Experimental Methods and Behavioral Insights in Health Economics: Estimating Risk and Time Preferences in Health

Matteo M Galizzi[a], Glenn W Harrison[b], Marisa Miraldo[c]

a. London School of Economics: Department of Social Policy, and LSE Behavioural Science Hub; Old 2.35, Old Building, Houghton Street, WC2A 2AE London (UK). Email: m.m.galizzi@lse.ac.uk
b. Georgia State University: Department of Risk Management & Insurance, and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University; 35 Broad Street, 11th Floor, Atlanta, GA 30303 (US). Harrison is also affiliated with University of Cape Town: School of Economics. Email: gharrison@gsu.edu.
c. Imperial Business School: Department of Management. Exhibition Road, SW7 2AZ, London. Email: m.miraldo@imperial.ac.uk.

Abstract

The use of behavioral insights and experimental methods has recently gained momentum among health policy-makers. There is a tendency, however, to reduce behavioral insights applications in health to "nudges," and to reduce experiments in health to "randomized controlled trials" (RCTs). We argue that there is much more to behavioral insights and experimental methods in health economics than just nudges and RCTs. First, there is a broad and rich array of complementary experimental methods spanning the lab to the field, and all of them could prove useful in health economics. Second, there are a host of challenges in health economics, policy, and management where the application of behavioral insights and experimental methods is timely and highly promising. We illustrate this point by describing applications of experimental methods and behavioral insights to one specific topic of fundamental relevance for health research and policy: the experimental elicitation and econometric estimation of risk and time preferences. We start by reviewing the main methods of measuring risk and time preferences in health. We then focus on the "behavioral econometrics" approach to jointly elicit and estimate risk and time preferences, and we illustrate its state-of-the-art applications to health.

Keywords: behavioral experiments in health; risk preferences; time preferences; randomized controlled trials; technical paper.

Introduction

In the last few years, insights from behavioral economics have gained momentum among health policy-makers. Governments in developed countries have constituted "behavioral insights teams," within their civil services, including the so-called "Nudge Unit" in the UK Cabinet Office, the behavioral science teams within the UK Department of Health, the NHS, and Public Health England, and analogous initiatives in other governments and international institutions (Dolan and Galizzi, 2014a; Sunstein, 2011).

At the same time, experimental methods have attracted attention among researchers and policy-makers. Most researchers and decision-makers in health economics, policy, and management, however, still implicitly associate experiments with developing countries, where the priorities and organization of health policies and systems are centered around access to care and the outspread of infectious diseases. Almost invariably, these experiments are also viewed as synonymous with "randomized controlled trials" (RCTs).

There is a tendency to reduce the remit of "behavioral" applications to health to "nudges," to subtle changes in the decision environment or "choice architecture," that trigger changes in health behavior at an automatic or unconscious level (Galizzi, 2014). We argue that there is much more to behavioral insights and experimental methods in health economics than just nudges and RCTs. We argue this for two reasons.

First, from a methodological perspective the current emphasis on RCTs is misleading because it overlooks the richness of a broad array of experimental methods spanning the lab to the field. The discussion about experimental methods in health economics should be centered around the advantages, disadvantages, and potential scope of the different types of experiments proposed by Harrison and List (2004), each of which could prove useful in health economics. The different types of experiments, including RCTs, should be seen as complementary rather than substitutes. More generally experiments should be seen as complementary to theory and econometrics.

Second, in health economics, policy, and management experimental methods and behavioral insights go beyond nudges and are applied to contexts other than developing countries. Many challenging applications are in high and middle income countries, that are redesigning their health systems to address the pressing challenges brought about an ageing populations and increased prevalence of non-communicable diseases.

Why does this focus on RCTs and nudges matter when we are concerned about developments in health econometrics? Advocates of nudges and RCTs often argue that in assessing interventions one should "let the data speak", that is design an experiment that lets the data speak directly to those wanting policy evaluations. Simple "average treatment effects", ideally, become just a matter of arithmetic. This is viewed as a positive thing as it allows policy evaluations that do not depend on econometric assumptions, beyond the basics of randomization. The same agnosticism extends to the use of formal theory. In what follows, however, we will argue that theory, behavioral experiments, and econometrics are complementary methods, rather than substitutes.

