Article:

Pedersen, LB, Hess, S orcid.org/0000-0002-3650-2518 and Kjaer, T (2016) Asymmetric information and user orientation in general practice. Exploring the agency relationship in a best-worst scaling study. Journal of Health Economics, 50. pp. 115-130. ISSN 0167-6296

https://doi.org/10.1016/j.jhealeco.2016.09.008

Asymmetric information and user orientation in general practice. Exploring the agency relationship in a best–worst scaling study

Line Bjørnskov Pedersen[1,2], Stephane Hess[3], Trine Kjær[1]

[1] Centre of Health Economics Research (COHERE), Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[2] Research Unit for General Practice, University of Southern Denmark, J.B. Winsløwsvej 9A, 1, 5000 Odense C, Denmark

[3] Institute for Transport Studies & Choice Modelling Centre, University of Leeds, Lifton Villas, 1-3 Lifton Place, Leeds LS2 9JT, UK

Abstract

This study uses a best–worst scaling experiment to test whether general practitioners (GPs) act as perfect agents for the patients in the consultation; and if not, whether this is due to asymmetric information and/or other motivations than user orientation. Survey data was collected from 775 GPs and 1379 Danish citizens eliciting preferences for a consultation. Sequential models allowing for within-person preference heterogeneity and heteroskedasticity between best and worst choices were estimated. We show that GPs do not always act as perfect agents and that this non-alignment stems from GPs being both unable and unwilling to do so. Unable since GPs have imperfect information about patients' preferences, and unwilling since they are also motivated by other factors than user orientation. Our findings highlight the need for multi-pronged strategies targeting different motivational factors to ensure that GPs act in correspondence with patients' preferences in areas where alignment is warranted.

Keywords: Agency theory, general practice, asymmetric information, user orientation, best–worst scaling case 3

1. Introduction

The patient–doctor relationship describes one of the cornerstones in the universal provision of health care services. Understanding and correct mapping of the relationship is important as it provides a basis for the design of efficient economic incentive schemes, and in the analysis of demand for and optimal delivery of health care services. This is especially true for primary care, as demonstrated by the substantial research within this area over the last decades including – but not limited to – research on remuneration systems and how they affect GP behavior (e.g., Gosden et al. 2000; Scott and Shiell 1997; van den Berg et al. 2009; Henning-Schmidt et al. 2011), GPs' roles as gatekeepers (e.g., Brekke et al. 2007; Dusheiko et al. 2006), supplier-induced demand (e.g., Dijk et al. 2013; Labelle et al. 1994), and patient moral hazard (e.g., Dijk, Berg, Verheij, Spreeuwenberg, Groenewegen, and Bakker 2013; Doran et al. 2005). During the last decades, this research has provided valuable input to health policy makers. Still, many issues remain unresolved including the influence of different motivational factors on GP behavior, and the GPs role in optimizing patients' pathways and in the delivery of patient-centered primary care services.

The patient–doctor relationship deviates from the assumptions made in traditional agency theory on several well-known dimensions (see, e.g., Blomqvist 1991; Gafni et al. 1998; Mooney and Ryan 1993; Ryan 1994; Scott 2000). First, the relationship is characterized by double-sided asymmetric information; that is, GPs have incomplete information on patients' health status, behavior and preferences for treatment on one side, and patients have incomplete information on clinical diagnosis and treatment options on the other side. Second, the GP is assumed to take the utility of the patient into account when maximizing his/her own utility; that is, having a genuine concern for the welfare of the patient, thereby acting as an agent for the patient. This altruistic motivation has in the literature been referred to as 'user orientation' to signal the public service provider's interest in doing good for the individual recipient (the patient) as opposed to 'public service motivation' concerning the interest in doing good for collective entities (all patients) (Vandenabeele 2008; Andersen et al. 2011; Jensen and Andersen 2015). Accordingly, GPs can be considered double agents, meaning that they have to satisfy two principals – the individual patient consulting the GP, and the health care authorities contracting with the GPs and representing the joint interest of all patients. Given this triangular nature of the agency relationship, it is likely that the interests of the GP and the individual patient are not always perfectly aligned. Thus, non-alignment could stem from uncertainty about the patient's preferences (that is, asymmetric information) and/or conflict between user oriented motivation and other motivations such as personal financial incentives[1], public service motivation[2], and/or paternalism (e.g. focusing more narrowly on the patient's health benefits). Consequently, GPs are expected to make trade-offs between the individual and the collective interest which may cause GPs to refrain from acting as perfect agents for the patients in order to optimize scarce resources[3].

Inspired by the work of (among others) Cheraghi-Sohi et al. (2008) and Scott and Vick (1999), we set out to further investigate the GP–patient encounter by studying both patients' and GPs' preferences for characteristics of a consultation using a stated preference (SP) methodology. With the application of a split questionnaire design, we also elicited GPs' perceptions of patients' preferences by asking them to choose the type of consultation they believe their typical patient would prefer. With this approach, we seek to identify the extent to which GPs are well-informed about patients' preferences (whether there is any asymmetric information), and to what extent GPs refrain from being perfect agents for the patients. This enables us to answer the following three research questions: (1) Do GPs act as perfect agents for the patients (comparison of GPs' preferences with patients' preferences)? If not, is this due to (2) uncertainty about patients' preferences and therefore asymmetric information between GPs and patients (comparison of GPs' perceptions of patients' preferences with patients' preferences), and/or (3) deviations caused by other motivational factors than user orientation (comparison of GPs' perceptions of patients' preferences with GPs' preferences)? That GPs' actions are not always congruent with the preferences of their individual patients is expected due to GPs' double agency. That GPs do not always know patients' preferences is of more concern, since the presence of asymmetric information implies that GPs cannot act as perfect agents for their patients even in situations where they intend to do so. This potential problem is highly relevant in light of the widespread focus on patient satisfaction as an indicator of quality and the standard of services more broadly, and for the delivery of patient-centered care (Greenfield and Braithwaite 2008).

