiments, and lab-field experiments.

We next review the existing literature by addressing ten areas of current debate about behavioral experiments in health. A few of these areas apply to behavioral experiments more generally and, when this is the case, we explicitly notice this.

Question 1: What can behavioral experiments tell us that non-experimental methods in health cannot tell us already?

First of all, theory, experiments, and econometrics are complements, not substitutes, to each other (Falk and Heckman, 2009; Harrison, Lau and Rutström, 2015; Galizzi, Harrison, and Miraldo, 2017). In particular, the way in which behavioral experiments are sometimes contrasted with econometric analysis is misleading. In fact, running any type of controlled behavioral experiments in health (see more in question 3) is just the first step of the data collection process that should then feed into an appropriate econometric analysis of the experimental data. The broad range of behavioral experiments in health allows the researcher to gather rich data to delve empirically into the behavioral nuances and mechanisms of an observed change in health-related behavior. Indeed, as witnessed by the field of “*behavioral econometrics*”, experiments and econometric analysis are complementary, not substitute, methods (Hey and Orme, 1994; Andersen, Harrison, Lau and Rutström, 2008a, 2010, 2014; Harrison, Lau and Rutström, 2015). A similar point holds for the theoretical underpinnings of a behavioral experiment in health.

Second, a key advantage of behavioral experiments is the ability to tightly control experimental conditions. In physics and social sciences, testing theory is a basic component of experiments and the scientific method relies upon explicit tests of theory (Charness and Kuhn, 2011; Charness and Fehr, 2015). While secondary data are often rich and abundant, at the same time they might be confounded by a variety of environmental factors. For example, a health economist who pursues to test the effect of incentives inherent in performance pay on the physicians’ quality of medical care using secondary data might end up with confounded results

because institutions such as public monitoring and reporting of physicians' quality was introduced at the same time. Using secondary data, disentangling these factors seems prohibitively challenging, if not impossible. Taking a more general perspective, the key strength of behavioral experiments is the ability to test a specific theoretical model. One can then compare the behavioral predictions of the model to what happens. If a specific theory is rejected it is then relatively neat to test competing explanations. For example, rational decision-making theory might not be suitable to explain the inconsistent choices of insurance plans in the US (e.g., Abaluck and Gruber, 2011). Competing behavioral decision-making theories might then be called upon whose alternative explanations can then be tested in controlled experiments. This is consistent with the open-minded approach of behavioral experiments in health, which consider principles and insights from both behavioral economics and conventional economics.

Third, another reason to run experiments is the unique opportunity to study behavior and practices analyzed in the theoretical health economics models that are difficult to observe using field data. An example is the effect of referral payments from specialists to primary care physicians on primary care physicians' referral behavior. While health economic theory (e.g., Pauly, 1979) suggests that referral fees enhance efficiency payments for referrals are largely forbidden, however, in almost all Western healthcare markets. In a lab experiment, Waibel and Wiesen (2017) explicitly test model predictions on physicians' diagnostic effort and referral decisions and find that the introduction of referral payments increase efficiency, although not to the levels predicted by theory. Another example is unethical behavior in health care, for example, diagnostic related groups (DRG) upcoding. Admittedly, at an aggregate level there is plenty of evidence that DRG upcoding exists (e.g., Silverman and Skinner, 2004; Jürges and Köberlin, 2015). However, what drives unethical behaviors is largely unknown. A study by

Hennig-Schmidt, Jürges, and Wiesen (2017) complements field studies on DRG upcoding by analyzing dishonest behavior in a framed experiment in neonatology and by linking dishonest behavior to individuals' characteristics to explore what drives dishonesty. They find that audits and fines significantly reduce dishonest and that subjects' personality traits and integrity relate to dishonest behavior. A further area in health, in which behavioral experiments have contributed to, is to a better understanding of the behavioral effects of professional norms. Exogenously changing professional norms in the field seems prohibitively challenging, and (if possible) drawing inferences seems difficult due to numerous confounding factors. In an online experiment with a large medical student sample, Kesternich, Schumacher, and Winter (2015) analyze the effect of making the Hippocratic oath salient on patient-regarding altruism and distributional preferences. In a series of experiments with physicians (from internal medicine and pediatrics), Ockenfels and Wiesen (2017) investigate the effect of a professional framing on physicians' dishonest behavior (on behalf of themselves and others). Evidence from behavioral experiments in health which are "well-grounded" in theory are therefore not only useful to contrast behavior with model predictions but also to further stimulate the debate among healthcare policy-makers on practices with little or no field evidence.

Fourth, one of the key strengths of experiments, in general, is that a researcher can empirically study the causal effects of different *institutions*, as defined by their rules, actors, and incentives. Thus, behavioral experiments seem ideal to serve as a test bed for analyzing the effect of institutional changes related to health care. Understanding behavioral mechanisms in health-related decisions is imperative *before* designing and implementing large-scale behavioral interventions in the field or *ad hoc* healthcare policy interventions, as there might be unknown or unintended effects for providers and patients alike. In this sense, excluding behavioral experiments from the research toolkit of a health economist would be somewhat similar to

ignoring animal studies for medical or drug research: “*while results from animal studies do not always apply to humans, the ability to test many hypotheses cheaply under carefully controlled conditions provides an indispensable tool for the development of models that work in the real world*” (Charness and Kuhn, 2011, p.233).

Fifth, behavioral results from experiments might not only be insightful to better understand actual health-related decision-making and behavior, but also to inform the development of *behavioral economics theories in health contexts* (e.g. Köszegi 2003, 2006; Frank, 2007; Hansen, Anell, Gerdtham, and Lyttkens, 2015). The observation of actual human behavior in experiments enables the researcher to identify behavioral deviations from theory and thus to extend health economic theories by taking into account issues like human motivation, or behavioral phenomena like emotions or (patient-regarding) altruism. Two prominent research areas in which theory and experiments have already fruitfully complemented each other are: the matching markets for organ donations and for physicians and healthcare professionals (e.g. Roth, 2002; Roth and Peranson, 1999; Kessler and Roth, 2012; Kessler and Roth, 2014a,b; Herr and Normann, 2016; Li, Hawley and Schnier, 2013); and the design of mixed systems of public and private healthcare finance (e.g. Buckley, Cuff, Hurley, McLeod, Mestelman, and Cameron, 2012, 2016).

In sum, running behavioral experiments in health allows the researcher to better understand the causal effects of health-related policy interventions on individual and organizational behavior and to contrast findings with predictions from theoretical models. Behavioral experiments therefore nicely complement and bridge non-experimental methods, in particular theory and empirical econometric analysis, and could therefore help bringing closer together the different health research communities (Galizzi, 2017; Galizzi, Harrison, and Miraldo, 2017).

Question 2: Are behavioral experiments really new to health economics, policy, and management?

No. First, health economists are particularly well aware of the importance of using randomized controlled experiments. Modern evidence-based medicine and pharmacology are all based on randomized controlled trials (RCTs), starting from the pioneering work on scurvy by James Lind in 1747, to the first published RCT in medicine by Austin Bradford Hill and colleagues in 1948. Thanks to the ground-breaking contributions of Charles Sanders Peirce, Jerzy Neyman, Ronald A. Fisher and others, modern science has since long considered randomized controlled experiments as fundamental and important scientific methods. Far from novel, the idea of using randomized controlled experiments has been advocated for decades even for policy applications (Rubin, 1974; Ferber and Hirsch, 1978; Burtless, 1995).

Second, arguably one of the most influential studies in health economics is indeed based on a large-scale randomized controlled experiment. The RAND Health Insurance Experiment conducted in the US between 1974 and 1982, in fact, analyzed the effects of randomly allocated co-payment rates and health insurance contracts on healthcare costs and utilization of healthcare (Newhouse, Manning, Morris, Orr, Duan, Keeler, Leibowitz, Marquis, Marquis, Phelps, and Brook, 1981; Manning, Newhouse, Duan, Keeler, Leibowitz, and Marquis, 1987). As a major finding, Joseph P. Newhouse and colleagues documented that cost sharing reduced the over-utilization of medical care while it did not significantly affect the quality of care received by participating patients.

The spirit and the main features of the RAND Health Insurance Experiment has later inspired the design of the Oregon Health Insurance Experiment. The latter was conducted in 2008 with

uninsured low-income adults in Oregon. Adults allocated to the treatment group were given the chance to apply for Medicaid (via a lottery). This allowed researchers to analyze the effects of expanding access to public health insurance (Medicaid), for example, on the healthcare use and health of low-income adults. They found that the treatment group had substantively and statistically significantly higher healthcare utilization and a better self-reported health than the control group (Finkelstein, Taubman, Wright, Bernstein, Gruber, Newhouse, Allen, and Baicker, 2012; Finkelstein and Taubman, 2015).

The launch of the *Behavioural Experiments in Health Network* ([BEH-net](#)) in 2015 can be seen as the response to a fast-increasing demand to systematically use behavioral experiments in health economics, policy, and management. The network aims precisely at integrating and bringing closer together the research communities at the interface between experimental and behavioral economics, and health economics.

Question 3: What types of experiments are considered when referring to behavioral experiments in health?

There is an important initial conceptual distinction between behavioral experiment in health and Randomized Controlled Trials (RCTs). Many health practitioners and policy-makers, in fact, tend to automatically associate behavioral experiments with RCTs.

In the health policy debate, the term RCT is sometimes used to denote large-scale field experiments conducted with entire organizations (e.g. hospitals, villages) without necessarily allowing the stakeholders in those organizations to explicitly express their views or their consent to the proposed manipulations. This is a major conceptual and practical difference with

respect to proper RCTs in medicine or pharmacology, where subjects are *always* explicitly asked to give informed consent prior to take part into RCTs, and allowed to drop out with important ethical, political, and logistical implications. The term RCT is therefore conceptually inappropriate and practically misleading in a health economics, policy, and management context, since it conveys the false impression that subjects have been made aware of being part of an experiment and have been consulted and given their consent to it, when actually this may not be the case.

Moreover, even in the above narrow and inappropriate connotation, RCTs are only one specific type of experiment, namely field experiments. As mentioned, however, behavioral experiments in health purportedly cover the entire spectrum of experiments from the lab to the field. Harrison and List (2004) proposed an influential taxonomy of experiments along the lab-field spectrum that illustrates the diversity of experiments: i) *conventional lab* experiments involve student subjects, abstract framing, a lab context, and a set of imposed rules; ii) *artefactual field* experiments depart from conventional lab experiments in that they involve non-student samples; iii) *framed field* experiments add to artefactual field experiments a field context in the commodity, stakes, task or information; and, finally, iv) *natural field* experiments depart from framed field experiments in that subjects undertake the tasks in their natural environment, and subjects do not know that they take part into an experiment.

The main idea behind *natural field* experiments is equivalent to von Heisenberg's "uncertainty principle" in physics: the mere act of observation and measurement necessarily alters, to some extent, what is being observed and measured. In key areas for health economics, for example, there may be potential *experimenter demand effects*, where participants change behavior due to cues about what represents "appropriate" behavior for the experimenter (Levitt and List,

2007), for example, when deciding on provision of medical services; *Hawthorne effects*, where simply knowing they are part of a study makes participants feel important and improves their effort and performance (Levitt and List, 2011); and *John Henry effects*, where participants who perceive that they are in the control group exert greater effort because they treat the experiment like a competitive contest and they want to overcome the disadvantage of being in the control group (Cook and Campbell, 1979).

More recently, other types of experiments have been conducted in experimental economics, beyond lab, artefactual field, frame field, and natural field experiments. For example, *virtual experiments* combine controlled experiments with virtual reality settings (Fiore, Harrison, Hughes and Rutström, 2009). While not yet applied to health economics contexts, virtual experiments are a promising approach to make tradeoffs more salient and vivid in health and healthcare decision-making.