We illustrate this point by describing the applications of experimental methods and behavioral insights to one specific topic of fundamental relevance for health research and policy: the elicitation and estimation of risk and time preferences.

Section 2 discusses the links between RCTs and the broader range of types of experiment. Sections 3 and 4 illustrate the use of behavioral experiments in health, focusing on risk and time preferences. Section 3 contains reviews of the main methods to measure risk and time preferences in health, while Section 4 focuses on the "behavioral econometrics" approach to jointly elicit and estimate risk and time preferences, and illustrates applications to health.[1] Finally, Section 5 concludes.

Randomized Controlled Trials and Experiments in Health Economics

Many policy-makers tend to automatically associate *behavioral experiments in health* with RCTs, probably because one of the most influential reports by the UK's Behavioural

[1] The term "behavioral econometrics" refers to the application of econometric methods to study issues in behavioral economics. Many of the general methods are the same as in economics in general, but often the adaptation to behavioral issues stresses different aspects of general methods. Section 4 provides one rich example, the use of multiple experimental tasks and joint, full information estimation of latent preferences. Others include the use of mixture models as illustrated in Harrison and Rutström (2009).

Insights Team, when it was still within the Cabinet Office, stressed the need to conduct randomized controlled trials to develop public policy (Haynes, Service, Goldacre and Torgerson, 2012).

This emphasis on the use of RCTs as the exclusive or fundamental experimental method in health economics, policy, and management, calls for two conceptual clarifications.

First, health economists are particularly well aware that the use of RCTs in science is not a new development. Modern evidence-based medicine and pharmacology are all based on 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. Far from novel, the idea of using randomized controlled experiments has been advocated for decades even for policy applications (Burtless, 1995; Ferber and Hirsch, 1978; Manning et al., 1987; Rubin, 1974).

Second, while the term RCT is now widely *en vogue* in policy circles, it is often used in a quite peculiar way. In the health economics applications in developing countries, the term RCT is typically used to denote large-scale 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 difference with respect to 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 and policy 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.

Furthermore, in the current debate, RCTs are often improperly contrasted with theory, structural econometrics, and other empirical methods that do not entail randomization. These contrasts are artificial and misleading because they often overlook the fact that there is currently debate on whether randomized controlled evaluations allow one to draw causality statements about *latent* causes or effects, particularly welfare effects. There also remains lively debate on the virtues of "atheoretic versus structural" approaches to econometrics (Deaton, 2010; Harrison, 2010, 2013, 2014; Heckman, 2010; Keane, 2010a, 2010b; Leamer, 2010). Randomization does not *require* that one should give up theory or standard econometrics, although it does *allow* one to do so under certain idealized circumstances.

Furthermore, RCTs are only one specific type of experiment. The taxonomy of experiments proposed by Harrison and List (2004) illustrates the diversity of experiments: *conventional lab* experiments involve student subjects, abstract framing, a lab context, and a set of imposed rules; *artefactual field* experiments depart from conventional lab experiments in that they involve non-student samples; *framed field* experiments add to artefactual field experiments a field context in the commodity, stakes, task or information; and, finally, *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); *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).

Moreover, RCTs and other randomized experiments can also be potentially subject to *randomization biases*. These are sample selection biases related to the fact that knowing that random assignment is in place causes the type of individual participating in a randomized study to differ from participants in other studies in a way that affects inferences about some outcome of interest (Harrison, Lau and Rutström, 2009; Heckman and Smith, 1995; Kramer and Shapiro, 1984). This type of bias is fundamental to randomized evaluations, assuming that subjects know or even perceive that randomization to treatment is occurring.

Other types of experiments beyond lab, artefactual field, frame field, and natural field experiments, are *virtual experiments* that combine controlled experiments with virtual reality settings (Fiore, Harrison, Hughes and Rutström, 2009); and *lab-field experiments* that 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). While not yet applied to health economics contexts, virtual experiments are a promising approach to make tradeoffs more salient and vivid in health decision-making.

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, Heiss, McFadden and Winter, 2013). They are also part of the more general *behavioral data linking* approach, that is, the linkage of behavioral economics experiments with longitudinal surveys, administrative registers, biomarkers banks, scan data, and other big data sources (Andersen et al., 2016; Galizzi, Harrison and Miniaci, 2016). Data linkage poses new ethical, practical, and logistical challenges when it seeks to link surveys and behavioral experiments with healthcare registers.