Previous studies have used SP methods to investigate the concurrence of doctors' and patients' preferences (e.g., Carlsen and Aakvik 2006; Payne et al. 2011; Van den Hombergh et al. 2005; Vedsted et al.

---

[1] The literature has traditionally distinguished between extrinsic and intrinsic motivation, where intrinsic motivation is defined as the doing of an activity for the enjoyment of the activity itself in contrast to extrinsic motivation where an activity is done in order to attain some separable outcome (that is, external reward) such as financial incentives (Benabou and Tirole 2003; Ryan and Deci 2000; Frey and Jegen 2001).

[2] According to Jacobsen et al. (2013), and in line with Ryan and Deci (2000), public service motivation can be seen as a type of internalized extrinsic motivation.

[3] By perfect agency we refer to a situation in which the GP works as a perfect agent for the patient thus making the same decision as the patient would, were the patient to be party to the same clinical expertise as the doctor. This also includes taking the patient's cost, such as time, into account.

2002), whereas other SP studies have compared doctors' perceptions of patients' preferences with patient preferences (e.g., Cox et al. 2007; Marshall et al. 2009; Mühlbacher and Nübling 2010; Neuman and Neuman 2009; Pedersen et al. 2012). Evidence on the degree to which doctors know patients' preferences is mixed, as are results on whether doctors' and patients' utility functions are aligned, with the majority of studies suggesting that differences exist. However, none of these studies are capable of identifying the extent to which possible discrepancies stem from doctors pursuing other goals than to satisfy patient preferences (such as financial incentives or public service motivation), or from asymmetric information between GPs and patients[4]. One way to explore the agency relationship in greater detail is to include all three perspectives (patient, GP, and GP perception) in one study. Although significant information can be obtained from such a setting, none of the aforementioned studies have attempted to do this. The present study fills this gab in the literature by providing a first attempt at a combined investigation of the agency relationship in general practice using a SP framework[5].

We use the agency model to study preferences for characteristics of a consultation not involving treatment decisions. Hence, clinical considerations on diagnosis and treatment are of less concern, implying that asymmetry of information due to patients' lack of clinical knowledge can be disregarded as a potential (and acceptable) explanation for divergence in preferences. We survey a random sample of the Danish population and GPs and apply a best–worst scaling case 3 (BW3) experiment (Lancsar et al. 2005; Louviere et al. 2015; Louviere et al. 2004; Marley and Louviere 2005; Marley and Pihlens 2012). The BW3 experiment is a SP approach and a variant of the more traditional choice experiment. In a BW3 experiment respondents are presented with a number of choice sets and are, in each choice set, asked to choose not only the best alternative (as in a traditional choice experiment) but also the worst alternative among those available. BW3 has received attention in the literature since it provides richer information on preferences compared to traditional choice experiments, due to the possible insights into the full preference ordering of respondents. BW3 is ideal in exploration of the agency relationship since more knowledge is obtained on the existence of possible discrepancies in preferences compared to, for example, traditional choice experiments. Hence, the method is useful in obtaining additional choice information and in providing a better understanding of the process of preference formation. Furthermore it enables us to gain knowledge on for which specific elements in the consultation the asymmetry in information exists, and on which

[4] Other studies have focused on the trade-off between altruistic and financial incentives. This includes recent experimental research by Godager and Wiesen (2013), studying the degree of GPs altruism in choice of medical treatment.

[5] Only one study (Hirth et al. 2000) has previously attempted this approach, albeit with a specific policy focus and without analyzing the findings in the theoretical context of the agency relationship.

elements GPs deviate from acting as perfect agents for both best and worst choices. Our data is analyzed using a sequential best–worst approach via a multinomial logit (MNL) specification that takes account of within-person heterogeneity and heteroskedasticity; that is, allowing for differences in scale and utility within best–worst choices. To date, this approach has only been used in a few other studies (e.g., Scarpa et al. 2011; Louviere et al. 2015), and never in the area of health economics. Our results show that GPs predominantly know patients' preferences within a consultation, but that asymmetric information exists when it comes to patients' most preferred aspect of the consultation, which involves discussions on their general health status and lifestyle. Furthermore, we find evidence that GPs are not solely user oriented and refrain from being perfect agents for the patients. This is the case, for instance, with respect to time allocated to explain the problem during the consultation. These results are robust to various specification checks.

The remainder of the paper is structured as follows. Section 2 provides an overview of the organizational context of general practice in Denmark, and describes our theoretical model on the agency relationship in general practice. Section 3 describes the survey design, data collection, statistical models, and hypotheses to be tested. In Section 4 results are presented and these are further discussed in section 5. Section 6 offers our conclusions.

2. Theoretical underpinning

2.1. The setting

Like in most other European countries, GPs in Denmark are private entrepreneurs on contract with third-party payers; that is, the Danish Regions. Collective agreements between GPs and the Danish Regions are bargained approximately every second year. The collective agreements define, among other things, organization of the remuneration system. GPs in Denmark are currently remunerated by capitation (approximately one third of their income) and fee-for-service (approximately two thirds of their income). For each consultation the GPs undertake, they are paid a fixed fee of 135.64 DKK (corresponding to 18.22 EUR/22.96 USD as of October 2015). This is independent of the time spent on the consultation. During the consultation, the GP can provide non-contractible services such as advice regarding health problems, and contractible services including blood tests. The contractible efforts are reimbursed with a fixed fee per service undertaken.