Lab-field experiments consist of a first-stage intervention under controlled conditions (in the lab) linked to a naturalistic situation (in the field) where subjects are not aware that their behavior is observed. Lab-field experiments have been used to look at the unintended “*behavioral spillover*” effects of health incentives (Dolan and Galizzi, 2014b; 2015; Dolan, Galizzi and Navarro-Martinez, 2015) or at the *external validity* of lab-based behavioral measures (Galizzi and Navarro-Martinez, 2017).

Virtual experiments and *lab-field experiments* are part of the growing efforts to bridge the gap between the lab and the field in health economics applications (Hennig-Schmidt, Selten and Wiesen, 2011; Kesternich, Schumacher, Winter, 2015). They are also part of the more general “*behavioral data linking*” approach (Galizzi, Harrison, and Miraldo, 2017), that is, the linkage

of behavioral economics experiments with longitudinal surveys, administrative registers, biomarkers banks, apps, mobile devices, scan data, and other big data sources (Andersen, Cox, Harrison, Lau, Rutström and Sadiraj, 2015; Galizzi, Harrison and Miniaci, 2017). Data linkage poses new ethical, practical, and logistical challenges when it seeks to link surveys and behavioral experiments with health records and healthcare registers (Galizzi, Harrison, and Miraldo, 2017). Nonetheless, there is currently an extraordinary, and still largely untapped, potential to apply the experimental approach to an unprecedented host of health and healthcare contexts, and by linking and augmenting behavioral experiments in health with the very rich data sources available in health (see more in question 10).

Taken together, there is *not* one type of experiment for potential health economics and policy purposes. Rather, the broad spectrum of different types of experiments spanning the lab to the field can prove useful and complementary for health applications, as well as the most recent online, mobile, and “behavioral data linking” experiments.

Question 4: Is there a preferred type of behavioral experiment in health?

There is currently no consensus on which specific type of behavioral experiment is superior. The choice of the specific type of experiment depends on the specific research question. Lab, field, online, mobile, and “behavioral data linking” experiments all have strengths and weaknesses, and their relative merits have been systematically discussed elsewhere (Smith, 2002; Harrison and List, 2004; Guala, 2005; Levitt and List, 2007, 2009; Bardsley, Cubitt, Loomes, Moffatt, Starmer, and Sugden, 2009; Falk and Heckman, 2009; Harrison, 2013; Kagel, 2015).

For example, it is generally reckoned that *lab experiments* allow for high *internal validity* because of their ability to tightly control the environment and frame, minimize confounding factors, closely simulate conditions of theoretical models, and replicate past experiments. Furthermore, they provide insights into possible patterns prior to moving into the field, they uncover the mechanisms underlying decisions and behavior, and they require significantly fewer financial, time, and logistical resources than field experiments.

On the other hand, it is generally reckoned that *field experiments* can enhance the *external validity* of experimental results (see more about this in question 5), because observations are made with subjects, environments, situations, tasks, rules and stakes which are closer to the ones occurring in the real world (Brookshire, Coursey, and Schulze, 1987; Galizzi and Navarro-Martinez, 2017). Field experiments, however, come with lesser control and with several other limitations when used for policy purposes (Harrison, 2014). Moreover, they are inherently more difficult to replicate. This is a major limitation given the increasing attention to the *replicability* of experimental results in economics, psychology, and health sciences (Burman, Reed, and Alm, 2010; Dolan and Galizzi, 2014a; Open Science Collaboration, 2015; Camerer, Dreber, Forsell, Ho, Huber, Johannesson, Kirchler, Almenberg, Altmejd, Chan, and Heikensten, 2016; Galizzi, Harrison, and Miraldo, 2017).

More generally, it is important to reiterate the point that the different types of experiments are complementary not substitutes (Falk and Heckman, 2009; Harrison, Lau and Rutström, 2015; Galizzi, Harrison, and Miraldo, 2017). Recent behavioral experiments in health pick up on this important point and combine different types of experiments and thereby test for the validity of findings from lab experiments with conventional student samples. For example, Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, and Wiesen (2016a) combine lab and artefactual field

experiments to analyze the effect of fee-for-service and capitation regimes for medical service provision on the behaviors of medical students, non-medical students, and physicians. Across the board, they found that all subject pools responded to incentives similarly – namely, patients were over-treated under fee-for-service and undertreated under capitation. Physicians, however, responded less to the incentives inherent in these two payment schemes. In experiments with medical and non-medical students, Hennig-Schmidt and Wiesen (2014) and Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, and Wiesen (2017a) report similar findings. Wang, Iversen, Hennig-Schmidt, and Godager (2017) compare behavior of physicians and medical students in China and Germany. Comparing lab experimental data from medical students with artefactual field experimental data from a subsample of a representative sample of German resident physicians, Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, and Wiesen (2017b) find that performance pay crowds out patient-regarding motivation for both subject pools.

Taken together, the experimental studies that systematically account for potential differences in the subject pools indicate that the direction of a treatment effect does not differ between (medical or non-medical) students and medical professionals. Importantly, however, the intensity of a behavioral effect might vary across subject pools. Moreover, experimental designs in health typically abstract from the complexity of real-world settings in order to “isolate” treatment effects. A few experimental studies even employ neutral framings of the decision situation presumably for reasons of control and salience of the incentives: see for example, Green (2014) on provider incentives; Huck, Lünser, Spitzer, and Tyran (2016) on health insurance choice; and Mimra, Rasch, and Waibel (2016) on specialists’ second opinions. As experiments are “scalable”, adding more realism (health context) to these settings by employing medical frames seem desirable: Ahlert, Felder, and Vogt (2012) and Kesternich,

Schumacher, and Winter (2015) find behavioral differences in the intensity of subjects' responses when comparing neutral and medical frames.

In sum, researchers and health care policy-makers alike might clearly need to be cautious in drawing conclusions to real world settings when taking findings from behavioral experiments in health at face value. One might argue, however, that this key point applies to any type of behavioral experiments in health, not just to lab experiments, or to field experiments. Moreover, the existing evidence indicates somewhat similar main directions of experimental treatment effects across subject pools and it is therefore useful to inform debates in healthcare policy and management. For a more general discussion of external validity and generalizability of experimental findings, see the next question.

Question 5: Can we trust the external validity of behavioral experiments in health?

The point on the *external validity* of behavioral experiments in health is too important to be unduly misrepresented or over-simplified as is oftentimes done in the research and policy debate. Most issues related to the external validity of experiments are actually not unique to behavioral experiments in health, but are common to all the economic experiments in general, and for this reason we answer this question more generally.

The main observation is that external validity means different things from different points of view. In a first connotation, external validity refers to the within-subjects question of whether the outcomes of a behavioral experiment in health are representative of the corresponding outcomes of interest that would occur outside of the behavioral experiment for the same pool of subjects. From this perspective, as mentioned, external validity is often contrasted with

internal validity on the presumed ground that there is always an inherent trade-off between internal and external validity while one moves from the lab end to the field end of the spectrum in the Harrison and List (2004) taxonomy. This, however, is not always nor necessarily the case. In fact, to start with, if rigorously designed and conducted, *all* randomized controlled behavioral experiments in health are internally valid, whether they are lab, artefactual field, framed field, or natural field experiments. So, it is simply not true that internal validity is necessarily higher in lab than in natural field experiments. Moreover, it is also not true that the external validity (in the above explained connotation) is necessarily higher in natural field than in lab experiments. It obviously depends on how close is the correspondence between the outcomes measured in the behavioral experiment and the outcomes of interest that would occur outside of the behavioral experiments. In other words, it depends on what the final outcomes of interest are, and, ultimately, what the research question is. For example, imagine that the main outcomes of interest is about looking at how many calories subjects eat in a buffet, or how much time they wait until lighting up the next cigarette, which are both likely to be the outcomes of some automatic or “visceral” decision-making occurring without any conscious deliberation (Loewenstein, 1996). Then, a natural field experiment, where subjects do not know they are part of an experiment, would be a natural setting where to observe such behaviors (Dolan and Galizzi, 2014b). On the other hand, however, imagine that the main outcomes of interest is about studying how subjects trade-off and choose between different private health insurance schemes, or which groceries subjects purchase in an online supermarket, two highly deliberate decisions which are both likely to take place in online settings even when they occur outside of a behavioral experiment. Then, a conventional lab or an online experiment would be a natural setting where to observe such behaviors.

In a second, slightly different, connotation, external validity refers to the, still within-subjects, question of whether the outcomes of a behavioral experiment in health are good predictors of the corresponding outcomes of interest that would occur outside of the behavioral experiment for the same pool of subjects. For example, are healthy food choices in a behavioral experiment good predictors of a subject's healthy diet? Are experimental decisions over drugs, treatments, health insurances good predictors of analogous decisions outside the experiment? It is true that, in principle, the experimental decisions can move closer to the decisions taken outside the experiment as one moves from lab experiments to natural field experiments in the Harrison and List (2004) taxonomy. But it is also true that this depends, again, on what the ultimate outcomes of interest and research questions are. And, in principle, also this type of external validity question can affect the entire spectrum of behavioral experiments in health, from lab to natural field experiments. In fact, the only rigorous strategy to empirically address this question is to design and implement a longitudinal augmentation of the original behavioral experiment that enables to follow up over time in more naturalistic settings the same pool of subjects. When implemented in a systematic and transparent way, this strategy also allows the researcher to overcome a major limitation of the very few external validity analyses to date, namely the fact that they typically are *ad hoc* analyses. The typical analysis, in fact, reports the correlation between one specific experimental outcome and one specific variable outside the experiment and, when such a correlation is found to be statistically significant, it concludes that what found in the experiment is externally valid. Such an approach, however, lacks of systematization because it fails to provide full information on the whole set of pairwise correlations between *all* the experimental outcomes and *all* the variables outside the experiment, being they significant or not significant. Only a systematic and transparent testing and reporting of *all* such correlations would be the litmus test of the external validity of a behavioral experiment. Such exercises are rare in experimental economics and virtually not existent in health.

In a non-health context, for example, Galizzi and Navarro-Martinez (2017) systematically assess and report the associations between the whole set of eight social preferences experimental games and five different pro-social variables outside of the experiment, and found that only one out of forty pairwise correlations is statistically significant (none is if properly correcting for multiple hypothesis testing). They then relate their finding to a systematic review and meta-analysis of all the, published and unpublished, studies that have previously tested the external validity of those same experimental games, and conclude that the often proclaimed external validity of social preferences games is not supported by the empirical evidence: only the 39.7% of the reported pairwise correlations and 37.5% of the reported regressions find a statistically significant association between an experimental outcome and a variable outside the experiment.

The behavioral experiments in health literature still lacks a similar systematic and transparent approach to this dimension of the external validity question and, given the importance of this exercise for both research and policy purposes, we encourage more research in this direction (Galizzi, Machado, and Miniaci, 2016). For example, the lack of transparency and systematization is the main explanation behind the current sterile debate on whether lab-based behavioral economics measures for risk, time, and social preferences are externally valid in the health context. Very few studies in behavioral health economics have tied their hands by publishing a pre-analysis plan or a public protocol that clearly and explicitly states at the outset of the analysis which health behaviors in the field will be associated with the lab-based behavioral economics measures. Most studies only report *ad hoc* subsets of the correlations and regressions between those measures and the health behaviors, and do not report nor discuss how these results relate and compare to the whole set of statistically significant and not

significant associations. Moreover, systematic replication is almost inexistent in behavioral experiments in health. It is even argued by strong proponents of the lab experiments that preferences can only be measured in the lab. This is tantamount to state that the question of whether lab-based measures for those preferences are externally valid is a non-falsifiable question, which is the opposite of an evidence-based scientific approach to this key matter. If behavioral experiments in health are about to leave infancy for adulthood, they should better take seriously the lessons learned in the neighboring disciplines of medicine and health studies, where collective knowledge is systematically cumulated only through transparent replications, systematic reviews, and meta-analyses, as epitomized by major collective research infrastructures such as the *Cochrane Collaboration* or the *Campbell Collaboration*.