RCT is therefore an inappropriate and misleading term to be used to *define* behavioral experiments in health, as it does not convey key information on the exact nature, scope, and typology of the experiment at hand. There is no consensus on whether lab or field experiments are superior: both have strengths and weaknesses, and their relative merits have been systematically discussed elsewhere (Bardsley et al. 2009; Harrison, Lau and Rutström, 2015; Harrison and List, 2004; Levitt and List, 2007). There is not one experiment type 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.

In fact, the broad spectrum of lab to natural field experiments in the Harrison and List (2004) sense have already been applied in a range of developed countries and economies in transition contexts to a variety of health economics and policy areas, well beyond "nudges." For example, lab, artefactual field, framed field, and natural field experiments in health have investigated the effects of different providers' incentives and the role of altruism in Belgium, Germany, and Norway (Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz and Wiesen, 2016; Hennig-Schmidt et al., 2011; Kesternich et al., 2013); the role of different healthcare financing policies in Canada (Buckley et al. 2012); the choice of health insurance in Germany (Kesternich et al., 2013); the impact and the unintended spillover effects of financial incentives for health behaviors in the UK and the US (Dolan and Galizzi, 2014b, 2015; Dolan et al., 2015; Volpp et al., 2008; Volpp, Asch, Galvin and Loewenstein, 2011). In the next Sections, we illustrate the scope and potential of behavioral experiments in health by considering an area of fundamental interest for health economics research and policy: the measurement of risk and time preferences.

Measuring Risk and Time Preferences in Health: a Review of Methods

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; Van der Pol, 2011). Empirically, they have been found to associate with a number of health and healthcare behaviors, including: heavy drinking (Anderson and Mellor, 2008), BMI (Sutter, Kocher, Glätzle-Reutzler and Trautmann, 2013), healthy nutrition (Galizzi and Miraldo, 2017), and the uptake of vaccinations, preventive care, and medical tests (Bradford, 2010; Bradford, Zoller and Silvestri, 2010; Chapman and Coups, 1999).

Moreover, risk and time preferences directly inform the principles and practices of cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) in healthcare (Bleichrodt, Wakker and Johannesson, 1997). In particular, they inform 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 popular Standard Gamble (SG) and Time Trade Off (TTO) methods to elicit preferences for hypothetical health states (Attema and Brouwer, 2012; Bleichrodt and Johannesson, 2001).

Despite such a fundamental role of risk and time preferences in the health context, it is surprising that there is no consensus to date on a "gold standard" measurement methodology. A multitude of different methods have, in fact, 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.

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, it is unclear whether individual risk and time preferences are stable across the health and the monetary domains

(Chapman, 1996). Galizzi, Miraldo and Stavropoulou (2016) 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 priority in the research agenda of behavioral experimentalists in health.

In sub-section 3.1 we review the main methods to measure risk taking in health, while in sub-section 3.2 we review the main methods to measure time preferences in health. In the next Section 4 we focus on the *behavioral econometrics* approach to jointly elicit and estimate risk and time preferences, and we illustrate its main applications to health.

Measuring Risk Taking in Health

There are six main approaches to measure risk taking in health. The first approach uses *insurance market choices* to infer underlying risk preferences from different insurance contracts (Barseghyan, Prince and Teitelbaum, 2011; Einav, Finkelstein, Pascu and Cullen, 2012).

A second approach uses "risky" health behavior such as smoking, heavy drinking, or not using seat belts as indirect proxies for risk taking (Viscusi and Hakes, 2008; Viscusi and Hersch, 2001). In this *behavior-proxy* strategy, risk preferences are indirectly inferred from observed behavior rather than being directly measured.

A third approach assumes that risk taking is inherently domain-specific, and should therefore be measured by *domain-specific questionnaires* such as the Domain-Specific Risk-Taking scale (DOSPERT) (Weber, Blais and Betz, 2002). A disadvantage of this approach is that risk preferences are not directly measured but are inferred from self-reported engagement in "risky" behaviors.

The fourth approach, a simplified variant of the third, is based on self-assessed willingness to take risk generally and in specific domains using Likert scales (Dohmen et al., 2011). This *scale-based self-assessed* approach is simple and scores can be quantitatively compared across domains. However, the theoretical foundations of this "*direct scaling*" approach remain unclear. In particular, the procedure does not allow associating the different individual choices to specific ranges of risk aversion parameters under a theoretical framework. Moreover, the evidence on how these scores correlate with experimental measures for risk preferences is mixed at best (Galizzi, Machado and Miniaci, 2016; Szrek, Chao, Ramlagan and Peltzer, 2012).