All citizens in Denmark are automatically assigned a GP, and 87 percent of citizens visit their GP at least once per annum, with an average of seven contacts per year including e-mail and telephone consultations (Statistics Denmark 2011). There are no user fees for consulting a GP (consultations are free of charge), although some services must be paid for by the individual patient (for instance, vaccinations). However, access to primary care is rationed by barriers to entry since health care authorities decide the number of GPs given permission to practice, based on the number of Danish nationals. This, together with a structural shortage of GPs in Denmark, causes a scarcity in the supply of GP consultations – especially in the rural areas of Denmark. Normally, a non-acute consultation lasts 10–15 minutes.

This study is based in the contemporary primary care setting as described above. In the experiment, GPs and patients were asked to choose their most and least preferred consultations that vary regarding the content of the consultation. This has several implications for the theoretical model and interpretation of our findings – most importantly, it suggests that GPs' choices should be modeled and interpreted on the basis of the current remuneration system, with GPs facing an implicit time constraint.

2.2. Agency theory in general practice

According to neoclassical economic theory, agents and principals seek to maximize utility by acting in their own best interest. It is assumed that individuals have complete and fully ordered preferences and make rational choices that satisfy their preferences over other possible choices. In such a case, utility functions can be used to describe the preferences of the parties. In traditional labor economics, utility functions of employed agents consist, among other things, of wage and effort (Ashenfelter and Card 2010). We can develop this standard utility function to apply to remunerated agents in the health care sector, such as GPs. Accordingly, GPs derive utility from remuneration and disutility from effort. We focus on the patient–doctor encounter described in terms of a consultation.

Consider a model where the GP faces a choice of consultation from a set of mutually exclusive consultations. Effort, e, denotes a function of services provided during a consultation, $\mathbf{X}$. In the definition by McGuire (2000), effort includes all actions the GP undertakes within the consultation, including medical actions and non-contractible efforts such as diligence, care, and attentiveness. Due to our focus in the empirical analysis, effort is restricted to non-contractible effort, such as time spent listening to the patient's problem, time used for reaching a shared decision, time for talking about other health-related issues, etc.

Hence, we define $\mathbf{X}$ as a vector of the non-contractible actions the GP undertakes during the consultation[6]. In line with McGuire (2000), time spent in the consultation is an implicit dimension of the services provided, and thus varies with $\mathbf{X}$. As standard we assume that the patient derives utility from the actions undertaken by the GP. We define patient utility broadly allowing non-contractible effort to generate utility beyond direct health/treatment benefits through e.g. attentiveness and reassurance (Brouwer et al. 2008; Mooney and Ryan 1993). Let B represent the patient's true utility associated with a consultation, and let B* represent the GP's perception of the patient's utility associated with the consultation. Inspired by Ellis (1990), Ellis and McGuire (1986) and McClellan (1996), the GP incorporates the patient's perceived utility into his/her own utility function, denoted *U* (Evans 1984; Mooney and Ryan 1993). Furthermore we assume that the GP's choice of consultation, and therefore intensity of services, is based on the trade-off between the opportunity cost (of effort and time) and the utility gained from treating the patient and is allowed to vary across GPs. Consequently the GPs are assumed to react to different motivational factors with different weights in their utility functions, such as placing more or less importance on perceived patient utility. We therefore introduce α as a relative weight on the trade-off between different motivational factors of the GP. Finally we assume that GPs are remunerated for the consultation with a fixed fee, *p*, which is independent of the choice of consultation (i.e. time and effort spent on the consultation). Correspondingly, the maximization problem of the GP for a consultation can be written as:

$$(1) \qquad Max_{\mathbf{X}}U = \alpha\big(p - e(\mathbf{X})\big) + (1-\alpha)B^*\big(e(\mathbf{X})\big), \quad \alpha \in \{0;1\}, \quad \mathbf{X} \in (X_1, X_2, \dots, X_n)$$

where $B^* = B$ when the GP has perfect information about the patient's utility function. According to the model, α equals 0 if the GP is purely user orientated focusing on the individual patient's wellbeing. In this extreme case, the GP only derives positive utility from effort and does not take the opportunity costs into consideration[7]. Contrary, α equals 1 if the GP does not gain any utility from the effort invested in the individual patient. In this extreme case the GP is motivated by other factors than user orientation and seek to minimize effort invested in the individual patient in order to maximize income and/or time for other patients.

[6] Ellis (1990) defines $\mathbf{X}$ as intensity in services determined by patient type, whereas McClellan (1996) defines intensity of services as a dichotomous variable (less intensive and more intensive) that depends on the expected benefit to the patient.

[7] The opportunity cost of effort describes the time the GP could use on leisure activities and/or on treating other patients. The latter could be motivated by both personal financial incentives (as more consultations generate more income) and public service motivation (as the GP serves the joint interest of all patients; that is, society, by seeing more patients).