A third connotation of the external validity has to do with whether the outcomes of a behavioral experiment in health are representative of the corresponding outcomes of the population of interest. This “out-of-sample” connotation of external validity clearly requires that the pool of subjects involved in the behavioral experiment in health has been drawn from a representative sample of the population of interest. So, for example, if the behavioral experiment aims at concluding something about decisions or behaviors of medical doctors, it should involve a representative sample of medical doctors, while if it aims at inferring anything at a population level, the behavioral experiment should involve a representative sample of the population.

Moreover, the debate about the external validity of behavioral experiments in health should be more generally conducted within the broader framework of the debate in terms of *generalizability* — that is, the question of which other populations, settings, contexts or domains the findings from an experiment can be generalized to (Al-Ubaydli and List, 2015).

Importantly, the generalizability question equally applies to the whole spectrum of behavioral experiments in health, from lab to natural field experiments.

There are three conceptually distinct threats to generalizability of behavioral experiments in health. The first threat comes essentially from participation bias. Unlike natural field experiments, lab, artefactual field, and framed field experiments recruit subjects through an explicit invitation to take part in an experiment. As a result, there is bias because subjects who choose to participate in experiments may be inherently different in their underlying characteristics from subjects who choose not to take part. Health researchers and policy-makers should therefore be aware that, because of the participation bias, even if the initial sample of subjects is indeed representative of the target student population, the resulting sub-sample of actual respondents may not be.

The second threat comes from the fact that the environment, context, and frame of the experimental decisions and tasks in the lab may not be representative of real situations encountered by subjects in natural health and healthcare settings. This limitation can be easily overcome by redesigning tasks and contexts to more closely match naturalistic situations that subjects are more familiar with in real life — that is, to design framed field experiments in the sense of Harrison and List (2004). This strategy has been extensively employed in other application areas in experimental economics (e.g. Harrison, List, and Towe, 2007; Harrison and List, 2008), and has already been explored in the health economics area (e.g. Hennig-Schmidt, Selten, and Wiesen, 2011; Hennig-Schmidt and Wiesen, 2014; Galizzi, Miraldo, Stavropoulou, and van der Pol, 2016; Eilermann, Halstenberg, Kuntz, Martakis, Roth, and Wiesen, 2017).

The last threat to generalizability is that experimental subjects may not be representative of the general population, especially when are students or medical students (Levitt and List, 2007). To overcome this limitation, behavioral economists have started running *artefactual field experiments* with *representative samples* of the population (Harrison, Lau and Rutström, 2002, 2007; Andersen, Harrison, Lau and Rutström, 2008a, 2014; Galizzi, Machado, and Miniaci, 2016; Galizzi, Harrison and Miniaci, 2017). This is a promising avenue for behavioral experiments in health, given that the goals and priorities in designing health policies and health systems are typically set at a population level (Michie, 2008).

From the broader generalizability perspective, we can hardly see why the results of a natural field experiment with, say, female nurses in Tanzania, or health insurance customers in the rural Philippines should be considered as more generalizable than a lab experiment with medical students in Germany, or an artefactual field experiment with a representative sample of the population in the UK. Too often similar claims even forget to state what the population of interest is for the study.

More generally, Falk and Heckman (2009) state that causal knowledge requires a controlled variation. Whether a variation from, for example, a natural field experiment or a more controlled lab experiment is more informative depends on the research question and is still debated among researchers in social sciences. It is important to acknowledge, again, that empirical methods and different sources of data are complements. For example, both behavioral experiments, spanning the lab to the field, and econometric analyses of secondary data can all improve the state of knowledge in health economics research, with the issue of generalizability of results being applicable to all of them.

Taken together, behavioral and experimental health economists should take seriously the external validity and generalizability challenges by open-mindedly using all types of

experiments in the lab-field spectrum, by embracing a transparent and systematic approach in gathering and reporting evidence, including reporting of all the statistically significant and not significant correlations and regressions (rather than cherry-picked subsets of the positive results). We see this as a fundamental requisite for behavioral experiments in health as a field moving, in the years to come, from infancy to adulthood.

Question 6: What about experiments to elicit preferences in health?

One of the above-discussed defining features of behavioral experiments in health is that they entail directly observable and measurable behavioral responses. For example, experimental decisions, tasks, and measures to elicit preferences and willingness-to-pay are incentive-compatible in the sense that subjects pay some real consequence in terms of monetary or non-monetary outcomes for the choices they make. This raises the question of whether or not behavioral experiments in health also include the experimental studies that aim at eliciting *health-related* preferences.

Some distinctions should be made on this point. On the one hand, there is a vast literature in health economics that uses popular experimental methods such as the Standard Gamble (SG) or the Time Trade Off (TTO) to elicit preferences for hypothetical health states (Bleichrodt and Johannesson, 2001; Bleichrodt, 2002; Attema and Brouwer, 2012). Given the *hypothetical* nature of the choices about different health states, these experiments are similar in nature to the already discussed “stated preference experiments”, such as contingent valuation studies or “discrete choice experiments” (DCEs), which do not typically consider real behavior or incentive-compatible choice situations. Using the same argument, therefore, the experiments in this literature should not be considered behavioral experiments in health.

On the other hand, there is also a small, but growing, literature looking at the relationships between incentive-compatible experimental measures of risk and time preferences and health-related behaviors. Harrison, Lau and Rutström (2010), for example, elicit risk and time preferences of a representative sample of the Danish population and find no difference in the likelihood of smokers and non-smokers to exhibit hyperbolic discounting, no significant association of smoking with risk aversion among men, and no significant association of smoking with discount rates among women. Galizzi and Miraldo (2017) measure the risk preferences of a convenience sample of students and find that, while there is no association between smoking or BMI with the estimated risk aversion, the latter is significantly associated with the *Healthy Eating Index*, an indicator of overall nutritional quality. Harrison, Hofmeyr, Ross and Swarthout (2015) elicit risk and time preferences of a convenience sample of students at the University of Cape Town, and find that smokers and non-smokers differ in their baseline discount rates, but do not significantly differ in their present bias, risk aversion, or subjective perception of probabilities. In a longitudinal experiment with a representative sample of the UK population, Galizzi, Machado, and Miniaci (2016) systematically assess the external validity of different measures of risk preferences linked to the *UK Longitudinal Household Survey* (UKHLS), and find that the experimental measures are not significantly associated to subjects' BMI, eating, smoking, or drinking habits. Several other *ad hoc* analyses have associated risk and time preferences with heavy drinking (Anderson and Mellor, 2008), BMI (Sutter, Kocher, Glätzle-Rützler, and Trautmann, 2013), and the uptake of vaccinations, preventive care, and medical tests (Chapman and Coups, 1999; Bradford, 2010; Bradford, Zoller and Silvestri, 2010).

Given that all these latter experiments use incentive-compatible methods to elicit risk and time preferences, they should be considered as behavioral experiments in health. A common aspect of the latter group of experiments, however, is that they measure individual risk and time preferences over (risky or inter-temporal) monetary outcomes, and then link these to health-related behaviors. But what about the studies that elicit individual risk and time preferences for *health* outcomes, rather than for monetary outcomes?

We see the experiments eliciting risk and time preferences in health as an interesting middle ground between stated preferences experiments and behavioral experiments. When it comes to the measurement of risk and time preferences in the health domain, in fact, the current community of behavioral health experimentalists interprets behavioral experiments in health with a fair degree of tolerance, flexibility, and open-mindedness, and considers the elicitation of risk and time preferences in health a research field that is closely aligned with, and affine to, the core interests and methods of behavioral experiments in health.

This is not because the community disagrees with the traditional experimental economics view that answers to hypothetical questions can significantly differ from responses to incentive-compatible tests because “talk is cheap” if there are no real behavioral consequences (Battalio, Kagel and Jiranyakul, 1990; Cummings, Harrison and Rutström, 1995; Cummings, Elliott, Harrison and Murphy, 1997; Holt and Laury, 2002; Harrison, 2006). Moreover, behavioral health experimentalists are all well aware that, from a theoretical perspective, risk and time preferences are fundamental individual characteristics at the core of health behavior and decision-making (Gafni and Torrance, 1984). Risk and time preferences, in fact, directly inform the principles and practices of cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) in healthcare, and the assumptions beyond the Quality Adjusted Life Years

(QALY), the measure of health benefits that is commonly employed in CEA and CUA, and that relies on the above mentioned SG and TTO methods (Bleichrodt, Wakker and Johannesson, 1997).

Rather, it is because at the moment the literature on behavioral experiments in health lacks a systematic body of state-of-the-art consensus methods to measure health-related preferences with real non-monetary consequences. Given the fundamental role of risk and time preferences in the health context, it is actually surprising that there is no consensus to date on a “gold standard” measurement methodology.

A multitude of different methods have been proposed to measure risk and time preferences in health contexts, which are heterogeneous in terms of underlying theoretical frameworks, methodological features, and links to formal econometric analysis (Galizzi, Harrison, and Miraldo, 2017). A major challenge in converging to a consensus methodology to measure risk and time preferences in health is related to the fact that, to date, the different proposed methods are substantially disconnected. On the one hand, the current methods to measure preferences for health outcomes only entail hypothetical scenarios. On the other hand, all the incentive-compatible methods to measure preferences with real consequences are based on monetary outcomes. From both a conceptual and an empirical point of view, however, it is unclear whether individual risk and time preferences are stable across the health and the monetary domains (Chapman, 1996).

There have been a number of exploratory analyses of whether these preferences are indeed stable across the finance and the health domains. Galizzi, Miraldo and Stavropoulou (2016), for example, summarize the relatively limited number of studies that compare risk taking across

the health and other domains, and find that, despite the broad heterogeneity of methods and frames used in the literature, there is general evidence that there are differences across domains, and that these differences also emerge when real consequences are at stake.

The elicitation of risk and time preferences with incentive-compatible methods in the health domain is a promising and challenging task, and a collective priority in the research agenda of behavioral experimentalists in health. We expect this methodological and substantial gap to be filled soon by the increasingly collaborative community of behavioral experimentalists in health.

Question 7: Which topics are addressed by behavioral experiments in health?

A first area of experimental research that has recently received considerable attention are “nudges”, that is, changes in the “choice architecture” made to induce changes in health behavior, mainly at an unconscious or automatic level (Thaler and Sunstein, 2008). In the spirit of “asymmetric paternalism” (Loewenstein, Brennan, and Volpp, 2007; Loewenstein, Ash, and Volpp, 2013), many behavioral experiments have in fact applied nudges to health and healthcare behavior spanning from risky behaviors in adolescents (Clark and Loheac, 2007) to exercise (Calzolari and Nardotto, 2016), from food choices (Schwartz, Riis, Elbel, and Ariely, 2012; Schwartz, Mochon, Wyper, Maroba, Patel, and Ariely, 2014; Milkman, Minson, and Volpp, 2014; VanEpps, Downs, and Loewenstein, 2016a,b) to drugs compliance (Vervloet et al. 2012), from medical decision making (Schwartz and Chapman, 1999; Brewer, Chapman, Schwartz, and Bergus, 2007; Ansher, Ariely, Nagler, Rudd, Schwartz, and Sha et, 2014) to dentists’ services (Altman and Traxler, 2014).

There is, however, much more in behavioral experiments in health than just nudging (Galizzi, Harrison, and Miraldo, 2017; Oliver, 2017). Behavioral experiments in health can, in fact, uncover the behavioral mechanisms beyond the change in health behavior, and thus inform the design and the implementation of a series of other types of health policies including informational campaigns, salient labeling and packaging of healthy food items, the use of financial and non-financial incentives, the design of effective tax, subsidy, health insurance plans, and regulatory schemes (Galizzi, 2014, 2017).