Two common features of the last three approaches are that they are not incentive-compatible, in the sense that the measures are merely hypothetical and they entail no real consequence to subjects, and that they involve purely self-reported scale measures rather than explicit tradeoffs.

The fifth approach encompasses a family of methods that directly measure risk preferences with tasks involving explicit trade-offs, rather than self-reported scales. Within this

trade-off approach a common method is the *certainty equivalent* (CE) method (Prosser and Wittenberg, 2007; Wakker and Deneffe, 1996). CE questions have also been included in surveys such as the US Health and Retirement Survey, with mixed evidence on their links with other risk preference measures and with risky health behaviors (Anderson and Mellor, 2009). Other methods within this trade-off approach are the *probability equivalent* (PE) method, which is also at the heart of the SG method commonly used to measure utilities of health states, and the *gamble tradeoff* (GTO) method.

The final approach to measure risk preferences in health uses *incentive-compatible* (IC) tests involving real rewards to respondents. A "hypothetical response bias" has been documented in the elicitation of risk preferences, with hypothetical methods showing significantly less risk aversion than methods with real rewards (Harrison, 2006). Since measuring risk preferences in health with real health consequences is challenging, most IC methods offer monetary rewards, rather than health rewards, and compare elicited risk preferences to health behaviors (Anderson and Mellor, 2008).

The most common IC measurement procedures for risk preferences for monetary outcomes are reviewed by Harrison and Rutström (2008) and Charness, Gneezy and Imas (2013). The multiple price list (MPL) design presents subjects with a series of ordered questions, each reproducing a choice between two lotteries. This method fully accounts for an individual being risk averse, risk neutral, or risk seeking, whereas other methods cannot empirically distinguish between risk neutrality and risk seeking behavior, and some do not even allow one to discriminate between different degrees of risk seeking. A second major advantage of the MPL method is that it allows the researcher to structurally estimate the underlying risk preferences for a sample of subjects. In particular, the *behavioral econometrics* approach developed by Harrison and Rutström (2008) and Andersen, Harrison, Lau and Rutström (2008a) uses Maximum Likelihood (ML) to estimate the risk aversion parameters assuming a range of Expected Utility Theory (EUT) and non-EUT models (see Section 4.1). The binary choice battery approach shares these advantages, but also allows the estimation of EUT and non-EUT models of risk preferences at the level of the individual (Harrison and Ng, 2016).

Measuring Time Preferences in Health

Several reviews examine different measures of time preferences (Andersen, Harrison, Lau and Rutström, 2014; Frederick, Loewenstein and O'Donoughe, 2002). Here we focus on its applications to health.

The empirical literature which estimates time preferences can be divided generally into three main groups: research that takes advantage of naturally occurring data; research that uses survey questions; and research utilizing experimental methods.

The first consists of using naturally occurring situations in which individuals must choose between alternatives with differential time dimensions, such that an underlying discount rate can be inferred.

A second approach consists of presenting subjects with a set of survey questions about hypothetical immediate and delayed payouts (e.g. Health and Retirement Survey). These survey questions typically lack real monetary payouts, although attempts have been recently made to link incentive-compatible experimental tests to household surveys (Galizzi et al., 2016; Tanaka, Camerer and Nguyen, 2010).

The third approach consists of using experimental methods to assess time preferences for health outcomes (Cairns, 1994; Dolan and Gudex, 1995; Van der Pol and Cairns, 2001, 2008). These methods typically use hypothetical tradeoffs. Galizzi, Miraldo, Stavropoulou, and Van der Pol (2016), for example, compare time preferences for hypothetical healthcare outcomes between doctors and their matched patients.

Coller and Williams (1999) appears to be the first incentive-compatible study to assess time preferences for monetary outcomes with real monetary payments. This approach, further developed by Harrison, Lau and Williams (2002) has become a popular standard method to elicit time preferences with real monetary incentives.