It is assumed that all variables are continuous and twice differentiable. Furthermore we assume that effort increases with actions undertaken in the consultation at an increasing rate, and that the perceived utility of patients increases with effort at a decreasing rate. Hence, assuming that $e(X_i)$ is separable in each action X_i with i = 1,….,*n*:

(2) $$e'(X_i) > 0; e''(X_i) > 0; B^{*\prime}(e) > 0; B^{*\prime\prime}(e) < 0$$

The solution to the maximization problem in equation (1) is found by taking the first-order derivative of *U* with respect to X_i. This entails that in optimality the marginal utility of increasing (perceived) patient utility equals the marginal utility of effort by performing one more task during the consultation. This is expressed in equation (3):

(3) $$(1\text{-}\alpha)\frac{\partial B^*}{\partial e}\frac{\partial e}{\partial X_i}=\alpha\frac{\partial e}{\partial X_i}$$

The model demonstrates the general trade-off the GP faces; more effort for a single patient (user orientation) entails an economic loss and thus a financial disincentive (in terms of lower payment per hour) and less time to consult with other patients (public service motivation), but also generates a positive effect from the perceived gain in utility for the individual patient (user orientation). How these effects are weighted against each other depends on the motivational factor α . In the empirical analyses we estimate the utility associated with the non-contractible services provided during the consultation, $\mathbf{X}$. Although our experimental setup does not allow us to estimate α , we do obtain implicit knowledge on α , and can distinguish between the extreme case (α = 0 and α = 1) and the non-extreme case $\alpha \in (0,1)$. In this regard our study complements the experimental findings by Godager and Wiesen (2013), where three groups with different alphas were identified.

Our focus is on three interrelated research questions: (1) do GPs act as perfect agents for the patients? And if not, is this due to (2) asymmetric information and/or (3) deviations due to other motivational factors than user orientation as a result of their double agency? If GPs have incomplete information on patients' preferences, they will not be able to correctly take account of patients' preferences, even if they have the intention to do so (that is, α < 1). On the other hand, GPs might be fully aware of patients' preferences for consultations, but decide (partly) not to act accordingly, with the extreme case of α =1. Based on the model (equations 1–3), we purport the following three hypotheses as a basis for the subsequent empirical analysis:

H1: *Perfect agent hypothesis*: The GP acts as a perfect agent for the patient if both of the following conditions are satisfied: 1) there is complete information (H2), and 2) the GP is purely user oriented (H3). This implies that $U = B$.

H2: *Complete information hypothesis*: The GP has complete information on the patient's preferences for the consultation, such that $B = B^*$. This is a necessary but not sufficient condition for H1 to hold.

H3: *User orientation hypothesis*: The GP is purely user oriented and does not take the opportunity cost of time and effort into consideration in the choice of consultation; that is, α = 0, and hence $U = B^*$. This is a necessary but insufficient condition for H1 to hold.

3. Methods

3.1. Survey design

Choice experiments have been used extensively in health economics to study preferences for health and for (the delivery of) health care services (see e.g. Clark et al. 2014). Recently, related methods that extend design and modelling of the traditional choice experiment have gained terrain, including BW3 experiments[8]. BW3 was introduced by Louviere et al. (2004)[9], and first presented in the health economics literature by Lancsar et al. (2005)[10]. It has recently been axiomatized and theoretically developed by Marley and Louviere (2005) and Marley and Pihlens (2012). In BW3 experiments respondents are presented with a number of hypothetical scenarios each consisting of three or more alternatives differentiated by the attributes of interest and asked to choose both the best and the worst options available. The best and worst terms are subsequently used to define the extremes of a latent, subjective continuum (Louviere et al. 2015). The BW3 experiment (as well as other best–worst scaling methods) provides richer information on preferences compared to traditional choice experiments (which only obtain information on the most preferred option in a choice set), since a fuller preference order of respondents can be obtained (Flynn 2010). This usually results in gains in statistical efficiency when maintaining the typical numbers of choice sets (Lancsar et al. 2013). Moreover, knowledge on respondents' worst case scenarios can be explicitly obtained. The BW3 model is the one that comes closest to traditional choice experiments entailing that the

[8] Also referred to as a best–worst discrete choice experiment (Lancsar et al. 2013; Lancsar and Louviere 2008) or best–worst multi-profile case (Flynn and Marley 2012; Marley and Pihlens 2012).
[9] Later published in Louviere et al. (2008).
[10] Later published in Lancsar et al. (2007).

design of the scenarios in the BW3 experiment is similar to the experimental design of a traditional choice experiment, whereas the econometric modelling differs.

We designed a BW3 experiment to elicit preferences for general practice consultations. We deliberately chose to model the meeting between the GP and the patient as this constitutes one of the most central meetings between health care providers and patients, and mimics an everyday situation in primary care. The experiment went through a thorough design phase. The attributes and levels, as well as the description of the choice situation, were carefully chosen on the basis of information gathered from: a literature review of studies investigating patients' preferences for attributes of GP consultations (e.g., Cheraghi-Sohi et al. 2008; Hjelmgren and Anell 2007; Scott and Vick 1999; Vedsted et al. 2002), visits to different general practices and observational studies of actual consultations, interviews with GPs and patients, and discussions with the Organisation of General Practitioners in Denmark. The final attributes included were: (1) time devoted to explaining the health care problem (*problem*), (2) discussion of possible treatments (*treatment*), (3) scheduling of follow-up visits (*follow-up)*, (4) dialogue about topics other than the specific health problem (*care*), and (5) conversation about patients' general health status and lifestyle (*health status*). The experiment was carefully designed to reflect the content of a standard practice consultation characterized by the specific steps GPs are trained to undertake in a normal (non-acute) consultation. The attributes were all unaffiliated with clinical diagnosis and specific treatment of the disease. Importantly, this obviates any disparities in preferences due to clinical considerations. Consequently, asymmetry of information due to patients' lack of clinical knowledge cannot constitute an explanatory factor in our analysis (as opposed to previous studies (e.g., Hirth et al. 2000)). Table 1 provides an overview of attributes and levels as they were described for GPs and patients, respectively. All attributes are qualitatively described and assigned two levels each.