In fact, a broad spectrum of lab to natural field experiments have already been applied to a variety of health economics, policy, and management areas, well beyond “nudges.” For example, behavioral experiments in health have investigated the effects of different co-payment rates and health insurance contracts on healthcare utilization and costs (Newhouse, Manning, Morris, Orr, Duan, Keeler, Leibowitz, Marquis, Marquis, Phelps, and Brook, 1981; Manning, Newhouse, Duan, Keeler, Leibowitz, and Marquis, 1987); the effects of public health insurance coverage on healthcare utilization and health outcomes (Finkelstein, Taubman, Wright, Bernstein, Gruber, Newhouse, Allen, and Baicker, 2012; Baicker, Taubman, Allen, Bernstein, Gruber, Newhouse, Schneider, Wright, Zaslavzky, and Finkelstein, 2013; Finkelstein and Taubman, 2015; Finkelstein, Taubman, Allen, Wright, and Baicker, 2016); the effects of different providers’ incentives and the role of altruism (Fan, Chen, and Kann, 1998; Hennig-Schmidt, Selten and Wiesen, 2011; Alhert, Felder, and Vogt, 2012; Godager and Wiesen, 2013; Green, 2014; Hennig-Schmidt and Wiesen, 2014; Kesternich, Schumacher and Winter, 2015; Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz and Wiesen, 2016a, 2016b, 2017a, 2017b; Kokot, Brosig-Koch and Kairies-Schwarz, 2017); the role of audit, transparency, compliance, and gender bias in healthcare management (Godager, Hennig-Schmidt and Iversen, 2016; Jakobsson, Kotsadam, Syse, and Øien, 2016; Hennig-Schmidt,

Jürges, and Wiesen, 2017; Lindeboom, van der Klaauw, and Vriend, 2017); the role of different healthcare financing policies (Buckley, Cuff, Hurley, McLeod, Mestelman, and Cameron, 2012, 2015, 2016); two-part tariffs for physician services (Greiner, Zhang and Tang, 2017); provider competition (Brosig-Koch, Hehenkamp and Kokot, 2017; Kairies-Schwarz, Han, and Vomhof, 2017) the matching markets for organ donations and for physicians and healthcare professionals (Roth, 2002; Roth and Peranson, 1999; Kessler and Roth, 2012; Kessler and Roth, 2014a,b; Herr and Normann, 2016; Li, Hawley and Schnier, 2013); the role of subsidies for diagnostic tests and new health products (Dupas, 2014a,b; Cohen, Dupas, and Schaner, 2015; Duflo, Dupas, and Kremer, 2015; Dupas, Hoffman, Kremer, and Zwane, 2016); the choice of health insurance (Schram and Sonnemans, 2011; Buckley, Cuff, Hurley, McLeod, Nuscheler, and Cameron, 2012; Kesternich, Heiss, McFadden and Winter, 2013; Huck, Lünser, Spitzer, and Tyran, 2016; Kairies-Schwarz, Kokot, Vomhof, and Weßling, 2017); the economic and behavioral determinants of vaccination (Milkman, Beshears, Choi, Laibson, and Madrian, 2011; Tsutsui, Benzion, & Shahrabani, 2012; Bronchetti, Huffman, and Magenheimer, 2015; Massin, Ventelou, Nebout, Verger, and Pulcini, 2015; Böhm, Betsch, and Korn, 2016; Böhm, Meier, Lars Korn, Betsch, 2017; Binder and Nuscheler, 2017); the effects of different types of HIV risk information and of SMS interventions on HIV treatment adherence (Dupas, 2011; Rana, Haberer, Huang, Kambugu, Mukasa, Trirumurthy, Wabukala, and Linnemayr, 2015); the use of financial incentives for smoking cessation (Volpp et al., 2006, 2009; Gine', Karlan, and Zinman, 2010; Halpern, French, Small, Sausgiver, Harhay, Audran-McGovern, Loewenstein, Brennan, Asch, and Volpp, 2015; Halpern, French, Small, Sausgiver, Harhay, Audran-McGovern, Loewenstein, Asch, and Volpp, 2016), physical exercise (Charness and Gneezy, 2009; Royer, Stehr, and Sydnor, 2015), weight loss (Volpp, John, Troxel, Norton, Fassbender, and Loewenstein, 2008; John, Loewenstein, Troxel, Norton, Fassbender, and Volpp, 2011; Rao, Krall, and Loewenstein, 2011; John, Loewenstein, and Volpp, 2012;

Kullgren, Troxel, Loewenstein, Asch, Norton, Wesby, Tao, Zhu, and Volpp, 2013; Kullgren, Troxel, Loewenstein, Norton, Gatto, Tao, Zhu, Schofield, Shea, Asch, Pellathy, Driggers, and Volpp, 2016), healthy eating (Loewenstein, Price, and Vopp, 2016), warfarin adherence (Volpp, Loewenstein, Troxel, Doshi, Price, Laskin, and Kimmel, 2008; Kimmel, Troxel, Loewenstein, Brensinger, Jaskowiak, Doshi, Laskin, and Volpp, 2012; Kimmel, Troxel, French, Loewenstein, Brensinger, Meussner, Doshi, Laskin, and Volpp, 2016), glucose control (Long, Jahnle, Richardson, Loewenstein, and Volpp, 2012), home-based health monitoring (Sen, Sewell, Riley, Stearman, Bellamy, Hu, Park, Yao, Zhu, Loewenstein, Asch, and Volpp, 2014), mental exercises (Schofield, Loewenstein, Kopisc, and Volpp, 2015), immunization coverage (Banerjee, Duflo, Glennerster, and Kothari, 2010), nursing services (Banejee, Duflo, and Glennerster, 2007), and medical drugs (Samper and Schwartz, 2013); the unintended carryover and spillover effects of financial incentives and nudges in health (Muller et al., 2009; Wisdom, Downs, and Loewenstein, 2010; Chiou, Yang, Wan, 2011; Dolan and Galizzi, 2014, 2015; Dolan, Galizzi, and Navarro-Martinez, 2015); the behavioral effect of decision support systems and feedback (Cox, Sadiraj, Schnier, and Sweeney, 2016a; Eilermann, Halstenberg, Kuntz, Martakis, Roth, and Wiesen 2017); the elicitation of risk and time preferences in health and their links with health-related behaviors (Cairns, 1994; Cairns and Van der Pol, 1997; Chapman and Elstein, 1995; Dolan and Gudex, 1995; Chapman, 1996; Allison, Kiefe, Cooke, Gerrity, Oray, and Centor, 1998; Chapman and Coup, 1999; Van der Pol and Cairns, 1999, 2001, 2002, 2008, 2011; Bleichrodt and Johannesson, 2001; Anderson and Mellor, 2008; Attema and Brouwer, 2010, 2012, 2014; Bradford, 2010; Bradford, Zoller, and Silvestri, 2010; Harrison, Lau and Rutström, 2010; Attema, 2012; Attema, Bleichrodt, and Wakker, 2012; Szrek, Chao, Ramlagan, and Peltzer, 2012; Attema and Versteegh, 2013; Michel-Lepage, Ventelou, Nebout, Verger, and Pulcini, 2013; Sutter, Kocher, Glätzle-Rützler, and Trautmann, 2013; Harrison, Hofmeyr, Ross and Swarthout, 2015; Galizzi, Machado and Miniaci, 2016;

Galizzi, Miraldo, Stavropoulou, 2016; Galizzi, Miraldo, Stavropoulou, and van der Pol, 2016; Galizzi and Miraldo, 2017; Garcia Prado, Arrieta, Gonzalez, and Pinto Prades, 2017); the elicitation of preferences for retransplantation (Ubel and Loewenstein, 1995), and end-of-life decisions (Halpern, Loewenstein, Volpp, Cooney, Vranas, Quill, McKenzie, Harhay, Gabler, Silva, Arnold, Angus, and Bryce, 2013).

Nudges are just *one* of the many areas of applications. They tend to be relatively effective in the health contexts where people suffer from “internalities”, costs that we incur in because we fail to account for our future selves (Herrnstein, Loewenstein, Prelec, and Vaughan, 1993). Many other health situations are, however, also affected by externalities. Other policy tools, such as taxes, subsidies, and regulatory interventions, have been documented to deal effectively with externalities in health markets (Bhargava and Loewenstein, 2015; Galizzi, 2017). The application of behavioral experiments to these policy areas is at the moment almost inexistent, and we foresee an increase of applications in this key area.

Another related aspect for both research and policy purposes is concerned about the rationale informing the legitimacy of nudging people. There is a major gap in the literature in measuring the underlying preferences before the nudging interventions take place, and in monitoring their evolution (if any) before and after being nudged. This would help understanding heterogeneity in behavioral change, as well as identifying the behavioral channels, mechanisms, and mediating factors that are activated when individuals’ behavior is nudged. At the same time, it would inform the design of target-specific nudges, incentives, and behavioral regulatory tools (thus advancing the state-of-the-art evidence beyond knowing just “what works”). The issue of which set of preferences should be considered for drawing a welfare analysis of nudges and other behavioral interventions is one of the most relevant and pressing open questions from

both a conceptual and an empirical perspective, as well as another area of promising development for the next waves of behavioral experiments in health.

Question 8: How do framing and subject pool matter in behavioral experiments in health when analyzing healthcare professionals' behavior?

While a neutral framing of the experimental decision is appropriate in an experiment on decision-making in games of strategic interactions, a medical framing appears natural for behavioral experiments on decision-making in medical contexts. Kesternich, Schumacher, and Winter (2015) show that framing in a health context affects subjects' behavior in modified dictator and trilateral distribution games. In particular, in their health frame subjects decide in the role of physicians on the provision of medical services with consequences for real patients outside the lab (similar to Hennig-Schmidt, Selten, and Wiesen 2011) in the modified dictator game, whereas in a trilateral distribution game consequences for the insured bearing the cost of medical service provision are also added.

More generally, recent practice in behavioral experiments in health is aligned with the belief that unless you frame a decision situation – for example in a medical or health frame – a researcher cannot be sure how subjects in an experiment have perceived the decision situation (Harrison and List, 2008; Galizzi and Navarro-Martinez, 2017). It may thus be crucial in chosen effort tasks to set subjects in the context the experimenter wants to study in order to avoid the possibility that subjects may impose a context on the abstract experimental task that is different from the experimenter's intended context (e.g., Harrison and List, 2004; Engel and Rand, 2014). As Harrison and List (2004) notice, *“it is not the case that abstract, context-free experiments provide more general findings if the context itself is relevant to the performance of the subjects”*

(p. 1022).

It remains unclear however, whether a change in behavior due to framing more or less accurately reflects true behavior of healthcare professionals. One may argue that a healthcare professional in a health-framed study may be more willing to forgo earnings to avoid looking bad (“experimenter demand effect”, Zizzo, 2010). In practice, healthcare professionals in a neutrally framed decision situation could even become less responsive to how choices directly affect patients than when facing a series of choices in a framed task. Further, individuals may value their health and the health of others differently than any other good. Therefore, both unframed and framed experiments may misinterpret choices by healthcare professionals. For these reasons, it is important to study whether a medical framing in experiments more accurately reflects behavior in healthcare delivery (Kesternich, Schumacher, and Winter, 2015; Cox, Green, and Hennig-Schmidt, 2016b).

Further, different subject pools used in health-related experiments (non-medical students, medical students, physicians) may significantly change behavioral results, with non-medical students exerting less patient-regarding altruism (Hennig-Schmidt and Wiesen, 2014; Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, and Wiesen, 2017a). Considering fraudulent behavior in a routine task in neonatal intensive care units (entry of weights in the birth reports), Hennig-Schmidt, Jürges, and Wiesen (2017) found some evidence for more honest behavior of medical students compared to economics students. A few studies with healthcare professionals and medical students in developing countries correlate neutrally-framed social preferences with actual health-related behaviors (e.g., Kolstad and Lindkvist, 2012; Brock, Lange, and Leonhard, 2014).

Taken together, three promising avenue for behavioral experiments in health on this issue are (i) rigorously and systematically testing the behavioral effects of framing and subject pools; (ii) extending initial findings from the laboratory to field experiments, ideally with healthcare professionals (Cox, Sadiraj, Schnier, and Sweeney, 2016a; Eilermann, Halstenberg, Kuntz, Martakis, Roth, and Wiesen 2017; Leeds, Sadiraj, Cox, Gao, Pawlik, Schnier, and Sweeney, 2017); and (iii) linking findings from behavioral experiments to actual health-related behaviors. In this sense, it seems thus appropriate to call, again, for more systematic evidence – also from healthcare systems in developed countries – to be able to gather more conclusive predictions of providers’ behavior in the field.