The most general approach is to ask subjects a MPL of binary options between a "Smaller Sooner" (SS) monetary payout and a "Larger Later" (LL) amount, allowing one to estimate structural models of discounting function at the pooled and individual levels. When combined with a task to elicit risk preferences, it allows joint "structural" estimations of risk and time preferences considering a broad range of theoretical models following the maximum-likelihood *behavioral econometrics* approach. Andersen et al., (2008a) were the first to highlight that time and risk preferences should necessarily be elicited and estimated jointly, by noticing that the level of respondents' discount rate can not be inferred without knowing or assuming something about the curvature of the utility function (implied, under EUT, by their risk attitudes). In particular, as explained below, assuming risk neutrality when the decision maker is actually risk averse implies a significant overestimation of the latent discount rates.

Jointly Eliciting and Estimating Risk and Time Preferences: the *Behavioral Econometrics* Approach

General Framework

Using ML, the *behavioral econometrics* approach allows jointly eliciting and structurally estimating risk and time preferences for a broad range of models, encompassing EUT and non-EUT models for risk preferences, and exponential and non-exponential discounting models for time preferences. Here we illustrate the approach under EUT. We start by assuming that time preferences are discounted exponentially, and we then consider various models of non-constant discounting.

Let *U(.)* denote subjects utility function that is separable and stationary over time, and *D(t)* a generic discount factor. A subject is then indifferent between SS income option M_τ at time τ and a LL income $M_{\tau+t}$ at time $\tau + t$ (with $t > 0$) if and only if the sum of the discounted

utilities of receiving the monetary outcome M_τ at time τ in addition to the background consumption ω, and receiving nothing extra at time $\tau + t$ equals the sum of the discounted utilities of receiving nothing in addition to background consumption ω at time τ, and the outcome $M_{\tau+t}$ plus background consumption at time $\tau + t$, i.e.:

$$U(M_\tau + \omega) + D(t)U(\omega) = U(\omega) + D(t)U(M_{\tau+t} + \omega) \quad (1)$$

An Identification Problem

All empirical analyses of discounting models prior to Andersen, Harrison, Lau and Rutström (2008a) implicitly assumed that the subject is risk neutral, so that Eq. (1) could be simply written as:

$$M_\tau = D(t)M_{\tau+t} \quad (2)$$

If, however, the decision maker is not risk neutral, it is evident from an application of Jensen's inequality that the implied discount rate inferred from given choices over SS and LL combinations decreases if *U(M)* is concave in *M*. Therefore assuming neutrality when the decision maker is actually risk averse implies an overestimation of the implicit discount rate, due to diminishing marginal utility. More generally, there exists an identification problem in the sense that one cannot infer the level of respondents' discount rate without knowing or assuming something about their risk attitudes (i.e. the value of *U''(.)*). Consequently time and risk preferences should be jointly estimated.

Choice Under Risk

It is typically assumed that subjects have a CRRA utility function:

$$U(M_\tau, \omega) = (\omega + M_\tau)^{(1-r)}/(1-r) \quad (3)$$

where r is the CRRA coefficient and $r \neq 1$,. Within this CRRA functional form, $r = 0$ implies risk-neutral choices, $r > 0$ denotes risk aversion, and $r < 0$ implies risk-seeking behavior.

It is often assumed that the utility function is stable over time, and is perceived *ex ante* to be stable over time, so that the same *r* applies to utility over the lottery payoffs, which are paid immediately, as well as to utility over inter-temporal monetary payoffs, which may be paid with a delay.[2]

Individual Discounting

The general definition of the discount factor $D(t)$ does not impose any *a priori* restriction on the functional form of the utility function *U(.)*. It can be linear for risk neutrality or non-linear

[2] The question of whether risk preferences are indeed stable over time is directly investigated by Andersen, Harrison, Lau and Rutström (2008b) in longitudinal artefactual field experiments with representative samples in Denmark, and by Galizzi et al., (2016), and Galizzi et al., (2016) in the UK.

that has been shown to be of great significance for the estimation of the discounting rates. Similarly, the definition of the discount factor does not restrict *U(.)* to EUT, and in fact several non-EUT specifications can be considered.

One can initially assume that the discounting factor takes the *exponential* (*E*) specification for t≥ 0:

$$D^E(t) = 1/(1+\delta)^t \quad (4)$$

In this case, the discount rate is simply $d^E(t) = \delta$, with the key assumption that the rate is constant over time: the percentage rate at which utility today and utility tomorrow is discounted is exactly the same as the rate at which utility in 1 month and utility in 1 month and 1 day is discounted.