[TABLE 1]

As noted in Louviere and Lancsar (2009), traditional experimental designs can be used for BW3 experiments. We used an optimal orthogonal in the differences design (OOD) with zero priors (Street and Burgess 2007) generated by means of the software Ngene provided by ChoiceMetrics (ChoiceMetrics 2009). Eight choice tasks were created obtaining 100 percent D-optimality. To keep the questionnaire length at a minimum, avoid too many drop outs (specifically among GPs), and increase response rates, the design was blocked in two such that each respondent received four choice tasks. Blocking was performed by minimizing the average correlation between the blocking column and the attribute column.

Two questionnaires were constructed; one for the patient sample and one for the two GP samples (only differing with respect to the BW3 task). The patient questionnaire was initiated with a number of introductory questions about the respondents' use of and satisfaction with their GP, and questions about preferences for consultations. GPs on the other hand were asked questions about their attitudes towards interaction with patients in the consultation. Prior to the choice task, all respondents were presented with a thoroughly described situation in which they were asked to imagine that, due to mild but fairly long-lasting symptoms (headache and a sore throat), they (the patient) had made an appointment with the GP. The symptoms were described in general terms to keep the situation generic and to avoid leading the GPs to think in terms of specific diagnoses. Respondents were then asked to choose first their most preferred consultation from a choice of three types of consultations, and thereafter their least preferred consultation. In the sample of GPs answering on behalf of patients, GPs were asked to choose most and least preferred consultation according to what they believed the patient would have chosen. An example of the best–worst choice tasks including scenario description as they were presented to the three samples is shown in Appendix A. Subsequent to the choice tasks, questions on the respondents' personal characteristics were included in the questionnaire.

The patient questionnaire was tested in a cognitive interview (Jobe 2003), which led to minor changes. Afterwards, a web-based pilot study with 28 respondents, drawn as a convenience sample from the Danish population, was conducted. While answering the questionnaire, the respondents were encouraged three times to comment on the questionnaire and state if they felt something was missing. As none of the interviewees felt that important aspects had been omitted, it was concluded that all relevant attributes were included. The GP questionnaire was tested in three similar cognitive interviews with GPs from different general practices. The cognitive interviews led to several adjustments relating to the wording and options given in the questionnaire to make the questions realistic for the GPs, and in correspondence with the collective agreement. Afterwards, a paper based pilot study with seven GPs was conducted, which led to minor changes. None of the changes were related to the best–worst scaling exercise.

3.2. Data collection

Patients (that is, members of the general public) were recruited in May 2010 from an Internet panel whose members received an e-mail with a link to the questionnaire. The target sample was a representative sample of 1400 respondents of the general public above 18 years of age, and the link was deactivated when the quota had been met. Inconsistent respondents, defined as respondents choosing the same alternative as best and worst, were removed from the data, and the representativeness of the remaining respondents was tested with respect to age, gender, and geography using *t*-tests for proportions applying a 5 percent significance level. As respondents had to fill in all choice questions before they could proceed in the questionnaire, no incomplete choice tasks were generated.

The GP questionnaires were mailed to a random sample of 1822 GPs, corresponding to half of all GPs in Denmark, in September 2010. To avoid ordering and strategic bias, half of the sample received a questionnaire containing best–worst choice tasks, where they were asked to act in accordance with their own preferences, while the other half were asked to choose best and worst consultations on behalf of a typical patient. The GPs were randomly allocated to the two types of surveys. One reminder was sent out during the data collection process, together with a copy of the questionnaire. After removing incomplete and inconsistent respondents, the representativeness of the GPs with respect to age, gender, and geography was tested using *t* tests for proportions on 5 percent significance levels. In addition, a test for successful randomization between the two samples was performed on the same variables.

3.3. Sequential best–worst models

Data was analyzed in a stepwise process to model how each step affected preference estimates and model fit across the samples. A sequential best–worst approach was taken assuming that preferences were formed in congruence with the question order (i.e. best out of three options first, followed by worst out of the remaining two options). Yoo and Doiron (2013) showed how to model BW3 data in a latent class framework, whereas Lancsar et al. (2013) have demonstrated how to account for preference and scale heterogeneity between persons, assuming constant scale over best–worst choices. This paper shows how to take account of within-person heterogeneity and heteroskedasticity; that is, allowing for differences in scale and utility within best–worst choices assuming constant utility and scale between persons within the three samples. For all three samples (patients, GPs, and GPs acting on behalf of patients), we estimated four types of models in the following order: (1) a generic model where all coefficients across best–worst stages are assumed to be equal, (2) a model with stage-specific scale parameters that takes within-person heteroskedasticity into account, (3) a model with stage-specific scale parameters and separate alternative specific constants (ASCs), and (4) a model with stage-specific coefficients that also takes within-person heterogeneity into account. The stepwise exploration, and the motivation for it, is described in detail below. Separate parameters are estimated for each sample (patients, GPs, and GPs acting on behalf of patients). For simplicity, the exposition below is based on the patient sample but is easily generalized to any sample.

All models are specified using a linear MNL model leading to the following deterministic utility function for patient *n* in choice task *t* and for alternative *j*, with *j=1,…,3*:

(4)
$$V_{n,j,t}^{pt} = \delta_j^{pt} + \beta_1^{pt} X_{problem_{n,j,t}} + \beta_2^{pt} X_{treatment_{n,j,t}}$$
$$+ \beta_3^{pt} X_{follow-up_{n,j,t}} + \beta_4^{pt} X_{care_{n,j,t}} + \beta_5^{pt} X_{health\ status_{n,j,t}}$$

where δ_j^{pt} is the ASC for alternative *j*, where, for normalization, we set $\delta_{j|j=3}^{pt} = 0$. The five separate *X* variables can be grouped together into $X_{n,j,t}$ and refer to the non-contractible actions the GP undertakes during the consultation where β denotes the utility parameter associated with each action.