Question 9: Is health really different from other policy domains?

The specificity of health as a policy domain is self-evident. On the one hand, health is a very special area of policy application for obvious ethical and political reasons, and even more so for the application of randomized controlled behavioral experiments. Health, moreover, is a research and policy area that is uniquely rich in data: think about the millions of yearly entries in healthcare records and administrative registers (e.g. the *Health Episodes Statistics* in the UK); the large epidemiological cohorts and clinical randomized controlled trials; the complex databanks containing the genetic and epigenetic profiling at a population level. It is also unclear from both a conceptual and an empirical point of view whether behaviors and decisions in the health domain merely reflect behaviors and decisions in other domains in life, for example in the financial domain. As mentioned, the small experimental literature on cross-domains preferences seems to suggest that preferences are not stable across the health and the monetary domains. Also the literature on the use of financial incentives in health finds that their effects are less straightforward and universally applicable than in other fields of applications (Gneezy,

Meier, and Rey-Biel, 2011). A more specific example is about *health care provider* incentives. Recent experimental findings suggest that health care providers' behavior seems to be affected by pay for performance pay, but that this might also lead to adverse effects such as motivation crowding-out (e.g., Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, and Wiesen 2016a, 2017b, Oxholm, 2016). The latter pattern is, for example, not observed in other working domains (admittedly with different performance schemes), such as in field experiments with teachers (e.g., Muralidharan and Sundararaman, 2011). Therefore, a note of caution should be in order when extrapolating lessons from experiments in other fields and to generalize them to the health domain.

On the other hand, behavioral health economists should be careful in advocating a complete disconnection of health applications from other areas of application of behavioral and experimental economics. On the contrary they should continue arguing that there is much that can be learned from health applications, which is useful to other policy domains. This can also help to reduce the substantial disengagement between economic and medical journals. For example, generalist economic journals seem to regularly publish behavioral experiments in, for example, education, financial savings, energy consumption more regularly than in health, which are sometimes dismissed as "too field-specific". That health is of more, not less, general interest than other sub-fields of economics is directly confirmed by the stellar impact factor and international reputation and visibility of the medical journals.

Question 10: What can behavioral experiments in health tell us about long-term effects?

It is true that, at the moment, there is very little evidence on the long-term carryover effects and on the cross-behavioral spillover effects of nudges, incentives, and other health policy

interventions (Dolan and Galizzi, 2014a, 2015; Dolan, Galizzi, and Navarro-Martinez, 2015). This is also due to the fact that, in practice, it is difficult to design behavioral experiments that follow up subjects over time for periods of time longer than a couple of hours (in the lab) or a few weeks or months (in the field), or that track all the complex ramifications of an initial policy intervention on the whole set of targeted and non-targeted health behaviors.

There is, more generally, a sort of major gap and disconnection between two key sources of empirical evidence in health economics. On the one hand, the behavioral experiments in health are typically conducted with small samples of subjects and almost invariably centered around a single observation window or a single data collection point. On the other hand, very comprehensive longitudinal datasets exist in health in forms of administrative records for healthcare access (e.g. the *Health Episodes Statistics* in the UK), biomarkers banks (e.g., the *UK BioBank* in the UK), or medical records and biomarkers for epidemiological cohorts (e.g., *Constances* in France).

The time seems ripe to systematically link and integrate these two major data sources. The already mentioned recent spring of experiments on “*behavioral data linking*” has showed that it is indeed feasible to link and merge behavioral economics experiments with other data sources such as longitudinal surveys, online panels, administrative records, biomarkers and epigenetics banks, apps and mobile devices, smart cards and scan data, clinical randomized controlled trials, and other big data sources (Andersen, Cox, Harrison, Lau, Rutström and Sadiraj, 2015; Galizzi, Harrison and Miniaci, 2017). Given the inherent data-richness of health as a research and policy domain, we expect behavioral data linking to become a key building block of the next generation of behavioral experiments in health. This will contribute to further integrate and cross-fertilize insights, tools, and methods from behavioral, experimental, and

health economics, and to shape up a ground-breaking inter-disciplinary area at the interface between behavioral, medical, and data sciences.

Conclusions

In this article, we have reviewed the state of the art of behavioral experiments in health by critically discussing ten key areas of potential debate and misconception, by highlighting their theoretical and empirical rationale and scope, and by discussing the significant questions which remain.

As our discussions indicate, there are many areas within health economics where experimental methods can be applied fruitfully. To date, in fact, a broad spectrum of behavioral experiments from the lab to the field have already been applied to numerous different health-related areas including, for example, the effects of different co-payment rates and health insurance contracts on healthcare utilization and costs; the effects of public health insurance coverage on healthcare utilization and health outcomes; the effects of different providers' incentives and the role of altruism; the role of audit, transparency, compliance, and gender bias in healthcare management; the role of different healthcare financing policies; the matching markets for organ donations and for physicians and healthcare professionals; the role of subsidies for diagnostic tests and new health products; the choice of health insurance; the economic and behavioral determinants of vaccination; the effects of different types of HIV risk information and of SMS interventions on HIV treatment adherence; the use of financial incentives and nudges for smoking cessation, physical exercise, weight loss, healthy eating, warfarin adherence, glucose control, home-based health monitoring, mental exercises, immunization coverage, and medical drugs; the unintended carryover and spillover effects of financial incentives and nudges in

health; the behavioral effect of decision support systems and feedback; the elicitation of risk and time preferences in health and their links with health-related behaviors; the elicitation of preferences for re-transplantation and end-of-life decisions. Many other health-related areas are expected to follow in the next years in both developing and developed countries.

Tailoring and fine-tuning the broad spectrum of lab, field, online, mobile, and “behavioral data linking” experiments in order to address pressing health policy challenges and key research questions is, both methodologically and substantially, one of the most promising and exciting areas of applications of behavioral experiments to health economics. Also via the new international networks, the next cohort of behavioral experiments in health is likely to originate from a closer collaboration among behavioral and experimental economists, health economists, medical doctors, and decision-makers in health policy and management. This forthcoming generation of behavioral experiments in health will likely scale up the current endeavors to systematically link behavioral economics measures to other data sources, data of which health is naturally rich. In the next years to come, the promise and the research and policy impact of behavioral experiments in health is only destined to grow.

References

- Abaluck, J., & Gruber, J. (2011). Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review*, 101(4): 1180-1210.
- Ahlert, M., Felder, S., & Vogt, B. (2012). Which patients do I treat? An experimental study with economists and physicians. *Health Economics Review*. 2, 1-11.

- Allison, J.J., Kiefe, C.I., Cook E.F., Gerrity M.S., Orav, E.J., & Centor, R. (1998). The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Medical Decision Making*, 18, 320-329.
- Altmann, S. & Traxler, C. (2014). Nudges at the dentist. *European Economic Review*. 72, 19-38.
- Al-Ubaydli, O., & List, J.A. (2015). On the generalizability of experimental results in economics. *Handbook of Experimental Economic Methodology*. (eds. G.R. Fréchet and A. Schotter), Oxford University Press, Oxford, UK.
- Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2008a). Eliciting risk and time preferences. *Econometrica*. 76, 583-618.
- Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2008b). Lost in state space: are preferences stable? *International Economic Review*. 49, 1091-1112.
- Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2010). Preference heterogeneity in experiments: comparing the field and the laboratory. *Journal of Economic Behavior & Organization*. 73, 209-224.
- Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2014). Discounting behavior: a reconsideration. *European Economic Review*. 71, 15-33.
- Andersen, S., Cox, J.C., Harrison, G.W., Lau, M., Rutström, E.E., & Sadiraj, V. (2015). Asset Integration and Attitudes to Risk: Theory and Evidence. Working paper, GSU.
- Anderson, L.R., & Mellor, J.M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*. 27, 1260-1274.
- Anderson, L.R., & Mellor, J.M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*. 39(2), 137-160.
- Andreoni, J. (1988): Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*. 37, 291–304.

- Ansher, C.A., Ariely, D., Nagler, A., Rudd, M., Schwartz, J.A., & Shah, A. (2014). Better medicine by default. *Medical Decision Making*, 34(2), 147-158.
- Attema, A.E. (2012). Developments in time preference and their implications for medical decision making. *Journal of the Operational Research Society*. 63, 1388–1399.
- Attema, A.E., Bleichrodt, H., & Wakker, P.P. (2012). A direct method for measuring discounting and QALYs more easily and reliably. *Medical Decision Making*, 32(4), 583-593.
- Attema, A.E., & Brouwer, W.B.F. (2010). The value of correcting values: influence and importance of correcting TTO score for time preference. *Value in Health*, 13(8), 879-884.
- Attema, A.E., & Brouwer, W.B.F. (2012). A test of independence of discounting from quality of life. *Journal of Health Economics*, 31, 22-34.
- Attema, A.E., & Brouwer, W.B.F. (2014). Deriving time discounting correction factors for TTO tariffs. *Health Economics*, 23(4), 410-425.
- Attema, A.E., Versteegh, M.M. (2013). Would you rather be ill now, or later? *Health Economics*, 22(12), 1496-1506.
- Baicker, C., Taubman, S.L., Allen, H.L., Bernstein, M., Gruber, J.H., Newhouse, J.P., Schneider, E.C., Wright, B.J., Zaslavzky, A.M., & Finkelstein, A.N.. (2013). The Oregon Experiment – Effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, 368, 1713-1722.
- Banerjee, A.V., Duflo, E., & Glennerster, R. (2007). Putting a band aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6(2-3), 487-500.
- Banerjee, A.V., Duflo, E., Glennerster, R., & Kothari, D. (2010). Improving immunization coverage in rural India: a clustered randomized controlled evaluation of immunization campaigns with and without incentives. *British Medical Journal*, 340, c2220.

- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2009). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.
- Battalio, R., Kagel, J., & Jiranyakul, K. (1990). Testing between Alternative Models of Choice under Uncertainty: Some Initial Results, *Journal of Risk and Uncertainty*. 3, 25-50.
- Bhargava, S. & Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review, papers and proceedings*, 105(5), 396-401.
- Bickel, Moody, & Higgins (2016). Some current dimensions of the behavioural economics of health-related behaviour change. *Preventive Medicine*, 92, 16-23.
- Binder, S. & Nuscheler, R. (2017). Risk-Taking in Vaccination, Surgery, and Gambling Environments - Evidence from a Framed Laboratory Experiment, *Health Economics*, forthcoming.
- Bleichrodt, H., (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics*. 11, 447-456.
- Bleichrodt, H., & Johannesson, M. (2001). Time preference for health: A test of stationarity versus decreasing timing aversion. *Journal of Mathematical Psychology*. 45, 265–282.
- Bleichrodt, H., Wakker, P., & Johannesson, M. (1997). Characterizing QALYs by risk neutrality. *Journal of Risk and Uncertainty*. 15, 107–114.
- Bohm, P., (1984). Revealing demand for an actual public good. *Journal of Public Economics*. 24(2), 135-151.
- Böhm, R., Betsch, C., & Korn, L. (2016). Selfish-Rational Non-Vaccination: Experimental Evidence from an Interactive Vaccination Game, *Journal of Economic Behavior & Organization*. 131 B, 183-195.
- Böhm, R., Meier, N., Korn, L. & Betsch, C., (2017). Behavioural Consequences of Vaccination Recommendations: An Experimental Analysis, *Health Economics*, forthcoming.