The discounting factor for the *quasi-hyperbolic* (*QH*) specification is defined as:

$$D^{QH}(t) = 1 \quad if \;\; t = 0 \quad (5a)$$
$$D^{QH}(t) = \beta/(1+\delta)^t \quad if \;\; t > 0 \quad (5b)$$

where $\beta < 1$ implies quasi-hyperbolic discounting, and $\beta = 1$ collapses to exponential discounting. The defining aspect of the *QH* specification is that the discount function has a jump discontinuity at $t = 0$, and for any $t > 0$ is exactly the same as the E specification. The discount rate for the QH specification is the value $d^{QH}(t)$ that solves $D^{QH}(t) = 1/(1+d^{QH})^t$, that is $d^{QH}(t) = [\beta/(1+\delta)^t]^{(-1/t)} - 1$ for $t > 0$. Thus for $\beta < 1$ a sharply declining discount rate is observed in the very short run, and thereafter the discount rate tends asymptotically to δ as the effects of the initial drop in the discounting factor diminishes. Since it does not vary with the time horizon once $t > 0$, the drop 1-β can be viewed as a fixed utility cost of discounting anything relative to the present.

There has been a whole family of *hyperbolic* specifications of the discounting function. The simplest specification (*HA*) assumes a discount factor given by

$$D^{HA}(t) = 1/t \quad (6)$$

with discount rate $d^{HA}(t) = t^{(1/t)} - 1$. A flexible *Weibull* (*W*) specification is based on the Weibull probability distribution, and is defined as

$$D^W(t) = \exp\left(-\rho t^{(1/\sigma)}\right) \quad (7)$$

for $\rho > 0$ and $\sigma > 0$. For $\sigma = 1$ the *W* specification collapses to the *E* specification. Coherently with the idea of non-linear subjective time perception, the parameter σ can be viewed as reflecting the "slowing down" or "speeding up" of time as perceived by the individual. The discount rate at time *t* in the *W* specification is $d^W(t) = \exp\left(\rho t^{(1-\sigma)/\sigma}\right) - 1$.

Experimental Design and Experimental Tests

In the experiments that jointly elicit and estimate risk and time preferences, subject typically answer two distinct sets of binary questions. At the end of the experimental session, one of the questions is randomly selected to be played for real, and subjects receive real money based on their preferred option in that question.

Time preferences questions are usually repeated both with and without a front-end delay (FED), to capture the possible occurrence of non-constant discount rates. Although ordered "multiple price lists" have been used to elicit risk and time preferences, more recent applications use unordered batteries of choices.

Econometrics

Following the behavioral econometrics approach, it is possible to directly estimate using ML a structural model of the latent choice process in which the underlying behavioral parameters defining time and risk preferences are estimated.

In the risk preferences tasks each of the lotteries considered has two possible monetary outcomes M_j, $j=1,2$ with probabilities $p(M_j)$ induced by the experimenter. Under EUT, the expected utility (EU) for lottery *i* is:

$$EU_i = \sum_{j=1,2}\left[p\left(M_j\right) * U\left(\omega + M_j\right)\right] \quad (8)$$

In the likelihood estimation, a simple stochastic specification is used, which allows some behavioral Fechner errors, and also accounts for "contextual errors": the EU for each lottery pair is calculated for candidate estimates of the parameters of the utility function *U(.)*, and the ratio:

$$\Delta EU = [(EU_A - EU_B)/\nu]/\mu \quad (9)$$

is calculated, where EU_A refers to Option A and EU_B to Option B; ν is a normalizing "contextual" term defined as the maximum utility over all prizes in that lottery pair minus the minimum utility over all prizes in that lottery pair; and μ is the "noise parameter" used to allow some errors from the perspective of the deterministic EUT model. As $\mu \longrightarrow 0$ this specification collapses in the deterministic choice EUT model, while as μ gets larger the choice essentially becomes random.