Ranked data can be expressed in an exploded logit formula. If respondents were asked to choose the best alternative from three options labeled A, B, and C followed by the worst alternative from the remaining two alternatives, then the probability of observing the ranking *A* > *B* > *C* for patient *n* in task *t* can be expressed as:

(5)
$$P_{n,t}(A > B > C) = \frac{e^{V_{n,A,t}^{pt}}}{e^{V_{n,A,t}^{pt}} + e^{V_{n,B,t}^{pt}} + e^{V_{n,C,t}^{pt}}} \cdot \frac{e^{-V_{n,C,t}^{pt}}}{e^{-V_{n,B,t}^{pt}} + e^{-V_{n,C,t}^{pt}}}$$

This is the product of two logit probabilities; that is, decomposing the ranking into two sequential choices. We first have the probability of respondent *n* choosing alternative *A* as the best alternative out of alternatives *A*, *B* and *C*. This leaves the choice of alternative *C* as being the worst out of alternatives *B* and *C*. This can be expressed as a binary logit probability, using the negatives of the utilities to reflect the fact that this is now the choice of the alternative with the lowest utility (rather than the highest). This is the generic best–worst model (Model 1), where scale and utility coefficients are assumed to be equal across best and worst stages (Louviere et al. 2015).

Flynn and Marley (2012) has pointed to an increasing amount of evidence suggesting that the variance of the error term is often not equal for best and worst data. This may be due to respondents being more consistent in worst (or best) choices, implying smaller error variance and larger estimates in absolute size compared to best (or worst) choices. To overcome potential bias in estimation, Louviere et al. (2015) and Flynn and Marley (2012) recommend estimating this term. In Model 2, the assumption about equal scale is therefore relaxed and heteroskedasticity across the two choice tasks is taken into account, and we use:

(6)
$$P_{n,t}(A > B > C) = \frac{e^{\mu_b V_{n,A,t}^{pt}}}{e^{\mu_b V_{n,A,t}^{pt}} + e^{\mu_b V_{n,B,t}^{pt}} + e^{\mu_b V_{n,C,t}^{pt}}} \cdot \frac{e^{-\mu_w V_{n,C,t}^{pt}}}{e^{-\mu_w V_{n,B,t}^{pt}} + e^{-\mu_w V_{n,C,t}^{pt}}}$$

where μ_b and μ_w are the scale parameters for the best and worst choices, respectively, where we normalize the scale to 1 for the best choice (μ_b) for identification reasons.

Model 3 allows for separate ASCs for best and worst choices. If the distribution of best frequencies differs significantly from the distribution of worst frequencies, the constant terms will differ. This was pointed out by Flynn et al. (2007) for best–worst scaling case 2, and is seen in our data sets to also apply to BW3 models.

We therefore now create two separate utility functions as:

(7) $$V^{pt}_{n,j,t,best} = \delta^{pt}_{j,best} + \beta^{pt}_1 X_{problem_{n,j,t}} + \beta^{pt}_2 X_{treatment_{n,j,t}} + \beta^{pt}_3 X_{follow-up_{n,j,t}} + \beta^{pt}_4 X_{care_{n,j,t}} + \beta^{pt}_5 X_{health\ status_{n,j,t}}$$

and

(8) $$V^{pt}_{n,j,t,worst} = \delta^{pt}_{j,worst} + \beta^{pt}_1 X_{problem_{n,j,t}} + \beta^{pt}_2 X_{treatment_{n,j,t}} + \beta^{pt}_3 X_{follow-up_{n,j,t}} + \beta^{pt}_4 X_{care_{n,j,t}} + \beta^{pt}_5 X_{health\ status_{n,j,t}}$$

We still have stage specific scale parameters and the choice probability for the *A>B>C* ranking is now given by:

(9) $$P_{n,t}(A > B > C) = \frac{e^{\mu_b V^{pt}_{n,A,t,best}}}{e^{\mu_b V^{pt}_{n,A,t,best}} + e^{\mu_b V^{pt}_{n,B,t,best}} + e^{\mu_b V^{pt}_{n,C,t,best}}} \cdot \frac{e^{-\mu_w V^{pt}_{n,C,t,worst}}}{e^{-\mu_w V^{pt}_{n,B,t,worst}} + e^{-\mu_w V^{pt}_{n,C,t,worst}}}$$

Louviere et al. (2015) point out that it may be naïve and misleading to assume that best and worst choices reflect mirror image values due to potential differences in decision rules in best and worst choices, and Giergiczny et al. (2013) show that this is indeed the case. Hence, Model 4 allows for within-person heterogeneity given relaxing of the assumption that the utility of choosing an option as worst is the exact negative of the utility of choosing it as best. This is the same as allowing utility coefficients to be stage specific (that is, estimating MNL models for each stage). We then have:

(10) $$V^{pt}_{n,j,t,best} = \delta^{pt}_{j,best} + \beta^{pt}_{1_{best}} X_{problem_{n,j,t}} + \beta^{pt}_{2_{best}} X_{treatment_{n,j,t}} + \beta^{pt}_{3_{best}} X_{follow-up_{n,j,t}} + \beta^{pt}_{4_{best}} X_{care_{n,j,t}} + \beta^{pt}_{5_{best}} X_{health\ status_{n,j,t}}$$

and

(11) $$V_{n,j,t,worst}^{pt} = \delta_{j,worst}^{pt} + \beta_{1_{worst}}^{pt} X_{problem_{n,j,t}} + \beta_{2_{worst}}^{pt} X_{treatment_{n,j,t}}$$
$$+ \beta_{3_{worst}}^{pt} X_{follow-up_{n,j,t}} + \beta_{4_{worst}}^{pt} X_{care_{n,j,t}} + \beta_{5_{worst}}^{pt} X_{health\ status_{n,j,t}}$$