- Bradford, W.D. (2010). The association between individual time preferences and health maintenance habits. *Medical Decision Making*. 30, 99–112.
- Bradford, W.D., Zoller, J., & Silvestri, G.A. (2010). Estimating the Effect of Individual Time Preferences on the Use of Disease Screening. *Southern Economic Journal*. 76, 1005–1031.
- Brewer, N.T., Chapman, G.B., Schwartz, J., & Bergus, G.R. (2007). Assimilation and contrast effects in physician and patient treatment choices. *Medical Decision Making*, 27, 203-211.
- Brock, J.M., Lange, A., & Leonard, K.L. (2016). Generosity and Prosocial Behavior in Healthcare Provision Evidence from the Laboratory and Field. *Journal of Human Resources*. 51, 133-162.
- Brookshire, D.S., Coursey, D.L. & Schulze, W.D. (1987). The external validity of experimental economics techniques: analysis of demand behavior. *Economic Inquiry*. 25, 239-250.
- Bronchetti, E.T., Huffman, D.B., & Magenheimer, E. (2015). Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination. *Journal of Economic Behavior & Organization*. 116, 270-291.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016a): Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision, *Journal of Economic Behavior & Organization*, 131 B, 17-23.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. & Wiesen, D. (2016b): Physician performance pay: Evidence from a laboratory experiment, *Ruhr Economic Papers*, No. 658.

- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017a): The effects of introducing mixed payment systems for physicians: Experimental evidence, *Health Economics*, 26, 26: 243–262.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. & Wiesen, D. (2017b): Physician performance pay: Experimental Evidence. *Unpublished manuscript*, University of Cologne.
- Brosig-Koch, J., Hehenkam, B. & Kokot, J. (2017). The effects of competition on medical service provision, *Health Economics*, forthcoming.
- Buckley, N.J., Cuff, K., Hurley, J., McLeod, L., Mestelman, S., & Cameron, D. (2012): An Experimental Investigation of Mixed Systems of Public and Private Health Care Finance, *Journal of Economic Behavior & Organization*, 84, 713-729.
- Buckley, N.J., Cuff, K., Hurley, J., McLeod, L., Nuscheler, R., & Cameron, D. (2012). Willingness-to-pay for parallel private health insurance: Evidence from a laboratory experiment. *Canadian Journal of Economics*. 45, 137-166.
- Buckley, N.J., Cuff, K., Hurley, J., Mestelman, S., Thomas, S. & Cameron, D. (2015). Support for Public Provision of a Private Good with Top-Up and Opt-Out: A Controlled Laboratory Experiment. *Journal of Economic Behavior & Organization*. 111, 177-196.
- Buckley, N., Cuff, K., Hurley, J., Mestelman, S., Thomas, S., & Cameron, D., (2016). Should I stay or should I go? Exit options within mixed systems of public and private health care finance. *Journal of Economic Behavior & Organization*. 131 B, 62-77.
- Burman, L.E., Reed, W.R., & Alm, J. (2010). A call for replication studies. *Public Finance Review*. 38, 787-793.
- Burtless, G. (1995). The case for randomised field trials in economic and policy research. *Journal of Economic Perspectives*. 9, 63-84.
- Cairns, J.A. (1994). Valuing future benefits. *Health Economics*. 3, 221–229.

- Cairns, J.A., & Van Der Pol, M. (1997). Constant and decreasing timing aversion for saving lives. *Social Science and Medicine*, 45(11), 1653-1659.
- Camerer, C., & Hogarth, R. (1999). The effects of financial incentives in experiments: a review and capital-production-labor framework. *Journal of Risk and Uncertainty* 18, 7–42.
- Camerer, C.F., Dreber, A., Forsell, E., Ho, T.H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351, 1433-1436.
- Cassar, A., & Friedman, D. (2004). *Economics lab: an intensive course in experimental economics*. Routledge.
- Chamberlin, E.H. (1948). An experimental imperfect market. *Journal of Political Economy*, 56, 95–108.
- Chapman, G.B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 22, 771.
- Chapman, G.B., & Coups, E.J. (1999). Time Preferences and Preventive Health Behavior Acceptance of the Influenza Vaccine. *Medical Decision Making*. 19, 307–314.
- Chapman, G.B., & Elstein, A.S. (1995). Valuing the future: temporal discounting of health and money. *Medical Decision Making*, 15, 373-386.
- Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*. 77, 909-931.
- Charness, G., & Kuhn, P. (2011). Lab Labor: What Can Labor Economists Learn from the Lab? in Ashenfelter, O. & Card, D. (Eds.) *Handbook of Labor Economics*, Elsevier, New York, vol. 4, 229-330.
- Charness, G., & E. Fehr (2015): From the lab to the real world, *Science*, 350, 512–513.
- Chiou, Yang, & Wan (2011). Ironic effects of dietary supplementation. *Psychological Science*, 22, 1081-1086.

- Clark, A., & Loheac, Y. (2007). 'It wasn't me, it was them!': Social influence in risky behavior by adolescents. *Journal of Health Economics*, 26, 763-784.
- Cohen, J., Dupas, P., & Schaner, S. (2015). Price subsidies, diagnostic tests, and targeting of malaria treatment. *American Economic Review*, 105(2), 609-645.
- Coller, M., & Williams, M.B. (1999). Eliciting individual discount rates. *Experimental Economics*. 2, 107-127.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally.
- Cox, J.C., Sadiraj, V., Schnier, K.E., & Sweeney, J.F. (2016). Incentivizing cost-effective reductions in hospital readmission rates. *Journal of Economic Behavior & Organization*, 131 B, 24-35.
- Cox, J.C., Green, E.P., & Hennig-Schmidt, H. (2016): Experimental and behavioral economics of healthcare, *Journal of Economic Behavior & Organization*, 131(B), A1-A4.
- Cummings, R.G., Harrison, G.W. & Rutström, E.E. (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible? *American Economic Review*. 85, 260-66.
- Cummings, R.G., Elliott, S., Harrison, G.W. & Murphy, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy*. 105, 609-621.
- de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health economics*. 21, 145-172.
- Dolan, P., & Galizzi, M.M. (2014a). Getting policy-makers to listen to field experiments. *Oxford Review of Economic Policy*. 30(4), 725-752.
- Dolan, P., & Galizzi, M.M. (2014b). Because I'm worth it. A lab-field experiment on spillover effects of incentives in health. *LSE CEP Discussion Paper CEPDP1286*, London School of Economics.

- Dolan, P., & Galizzi, M.M. (2015): Like ripples on a pond: behavioral spillovers and their consequences for research and policy. *Journal of Economic Psychology*. 47, 1-16.
- Dolan, P., Galizzi, M.M., & Navarro-Martinez, D. (2015). Paying people to eat or not to eat? Carryover effects of monetary incentives on eating behavior. *Social Science and Medicine*. 133, 153-158.
- Dolan, P., & Gudex, C. (1995). Time preference, duration and health state valuations. *Health Economics*. 4, 289–299.
- Duflo, E., Dupas, P., Kremer, M. (2015). Education, HIV and early fertility: experimental evidence from Kenya. *American Economic Review*, 105(9), 2257-2297.
- Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics*, 3(1), 1-36.
- Dupas, P. (2014a). Short-run subsidies and long-run adoption of new health products: evidence from a field experiment. *Econometrica*, 82(1), 187-228.
- Dupas, P. (2014b). Getting essential health products to their end users: subsidize, but how much? *Science*, 345(6202), 1279-1281.
- Dupas, P., Hoffmann, V., Kremer, M., & Zwane, A.P. (2016). Targeting health subsidies through a non-price mechanism: a randomized controlled trial in Kenya. *Science*, 353(6302), 889-895.
- Ellermann, K., Halstenberg, K., Kuntz, L., Martakis, K., Roth, B., and Wiesen, D. (2017). The effect of feedback on antibiotics provision in pediatrics: Experimental evidence, Working paper, University of Cologne.
- Engel, C., & Rand, D.G. (2014). What does “clean” really mean? The implicit framing of decontextualized experiments. *Economics Letters*, 122(3), 386-389.
- Falk, A., & Heckman, J.J. (2009): Lab experiments are a major source of knowledge in the social sciences. *Science*. 326, 535-538.

- Fan, C., Chen, K., & Kann, K. (1998): The Design of Payment Systems for Physicians under Global Budget - An Experimental Study, *Journal of Economic Behavior & Organization*, 34, 295–311.
- Fehr, E., Kirchsteiger, G. & Riedl, A., (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*. 108, 437-459.
- Ferber, R., & Hirsch, W.Z. (1978). Social experimentation and economic policy: A survey. *Journal of Economic Literature*. 16, 1379-1414.
- Finkelstein A., & Taubman S. (2015). Randomize evaluations to improve health care delivery. *Science*, 347, 720–722.
- Finkelstein, A.N., Taubman, S.L., Wright, B.J., Bernstein, M., Gruber, J.H., Newhouse, J.P., Allen, H.L., & Baicker, K. (2012). The Oregon Health Insurance Experiment: evidence from the first year. *Quarterly Journal of Economics*. 127, 1057-1106.
- Finkelstein, A.N., Taubman, S.L., Allen, H.L., Wright, B.J., & Baicker, K. (2016). Effect of Medicaid coverage on ED use: further evidence from Oregon’s Experiment. *New England Journal of Medicine*, 375, 1505-1507.
- Fiore, S.M., Harrison, G.W., Hughes, C.E., & Rutstrom, E.E. (2009). Virtual experiments and environmental policy. *Journal of Environmental Economics and Management*. 57, 65-86.
- Frank, R.G. (2007). *Behavioral economics and health economics*. In: Diamond, P., Vartianinen, H. (Eds.), *Behavioral Economics and Its Applications*. Princeton University Press.
- Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge University Press.
- Fuchs, V. (2000). The future of health economics. *Journal of Health Economics* 19, 141 - 157.
- Gafni, A., & Torrance, G.W. (1984). Risk attitude and time preference in health. *Management Science*. 30, 440–451.

- Galizzi, M.M. (2014). What is really behavioural in behavioural health policies? And, does it work? *Applied Economics Perspectives and Policy*. 36(1), 25-60.
- Galizzi, M.M. (2017). Behavioral Aspects of Policy Formulation: Experiments, Behavioral Insights, Nudges. In Howlett M, Mukherjee I (Eds.) *Handbook of Policy Formulation*, Handbooks of Research on Public Policy, Edward Elgar Publishing, Cheltenham, UK, ISBN: 9781784719319.
- Galizzi, M.M. & Navarro-Martinez, D. (2017). On the external validity of social-preference games: A systematic lab-field study, *Management Science*, forthcoming.
- Galizzi, M.M., Harrison, G.W., & Miniaci, R. (2017). Linking experimental and survey data for a UK representative sample: structural estimation of risk and time preferences. *Unpublished manuscript*, London School of Economics.
- Galizzi, M.M., Harrison, G.W., & Miraldo, M. (2017). Experimental methods and behavioral insights in health economics: estimating risk and time preferences in health. In Baltagi, B., Moscone, F. (Eds) *Health Econometrics in Contributions to Economic Analysis*, Emerald Publishing, UK.
- Galizzi, M.M., Machado, S.R., & Miniaci, R. (2016). Temporal stability, cross-validity, and external validity of risk preferences measures: experimental evidence from a UK representative survey. *LSE Research Online Working Paper 67554*. Department of Social Policy, London School of Economics. <http://eprints.lse.ac.uk/67554/>.
- Galizzi, M.M., & Miraldo, M. (2017). Are you what you eat? Healthy behaviour and risk preferences. *The B.E. Journal of Economic Analysis and Policy*, forthcoming.
- Galizzi, M.M., Miraldo, M., & Stavropoulou, C. (2016). In sickness but not in wealth: field evidence on patients' risk preferences in the financial and health domain. *Medical Decision Making*. 36, 503–517.