The latent index ΔEU is then linked to observed choices using a logistic cumulative probability distribution function $\Lambda(\Delta EU)$, so that Prob(choose lottery A) = $\Lambda(\Delta EU)$. Thus the likelihood of the series of lottery choices, conditional on the EUT and CRRA specifications, depends on r, μ, and the observed choices. Since in the experimental tasks subjects are not allowed to express indifference between the two options, the conditional log-likelihood for a generic individual taking part in the experiment is:

$$lnL^{RP}(r, \mu; y, \omega, X) = \sum_i [(\ln \Lambda(\Delta EU) \,|y_i = 1) + (\ln(1 - \Lambda(\Delta EU)) \,|y_i = -1)] \quad (10)$$

where $y_i = 1$ ($y_i = -1$) denotes the choice of Option B (A) lottery in risk preferences task *i*. The CRRA r is allowed to linearly depend on a vector X of individual characteristics, $r = r_o + R'X$ where r_0 and R are (scalar and vector) parameters to be estimated.

A similar specification is used for the discount rate choices. The latent index is given by the difference of the discounted utility of each of the two options, conditional on some assumed discount factor *D(t)*. The likelihood function for these choices is a function of both the risk parameter *r* and the discount factor *D(t)*. The same stochastic error specification as in the lottery choices can be used, but with a different Fechner error term υ. Moreover, in the context of the time preferences tasks, where the choices are over deterministic outcomes, it is not necessary to apply the "contextual" error correction. Hence the utility differences between the two options ΔPV is given by:

$$\Delta PV = (PV_A - PV_B)/\upsilon \quad (11)$$

where the discounted utility of Option A (PV_A) is:

$$PV_A = D(\tau)U(\omega + M_\tau) + D(\tau + t)U(\omega) \quad (12)$$

and the discounted utility of Option B (PV_B) is:

$$PV_B = D(\tau)U(\omega) + D(\tau + t)U(\omega + M_{\tau+t}) \quad (13)$$

where τ is equal to 1 month if there is front-end delay, and zero otherwise; and *t* can be 1, 3 or 12 months. Thus, the likelihood of the discount rate responses, conditional on the EUT, CRRA, and *D(t)* discount factor being true, depends on *r*, given the value of the observed choices. The conditional log-likelihood is:

$$lnL^{TP}(r, \mu, D(t), \upsilon; y, \omega, X) = \sum_i [(\ln \Lambda(\Delta PV) \,|y_i = 1) + (\ln(1 - \Lambda(\Delta PV)) \,|y_i = -1)] \quad (14)$$

where again $y_i = 1$ ($y_i = -1$) denotes the choice of Option B(A) in discount rate task *i*, and X is a vector of individual characteristics as discussed above.

The joint likelihood of the risk and time preferences responses is then:

$$lnL(r, \mu, D(t), \upsilon; y, \omega, X) = lnL^{RP} + lnL^{TP} \quad (15)$$

This expression can then be maximized using standard numerical methods.

The lab experiments that elicit risk and time preferences for student samples typically assume that ω is set equal to zero for all subjects. Experimental elicitations of risk and time preferences with representative samples of the population either set this level equal to average daily income level in the country, or take advantage of the fact of being linked to an asset portfolio from administrative records or to a household survey, to calculate individual-specific values of the background consumption ω. (Andersen et al., 2016, and Galizzi, et al., 2016, respectively).

Applications to Health

Even though several studies have looked at the links between risk and time preferences for monetary outcomes and health behaviors (Anderson and Mellor, 2008; Sutter et al., 2013; Szrek et al. 2012), relatively few studies have done so using the behavioral econometrics approach.

In an artefactual field experiment, Harrison, Lau and Rutström (2010) 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.

In a lab experiment, 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.

In another lab experiment, 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 an artefactual field experiment with a representative sample of the UK population, Galizzi et al. (2016) assess the external validity of three different measures of risk preferences including the MPL method, and find that the MPL measure of risk aversion significantly predict respondents' BMI and the consumption of fruit and vegetables, but is not significantly associated to their smoking or drinking habits, nor to the consumption of junk food. By structurally estimating the risk and time preferences of the same UK representative sample, Galizzi et al. (2016) find evidence of non-exponential discounting, and look comprehensively at the associations between various estimated models of risk and time preferences and a broad range of health behaviors contained in the linked UK Longitudinal Household Survey.

Conclusions

Risk and time preferences are of fundamental interest for health economics research and policy as they are the ultimate drivers of a wide range of health-related behaviors and directly inform the principles of cost-benefit analysis moving beyond cost-effectiveness analysis. We reviewed and assessed the main methods for the measurement of risk and time preferences. We then illustrated how risk and time preferences for monetary outcomes can be elicited using incentive-compatible experimental methods, and how a range of models of risk and time preferences can be structurally estimated using these experimental data. We also described the major applications to date on the links between structurally estimated risk and time preferences and health behavior.