The full set of stage specific parameters now directly allow for scale differences, and we no longer need an additional scale parameter, such that the choice probability for the *A*>*B*>*C* ranking is now given by:

(12) $$P_{n,t}(A > B > C) = \frac{e^{V_{n,A,t,best}}}{e^{V_{n,A,t,best}} + e^{V_{n,B,t,best}} + e^{V_{n,C,t,best}}} \cdot \frac{e^{-V_{n,C,t,worst}}}{e^{-V_{n,B,t,worst}} + e^{-V_{n,C,t,worst}}}$$

All models were estimated with robust standard errors clustered at the respondent level to account for the panel structure of the data.

The different model specifications were tested against each other using log-likelihood (LL) ratio tests. For the patient sample, the stepwise exploration showed statistical significant improvements in model fit for each step, with the greatest improvement with the introduction of separate ASCs (Model 3). For the GP sample and the sample of GPs answering on behalf of patients, no statistical significant improvements in fit were found from Model 1 to Model 2 (the scale parameter was insignificant), while Model 3 was found to be superior to Model 2, and Model 4 was superior to Model 3 in both samples. Hence, in the GP samples, within-person heterogeneity was the main driver for statistical superiority of Model 4, while the presence of both within-person heteroskedasticity and heterogeneity made Model 4 most appropriate for the patient sample. Importantly, however, main results are robust across the different specifications, and we therefore decided in the subsequent analysis to focus on Model 4 taking both within-person heteroskedasticity and heterogeneity into account. Other results are available from the authors upon request.

3.4. Hypotheses to be tested

We set out to empirically test the theoretical model described in terms of the three hypotheses previously outlined. We elicited preferences for the non-contractible effort vector, $\mathbf{X}$, defined in terms of the five attributes displayed in equation (10) and (11) and Table 1. All three null hypotheses were tested using the LL ratio test on equality in utility parameters while controlling for differences in scale parameters across samples (Swait and Louviere 1993)[11]. In the following, let vector $\boldsymbol{\beta} = (\delta_j, \beta_{1,\ldots}\beta_5)$ represent the utility parameters of the attribute variables as depicted in the above equations. Furthermore let $\boldsymbol{\beta}^{pt}$, $\boldsymbol{\beta}^{GP}$ and $\boldsymbol{\beta}^{GP\_pt}$ represent the vector of the utility parameters of the three samples respectively (patients, GPs and GPs answering on behalf of patients).

For H1 (*Perfect agent hypothesis)* we tested the following null hypothesis for best and worst choices respectively:

(13) $$\mathrm{H}_0^1: (\boldsymbol{\beta}^{pt} - \boldsymbol{\beta}^{GP}) = 0$$

H_0^1 will be rejected if GPs and patients differ systematically in their choices of best and worst consultations. Accordingly, a rejection of H_0^1 entails that GPs do not act as perfect agents for their patients. Whether this disparity is attributed to information asymmetry and/or caused by over motivations was further tested in H2 and H3.

For H2 (*Complete information hypothesis*), we tested the following null hypothesis:

(14) $$\mathrm{H}_0^2: (\boldsymbol{\beta}^{pt} - \boldsymbol{\beta}^{GP\_pt}) = 0$$

H_0^2 will be rejected if the GPs' choices of best and worst consultations when answering on behalf of the patient differ significantly from the patients' choices. This entails that asymmetric information between GPs and patients is present, implying that GPs are not able to correctly take account of patients' preferences even if they intend to do so (that is, $\alpha < 1$).

Finally, H3 (*User orientation hypothesis)* was tested against the following null hypothesis:

(15) $$\mathrm{H}_0^3: (\boldsymbol{\beta}^{GP} - \boldsymbol{\beta}^{GP\_pt}) = 0$$

H_0^3 will be rejected if preference estimates for GPs' choices of consultations when answering on behalf of the patient differ significantly from GPs' own choices. Thus, if H_0^3 is rejected this will entail that $\alpha > 0$, implying that GPs take factors other than patients' preferences into consideration in their choice of consultation. This suggests that GPs refrain from acting as perfect agents due to other motivational factors, such as opportunity costs of time and effort and economic benefits.

[11] More specifically, the LL test is conducted using the following formula: $-2(\mathrm{LL}_\mu - (\mathrm{LL}_1 + \mathrm{LL}_2)), \chi^2(\mathrm{k}+1)$, where LL_μ is the log likelihood estimate from the pooled model allowing for differences in scale, LL_1 and LL_2 is the log likelihood estimates for the separate models, and *k* is the number of parameters (Swait and Louviere 1993).

In addition to the overall testing of the hypotheses, we explored possible discrepancies in utility parameters across samples by performing pairwise comparisons of attribute estimates. This was done via graphical explorations of patterns in rank orders and overlap in confidence intervals within samples. Since it is not possible to compare the absolute size of utility coefficients across samples due to differences in scaling (Train 2003), we calculated standardized relative importance scores with the most important attribute set at 1.00. This approach is widely used in the health economics literature (e.g. Wijnen et al. 2015; Hauber et al. 2009). Supplementary, to test for the influence of particular attributes on overall differences across samples, we calculated marginal rate of substitution (MRS) matrices for all possible combinations of attributes and tested for statistical significant differences in MRS across samples using *t* tests (results are available upon request).