- Galizzi, M.M., Miraldo, M., Stavropoulou, C., & Van Der Pol, M. (2016). Doctors-patients differences in risk and time preferences: a field experiment. *Journal of Health Economics*, 50, 171-182.
- Galizzi, M.M., Godager, G., Linnosmaa, I., Tammi, T., & Wiesen, D. (2015). Provider altruism in health economics. THL Discussion paper, No. 4.
- Galizzi, M.M. & Wiesen, D. (2017). Behavioral experiments in health: an introduction. *Health Economics*, forthcoming.
- Garcia Prado, A. Arrieta, A., Gonzalez, P. & Pinto Prades, J.L. (2017). Risk Attitudes in Medical Decisions for Others: An Experimental Approach. *Health Economics*, forthcoming.
- Gine, Karlan, D., Zinman, (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2, 213-235.
- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*. 25, 191-209.
- Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32, 1105-1116.
- Godager, G., Hennig-Schmidt, H., & Iversen, T. (2016). Does performance disclosure influence physicians' medical decisions? An experimental study. *Journal of Economic Behavior & Organization*, 131, 36-46.
- Green, E.P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization*. 106, 367-378.
- Green, D. P., & Gerber, A. S. (2003). The underprovision of experiments in political science. *The Annals of the American Academy of Political and Social Science*, 589, 94-112.
- Greiner, B., Zhang, L., & Tang, C. (2017). Separation of Prescription and Treatment in Health Care Markets-A Laboratory Experiment. *Health Economics*, forthcoming.

- Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press, Cambridge, UK.
- Halpern, S.D., Loewenstein, G., Volpp, K.G., Cooney, E., Vranas, K. Quill, C.M., McKenzie, M.S., Harhay, M.O., Gabler, N.B., Silva, T., Arnold, R., Angus, D.C., & Bryce, C. (2013). Default Options In Advance Directives Influence How Patients Set Goals For End-Of-Life Care. *Health Affairs*, 32(2).
- Halpern, S.D., French, B., Small, D.S., Sausgiver, K., Harhay, M.O., Audran-McGovern, J., Loewenstein, G., Brennan, T.A., Asch, D.A., & Volpp, K.G. (2015). A Randomized Trial of Four Financial Incentive Programs for Smoking Cessation. *New England Journal of Medicine*, 372(22), 2108-2117.
- Halpern, S.D., French, B., Small, D.S., Sauysgiver, K., Harhay, M.O., Audrain-McGovern, J., Loewenstein, G., Asch, D.A., & Volpp, K.G. (2016). Heterogeneity in the effects of reward- and deposit-based financial incentives on smoking cessation. *American Journal of Respiratory and Critical Care Medicine*, 194(8), 981-988.
- Hanoch, Y., Barnes, A.J., & Rice, T. (2017). *Behavioral Economics and Healthy Behaviors: Key Concepts and Current Research*. London, UK: Routledge.
- Hansen, F., Anell, A., Gerdtham, U. G., & Lyttkens, C. H. (2015). The future of health economics: the potential of behavioral and experimental economics. *Nordic Journal of Health Economics*. 3, 68-86.
- Harrison, G.W. (2006). Hypothetical bias over uncertain outcomes. In J.A. List (Ed.) *Using Experimental Methods in Environmental and Resource Economics*, Northampton, MA: Edward Elgar.
- Harrison, G.W. (2010). The behavioral counter-revolution. *Journal of Economic Behavior and Organization*, 73, 49-57.

- Harrison, G.W. (2013). Field experiments and methodological intolerance. *Journal of Economic Methodology*. 20(2), 103-117.
- Harrison, G.W. (2014). Cautionary notes on the use of field experiments to address policy issues. *Oxford Review of Economic Policy*. 30(4), 753-763.
- Harrison, G.W., Hofmeyr, A., Ross, D., & Swarthout, J.T. (2015). Risk preferences, time preferences, and smoking behavior. *CEAR Working Paper 2015-11*. Robinson College of Business. Georgia State University.
- Harrison, G.W., Lau, M.I., & Williams, M.B. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*. 92, 1606-1617.
- Harrison, G.W., Lau, M.I. & Rutström, E.E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*. 109, 341-368.
- Harrison, G.W., Lau, M.I., & Rutström, E.E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization*. 70, 498-507.
- Harrison, G.W., Lau, M.I., & Rutström, E.E. (2010). Individual discount rates and smoking: evidence from a field experiment in Denmark. *Journal of Health Economics*. 29, 708-717.
- Harrison, G.W., Lau, M.I., & Rutström, E.E. (2015). Theory, experimental design, and econometrics are complementary (and so are lab and field experiments). In G. Frechette & A. Schotter (Eds.), *Handbook of experimental economic methodology*. Oxford: Oxford University Press.
- Harrison, G.W., & List, J.A. (2004). Field experiments. *Journal of Economic Literature*. 42, 1009-1055.
- Harrison, G.W., & List, J.A. (2008). Naturally occurring markets and exogenous laboratory experiments: A case study of the winner's curse. *The Economic Journal*, 118, 822-843.

- Harrison, G.W., List, J.A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*. 75, 433-458.
- Harrison, G.W., & Rutström, E.E. (2008). Risk aversion in the laboratory. In J.C. Cox & G.W. Harrison (Eds.), *Risk Aversion in Experiments*. Bingley, UK: Emerald Research in Experimental Economics, Volume 12.
- Haynes, L., Service, O, Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomized controlled trials. Cabinet Office Behavioural Insights Team.
- Heckman, J.J. (2010). Building bridges between structural and program evaluation approaches to evaluating policies. *Journal of Economic Literature*. 48(2), 356-398.
- Heckman, J.J., & Smith, J.A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*. 9(2), 85-110.
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behavior: an experimental investigation. *Journal of Health Economics*. 30(4), 637-646.
- Hennig-Schmidt, H., & Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students, *Social Science & Medicine*, 108, 156-165.
- Hennig-Schmidt, H., Jürges, H., & Wiesen, D. (2017). Dishonesty in healthcare practice: A behavioral experiment on upcoding in neonatology, Working paper, University of Cologne.
- Herr, A., & Normann, H.-T. (2016). Organ donation in the lab: Preferences and votes on the priority rule, *Journal of Economic Behavior & Organization*. 161, Part B, 139-149.

- Herrnstein, R., Loewenstein, G., Prelec, D., & Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, 6, 149-185.
- Hey, J.D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*. 62, 1291-1326.
- Holt, C. & Laury, S.K. (2002) Risk Aversion and Incentive Effects. *American Economic Review*. 92,1644-1655.
- Huck, S., Lünser, G., Spitzer, F., & Tyran, J.R. (2016). Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. *Journal of Economic Behavior & Organization*, 131, 78-105.
- Jakobsson, N., Kotsadam, A., Syse, A., & Øien, H. (2016): Gender Bias in Public Long-term Care? A Survey Experiment among Care Managers, *Journal of Economic Behavior & Organization*. 131, Part B, 126-138.
- John, L., Loewenstein, G., Troxel, A., Norton, L., Fassbender, J., & Volpp, K. (2011). Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial. *Journal of General Internal Medicine*, 621-626.
- John, L., Loewenstein, G., & Volpp, K. (2012). Empirical observations on longer-term use of incentives for weight loss. *Preventive Medicine*, 55(1) S68-S74.
- Jürges, H., & Köberlein, J. (2015). What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *Journal of Health Economics*. 43, 13-26.
- Kagel, J.H., Battalio, R.C., Rachlin, H., & Green, L. (1981). Demand Curves for Animal Consumers, *Quarterly Journal of Economics*. 96 1-15.
- Kagel, J.H. (2015). Laboratory experiments: The lab in relationship to field experiments, field data, and economic theory. In G. Frechette and A. Schotter (Eds.), *Handbook of Experimental Economic Methodology*. Oxford University Press: Oxford.

- Kairies-Schwarz, N., Kokot, J., Vomhof, M., & Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory – An experiment, *Journal of Economic Behavior & Organization*. 137, 374-397.
- Keane, M.P. (2010a). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*. 156, 3-20.
- Keane, M.P. (2010b). A structural perspective on the experimentalist school. *Journal of Economic Perspectives*. 24(2), 47-58.
- Kessler, J.B., & Roth, A. (2012). Organ Allocation Policy and the Decision to Donate. *American Economic Review*, 102, 2018-2047.
- Kessler, J.B., & Roth, A. (2014a). Getting More Organs for Transplantation. *American Economic Review*. 104, 425-430.
- Kessler, J.B., & Roth, A. (2014b). Loopholes undermine donation: An experiment motivated by an organ donation priority loophole in Israel, *Journal of Public Economics*. 114, 19-28.
- Kesternich, I., Heiss, F., McFadden, D., & Winter, J. (2013). Suit the action to the word, the word to the action: hypothetical choices and real decisions in Medicare Part D. *Journal of Health Economics*. 32, 1313-1324.
- Kesternich, I., Schumacher, H., & Winter, J. (2015). Professional norms and physician behavior: homo oeconomicus or homo hippocraticus? *Journal of Public Economics*. 13, 1–11.
- Kimmel, S.E., Troxel, A.B., Loewenstein, G., Bensinger, C.M., Jaskowiak, J., Doshi, J.A., Laskin, M., & Volpp, K. (2012). Randomized Trial of Lottery-Based Incentives to Improve Warfarin Adherence. *American Heart Journal*, 164(2), 268-274.

- Kimmel, S.E, Troxel, A.B., French, B., Loewenstein, G., Brensing, C.M., Meussner, C., Doshi, J.A., Laskin, M., & Volpp, K. (2016). A randomized trial of lottery-based incentives and reminders to improve warfarin adherence: The Warfarin Incentives (WIN2) Trial. *Pharmacoepidemiology and Drug Safety*, 25, 1219-1227.
- Kokot, J., Brosig-Koch, J., & Kairies-Schwarz, N. (2017). Sorting into payment schemes and medical treatment: A laboratory experiment, *Health Economics*, forthcoming.
- Kolstad, J. R., & Lindkvist, I. (2012). Pro-Social Preferences and Self-Selection into the Public Health Sector: Evidence from an Economic Experiment. *Health Policy and Planning*. 28, 320–327.
- Köszegi, B. (2003). Health anxiety and patient behavior. *Journal of Health Economics*. 22, 1073-1084.
- Köszegi, B. (2006). Emotional agency. *Quarterly Journal of Economics*. 121,121-155.
- Kramer, M., & Shapiro, S. (1984). Scientific challenges in the application of randomized trials. *Journal of the American Medical Association*. 252, 2739-2745.
- Krawczyk, M. (2011). What brings your subjects to the lab? A field experiment. *Experimental Economics*, 14, 482-489.
- Kullgren, J.T., Troxel, A.B., Loewenstein, G., Asch, D.A., Norton, L.A., Wesby, L., Tao, Y., Zhu, J., & Volpp, K.G. (2013). Individual vs. Group-Based Incentives for Weight Loss: A Randomized, Controlled Trial. *Annals of Internal Medicine*, 158(7), 505-514.
- Kullgren, J.T., Troxel, A.B., Loewenstein, G., Norton, L.A., Gatto, D., Tao, Y., Zhu, J., Schofield, H., Shea, J.A., Asch, D.A., Pellathy, T., Driggers, J., & Volp, K.G. (2016). A randomized controlled trial of employer matching of employees' monetary contributions to deposit contracts to promote weight loss. *American Journal of Health Promotion*, 30(6), 441-452.

- Lagarde, M. & Blaauw, D. (2017). Physicians' responses to financial and social incentives: A medically framed real effort experiment, *Social Science & Medicine*, 179, 147-159.
- Leamer, E.E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives*. 24(2), 31-46.
- Leeds, I.L., Sadiraj, V., Cox, J.C., Gao, X.S., Pawlik, T.M., Schnier, K.E. & Sweeney, J.F. (2017). Discharge decision-making after complex surgery: Surgeon behaviors compared to predictive modeling to reduce surgical readmissions. *American Journal of Surgery*, 213(1), 112-119.
- Levitt, S.D., & List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*. 21(2), 153-174.
- Levitt, S.D., & List, J.A. (2009), Field experiments in economics: The past, the present, and the future, *European Economic Review*, 53(1), 1-18.
- Levitt, S.D., & List, J.A. (2011). Was there really a Hawthorne effect at the Hawthorne Plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics*. 3, 224-238.
- Li, D., Hawley, Z., & Schnier, K. (2013). Increasing organ donation via changes in the default choice or allocation rule, *Journal of Health Economics*. 32, 1117-1129.
- Lindeboom, M., van der Klaauw, B., & Vriend, S. (2016): Audit rates and compliance: A field experiment in care provision, *Journal of Economic Behavior & Organization*. 131B, 160–173.
- Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65, 272-92.
- Loewenstein, G., Brennan, Volpp, K. (2007). Asymmetric paternalism to improve health behaviors. *JAMA*, 298, 2415-2417.