We have demonstrated that there is much more to behavioral insights and experimental methods in health economics than just nudges or RCTs. As we demonstrated for the case of estimating risk and time preferences tailoring and fine-tuning the broad spectrum of behavioral experiments from the lab to the field to address pressing health policy and research questions in developed countries is, both methodologically and substantially, one of the most challenging, promising, and exciting areas of applications of experimental methods and behavioral insights to health economics.

References

Andersen, S., Cox, J.C., Harrison, G.W., Lau, M., Rutström, E.E., & Sadiraj, V. (2016). Asset integration and attitudes to risk: theory and evidence. *CEAR Working Paper 2011-17*, Georgia State University.

Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2008a). Eliciting risk and time preferences. *Econometrica*. 76(3), 583-618.

Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2008b). Lost in state space: are preferences stable? *International Economic Review*. 49(3), 1091-1112.

Andersen, S., Harrison, G.W., Lau, M., & Rutström, E.E. (2014). Discounting behavior: a reconsideration. *European Economic Review*. 71, 15-33.

Anderson, L.R., & Mellor, J.M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*. 27(5), 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.

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.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2009). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.

Barseghyan, L., Prince, J., & Teitelbaum, J.C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*. 101, 591–631.

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.

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.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016): Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision, *Journal of Economic Behavior & Organization*, 131, 17-23.

Buckley, N., 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.

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.

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.

Charness, G., Gneezy, U., & Imas, M. A. (2013). Experimental methods: extra-laboratory experiments – extending the reach of experimental economics, *Journal of Economic Behavior & Organization.* 91, 93-100.

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.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48, 424-455.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G.G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*. 9(3), 522-550.

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.

Einav, L., Finkelstein, A., Pascu, I., Cullen, M. (2012). How general are risk preference? Choices under uncertainty in different domains. *American Economic Review*. 102(6), 2606-2638.

Ferber, R., & Hirsch, W. Z. (1978). Social experimentation and economic policy: a survey. *Journal of Economic Literature.* 16(4), 1379-1414.

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.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*. 351–401.

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., Harrison, G. W., & Miniaci, R. (2016). 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., 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.

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*, in press.

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.

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., & Rutström, E.E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization* 70(3), 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(5), 708-717.

Harrison, G.W., Lau, M.I., & Rutstrom, 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., Lau, M.I., & Williams, M.B. (2002). Estimating individual discount rates in Denmark: a field experiment. *American Economic Review*, 92(5), 1606-1617.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*. 42, 1009-1055.

Harrison, G.W., & Ng, J.M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance*, 83(1), 91-120.

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.

Harrison, G.W., & Rutström, E.E. (2009). Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics*, 12, 133-158.

Haynes, L., Service, O, Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomized controlled trials. Cabinet Office Behavioural Insights Team:

https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials.

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.

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.

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.

Kramer, M., & Shapiro, S. (1984). Scientific challenges in the application of randomized trials. *Journal of the American Medical Association*. 252, 2739-2745.

Leamer, E.E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives*. 24(2), 31-46.

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. (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.

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.

Prosser, L.A., & Wittenberg, E. (2007). Do risk attitudes differ across domains and respondent types? *Medical Decision Making*. 27, 281–287.

Rubin, D. B. (1974). Estimating the causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*. 66, 688-701.

Sunstein, C. R. (2011). Empirically informed regulation. *University of Chicago Law Review*. 78, 1349-1429.

Sutter, M., Kocher, M.G., Glätzle-Reützler, 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.

Tanaka T., Camerer C., & Nguyen Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*. 100(1), 557-571.

Van der Pol, M. (2011). Health, education and time preference. *Health Economics*. 20, 917–929.

Van der Pol, M., & Cairns, J. (2001). Estimating time preferences for health using discrete choice experiments. *Social Science and Medicine*. 52, 1459–1470.

Van der Pol, M., & Cairns, J. (2008). Comparison of two methods of eliciting time preference for future health states. *Social Science and Medicine*. 67, 883–889.

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, *JAMA*. 300(22), 2631-7.

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.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*. 42(8), 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.