- Loewenstein, G., Asch, D., Volpp, K. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs*, 32(7), 1244-1250.
- Loewenstein, G., Price, J., & Volpp, K.G. (2016). Habit Formation in children: Evidence from incentives for healthy eating. *Journal of Health Economics*, 45, 47-54.
- Loewenstein, G., Schwartz, J., Ericson, K., Kessler, J.B., Bhargava, S., Hagmann, D., Blumenthal-Barby, J., D'Aunno, T., Handel, B., Kolstad, J., Nussbaum, D., Shaffer, V., Skinner, J., Ubel, P., & Zikmund-Fisher, B.J. (2017). Behavioral Insights for Health Care Policy. *Behavioral Science and Policy* (in press).
- Long, J.A., Jahnle, E.C., Richardson, D.M., Loewenstein, G., & Volpp, K.G. (2012). A Randomized Controlled Trial of Peer Mentoring and Financial Incentive to Improve Glucose Control in African American Veterans. *Annals of Internal Medicine*. 156, 416-424.
- Manning, W.D., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A., & Marquis, S.M. (1981). Some interim results from a controlled trial on cost sharing in health insurance. *New England Journal of Medicine*, 305(25), 1501-1507.
- Manning, W.D., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A., & Marquis, S.M. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*. 77, 251-277.
- Marwell, G., & Ames, R.E., (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods. *Journal of Public Economics*, 15, 295-310.
- Massin, S., Ventelou, B., Nebout, A., Verger, P., & Pulcini, C. (2015). Cross-sectional survey: risk-averse French general practitioners are more favorable toward influenza vaccination. *Vaccine*, 33(5), 610-614.

- Michel-Lepage, A., Ventelou, B., Nebout, A., Verger, P., & Pulcini, C. (2013). Cross-sectional survey: risk averse French GPs use more rapid-antigen diagnostic tests in tonsillitis in children. *BMJ Open*, 3(10), e003540.
- Michie, S. (2008). Designing and implementing behavior change interventions to improve population health. *Journal of Health Services Research & Policy*, 13, 64-69.
- Miguel, E., Camerer, C.F., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B.A., Petersen, M., Sedlmayr, R., Simmons, J.P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30-31.
- Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108, 10415–10420.
- Milkman, K., Minson, J., & Volpp, K. (2014). Holding the hunger games hostage at the gym: an evaluation of temptation bundling. *Management Science*, 60(2), 283-299.
- Mimra, W., Rasch, A., & Waibel, C., (2016). Second opinions in markets for expert services: Experimental evidence. *Journal of Economic Behavior & Organization*. 131,106-125.
- Muller et al (2009). Tell me why: the influence of self-involvement on short term smoking behaviour. *Addictive Behaviors*, 34, 427-431.
- Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*. 119, 39-77.
- Newhouse, J.P., Manning, W.G., Morris, C.N., Orr, L.L., Duan, N., Keeler, E.B., Leibowitz, A., Marquis, K.H., Marquis, M.S., Phelps, C.E., & Brook, R.H., (1981). Some interim results from a controlled trial of cost sharing in health insurance. *New England Journal of Medicine*, 305, 1501–1507.

- Ockenfels, A., & Wiesen, D. (2017). Professional culture and dishonesty in medicine, *Unpublished manuscript*, University of Cologne.
- Oliver, A.J. (2017). *The Origins of Behavioural Public Policy*. Cambridge, UK: Cambridge University Press.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Oxholm, A.S. (2016). Physician response to target-based performance payment. University of Southern Denmark. COHERE Discussion Papers, No. 9, 2016.
- Pauly, M.V., (1979). The ethics and economics of kickbacks and fee Splitting. *Bell Journal of Economics*, 10, 344-352.
- Plott, C.R. (1982). Industrial Organization Theory and Experimental Economics. *Journal of Economic Literature*. 20, 1485–1572.
- Prosser, L.A., & Wittenberg, E. (2007). Do risk attitudes differ across domains and respondent types? *Medical Decision Making*. 27, 281–287.
- Rana, Y., Haberer, J., Huang, C., Kambugu, A., Mukasa, B., Trirumurthy, H., Wabukala, P., and Linnemayr, S., (2015). Short message service (SMS)-based intervention to improve treatment adherence among HIV-positive youth in Uganda: focus group findings. *PLOS ONE*, 10(4), e0125187.
- Rao, G., Krall, J., & Loewenstein, G. (2011). An Internet-based Pediatric Weight Management Program with and without Financial Incentives: A Randomized Trial. *Childhood Obesity*, 7(2), 122-128.
- Rassenti, S.J., Smith, V.L., & Bulfin R.L. (1982). A combinatorial auction mechanism for airport time slot allocation, *Bell Journal of Economics*, 13(2), 402–417.
- Roberto, C.A., & Kawachi, I. (2015). *Behavioral Economics and Public Health*. Oxford, UK: Oxford University Press.

- Roth, A.E. (2002): The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica*, 70: 1341–1378.
- Roth, A.E., & Peranson, E. (1999): The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*. 89, 748-780.
- Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*. 7, 51-84.
- Rubin, D.B. (1974). Estimating the causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*. 66, 688-701.
- Ryan, M., McIntosh, E., & Shackley, P. (1998). Methodological issues in the application of conjoint analysis in health care. *Health Economics*. 7, 373-378.
- Ryan, M., & Farrar, S. (2000). Using conjoint analysis to elicit preferences for health care. *BMJ: British Medical Journal*. 320, 1530.
- Samper, L.A. & Schwartz, J.A. (2013). Price inferences for sacred versus regular goods: changing the price of medicine influences perceived health risk. *Journal of Consumer Research*, 39, 1343-1358.
- Sauermann, H., & Selten, R. (1959). Ein Oligopolexperiment. *Zeitschrift für die gesamte Staatswissenschaft*, 115, 427–471.
- Schofield, H., Loewenstein, G., Kopisc, J., & Volpp, K.G. (2015). Comparing the effectiveness of individualistic, altruistic, and competitive incentives in motivating completion of mental exercises. *Journal of Health Economics*. 44, 286-299.
- Schram, A., & Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review*. 55, 799–819.

- Schwartz, J.A., Riis, J., Elbel, B., & Ariely, D. (2012). Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. *Health Affairs*, 31, 2399-2407.
- Schwartz, J.A., Mochon, D., Wyper, L., Maroba, J., Patel, D., Ariely, D. (2014). Healthier by precommitment. *Psychological Science*, 25(2), 538-546.
- Schwartz, J.A., & Chapman, G.B. (1999). Are more options always better? The attraction effect in physicians' decisions about medications. *Medical Decision Making*, 19, 315-323.
- Sen, A.P, Sewell, T.B., Riley, E.B., Stearman, B., Bellamy, S.L., Hu, M.F., Park, J.D., Yao, Y., Zhu, J., Loewenstein, G., Asch, D.A., & Volpp, K.G. (2014). Financial Incentives for Home-Based Health Monitoring: A Randomized Controlled Trial. *Journal of General Internal Medicine*, 29(5), 770-777.
- Silverman, E., & Skinner, J. (2004). Medicare upcoding and hospital ownership, *Journal of Health Economics*. 23, 369-389.
- Smith V.L., (1976) Experimental economics: induced value theory. *American Economic Review*, 66, 274-279.^[1]_[SEP]
- Smith V.L., (1982). Microeconomic systems as experimental science. *American Economic Review*, 72, 923-955.
- Smith, V.L. (2002). Method in experiment: rhetoric and reality. *Experimental Economics*, 5, 91-110.
- Sunstein, C.R. (2011). Empirically informed regulation. *University of Chicago Law Review*. 78, 1349-1429.
- Sutter, M., Kocher, M.G., Glätzle-Rützler, D., & Trautmann, S.T. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review*, 103, 510–31.
- Szrek, H., Chao, L.-W., Ramlagan, S., & Peltzer, K. (2012). Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. *Judgment and Decision Making*. 7,

716. Tsutsui, Y., Benzion, U., & Shahrabani, S. (2012). Economic and behavioral factors in an individual's decision to get an influenza vaccination in Japan, *Journal of Socio-Economics*, 41(5), 594-602.
- Ubel, P. & Loewenstein, G. (1995). The efficacy and equity of retransplantation: an experimental survey of public attitudes. *Health Policy*, 34, 145-151.
- Thaler, R.H. & Sunstein, CR. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Princeton, NY: Princeton University Press.
- Van Der Pol, M., & Cairns, J.A. (1999). Individual Time Preferences for Own: an Application of a Dichotomous Choice Question with Follow-Up. *Applied Economics Letters*, 6, 649-654.
- Van der Pol, M., & Cairns, J.A. (2001). Estimating time preferences for health using discrete choice experiments. *Social Science and Medicine*. 52, 1459–1470.
- Van Der Pol, M., Cairns, J., (2002). A Comparison of the Discounted Utility Model and Hyperbolic Discounting Models in the Case of Social and Private Intertemporal Preferences for Health. *Journal of Economic Behavior and Organization*, 49, 79-96.
- Van der Pol, M., & Cairns, J.A. (2008). Comparison of two methods of eliciting time preference for future health states. *Social Science and Medicine*. 67, 883–889.
- Van Der Pol, M., & Cairns, J. (2011). Descriptive validity of alternative intertemporal models for health outcomes: an axiomatic test. *Health Economics*, 20, 770-782.
- Van Der Pol, M., & Ruggeri, M., 2008. Is risk attitude outcome specific within the health domain? *Journal of Health Economics*, 27, 706-717.
- VanEpps, E.M., Downs, J.S., & Loewenstein, G. (2016a). Calorie label formats: Using numbers or traffic lights to reduce lunch calories. *Journal of Public Policy and Marketing*, 35(1), 26-36.

- VanEpps, E.M., Downs, J.S., & Loewenstein, G. (2016b). Advance ordering for healthier eating? Field experiments on the relationship between time delay and meal healthfulness. *Journal of Marketing Research*, Vol LIII, 369-380.
- Vervloet et al. (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication. *JAMA*, 19, 696-704.
- Viscusi, W.K., & Hersch, J. (2001). Cigarette smokers as job risk takers. *Review of Economics and Statistics*. 83(2), 269–80.
- Viscusi, W., & Hakes, J.K. (2008). Risk beliefs and smoking behavior. *Economic Inquiry*. 46(1), 45–59.
- Volpp, K., John, L.K., Troxel, A.B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss, *Journal of American Medical Association*. 300, 2631-7.
- Volpp, K.G., Loewenstein, G., Troxel, A.B., Doshi, J., Price, M., Laskin, M., & Kimmel, S.E. (2008). A test of financial incentives to improve warfarin adherence. *Biomedical Central: Health Services Research*, 8:272.
- Volpp, K.G., Asch, D.A., Galvin, R., & Loewenstein, G. (2011). Redesigning employee health incentives: lessons from behavioural economics. *New England Journal of Medicine*. 365, 388-390.
- Waibel, C., & Wiesen, D. (2017) An experiment on referrals in health care, Working Paper, ETH Zurich.
- Wang, J., Iversen, T. Hennig-Schmidt, H. & Godager, G. (2017). How Changes in Payment Schemes Influence Provision Behavior. HERO Working Paper 2017/2.
- Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*. 42, 1131–50.

- Weber, E.U., Blais, A.R., & Betz, N.E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4), 263-290.
- Wisdom, J., Downs, J., & Loewenstein, G. (2009). Promoting healthy choices: information versus convenience. *American Economic Journal: Applied Economics*, 99(2), 159-164.
- Zizzo, D.J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13, 75–98.