4. Results

4.1. Descriptives

A total of 1435 respondents from the Danish population answered the patient questionnaire. After removal of inconsistent respondents, the final sample consists of 1379 respondents. The patient sample is representative on gender and geography, while young respondents aged 18–29 years are overrepresented (p=0.0000), and respondents aged 40–49 years and 60+ are underrepresented (p=0.0311 and p=0.0000, respectively). The characteristics of the study population as well as tests for representativeness are presented in Appendix B.

Of the 1822 distributed questionnaires to GPs, 969 were returned, resulting in a response rate of 53 percent. In total, 491 GPs answered the questionnaire on behalf of themselves, while 478 answered on behalf of the patients. After removal of incomplete and inconsistent respondents, the final sample consists of 384 GPs answering on behalf of themselves and 391 GPs answering on behalf of the patients. An overview of incomplete and inconsistent respondents is provided in Appendix C. The sample of GPs is representative with respect to gender, age, and region, except that there are slightly more GPs in the sample from the Central Denmark Region (p=0.025) and slightly less from Region Zealand (p=0.023). The characteristics of the study population as well as tests for representativeness are presented in Appendix B. Randomization of the two samples of GPs is deemed successful since the samples are similar on all variables, except that in the GP sample there are slightly more GPs from the Capital Region of Denmark (p=0.022).

4.2. Regression results

Table 2 shows the results from the BW3 models (Model 4). For best choices, all attributes have the same signs across samples and are statistically significant (except for the health status attribute in the GP sample). Interestingly, the utility for follow-up visits is negative for all three samples, and is consistently ranked as least important. For worst choices signs are also similar across samples, except for the follow-up for patients, which is negative but insignificant. It should be noted that the larger standard errors and thereby confidence intervals in the GP samples are presumably due to the smaller sample sizes compared to the patient sample.

[TABLE 2]

H1: *Perfect agent hypothesis*

Based on the results from the LL test for equality in utility parameters we reject the null hypothesis of similar preference patterns for GPs and patients (statistical value = 120.30 vs. critical chi-square value of 24.996). Hence, H1 (that $U = B$) is rejected at the overall level, entailing that we cannot confirm that GPs are perfect agents for the patients.

To explore these findings further, Figure 1 shows the rank order of the attributes and the confidence intervals of the utility coefficients for best and worst choices for patients and GPs. It is seen that GPs' preferences for best consultations are markedly different to patients' preferences for best consultations. GPs value shared decision-making on treatment (treatment_best) and proper time to explain the problem (problem_best) significantly more than the other attributes. Patients, on the other hand, value discussions about their health status and lifestyle (health status_best) most, closely followed by time devoted to explaining the problem (problem_best) and making shared decisions on treatment (treatment_best). Our results demonstrate that GPs do not act as perfect agents for patients in all aspects of the consultation, corresponding with the rejection of H1 from the LL test.

Furthermore, we found discrepancies in preferences to be largest in choice of best consultations. In choices of worst consultations, rank orders seem to follow the same pattern across the two samples. Both GPs and patients find a consultation where they do not talk about health status and lifestyle (health status _worst), and where shared decision-making does not take place (treatment_worst) as the worst (although the latter is more pronounced in the GP sample). This suggests that patients and GPs share a common understanding of what characterizes a bad consultation (that is, what should not be ignored), but that preferences are more dispersed when it comes to defining a good consultation.

[FIGURE 1]

H2: *Complete information hypothesis*

Based on the test statistics from the LL test between patients, and GPs answering on behalf of patients, we reject the null hypothesis of equality in utility parameters between the two samples (statistical value = 77.78 vs. critical chi-square value of 24.996). Hence, H2 (that $B = B^*$) can be rejected at the overall level, indicating that asymmetric information exists and that GPs have imperfect information about patients' preferences.

Figure 2 shows that GPs manage to get patients' rank order right for all attributes except for the patients' most important attribute. Thus, our results seem to indicate that GPs predominantly know patients' preferences for consultations, but that there is evidence of asymmetric information on patients' most preferred aspect in the consultation; that is, talking about health status and lifestyle. The fact that GPs underestimate the value for patients of being asked about general health status is found to be the main reason for the overall rejection of H2.[12] Our results demonstrate that even if GPs are purely user oriented, taking patients' preferences into account (α = 0), they are not fully capable of acting as perfect agents due to uncertainty about preferences. Notably, this discrepancy is not observed in the assessment of preferences for worst consultation, where GPs are aware that the worst consultation for patients is one in which patients are not asked about health status and lifestyle.

[FIGURE 2]

H3: *User orientation hypothesis*

Given the results of the LL test, we reject the null hypothesis of overall equality in utility parameters between GPs and GPs answering on behalf of patients, implying that GPs' perceptions of patient preferences significantly differ from their own preferences (statistical value = 76.10 vs. critical chi-square value of 24.996). This entails that H3 (that $U = B^*$) is rejected and henceforth $\alpha > 0$. This indicates that GPs deviate from being a perfect agent for the patients (independently of any asymmetric information that might be present) as a response to other motivational factors such as financial incentives and public service motivation (captured in the model as increasing opportunity costs in effort invested in the individual patient).

Figure 3 shows that best choices especially differ with respect to time for explaining the problem (problem_best) and reaching a shared decision on treatment (treatment_best). These are reversed in rank

[12] This was tested by calculating MRS matrices for all possible combinations of attributes and testing for statistical significant differences in MRS across samples using t-tests.

order between the two samples, constituting the main explanation for the rejection of H3.[13] For worst choices, there are also some notable differences. This is especially pronounced for the shared decision-making attribute (treatment_worst), which GPs value more strongly than their patients, according to their perception.

[FIGURE 3]

5. Discussion

Our study demonstrates that BW3 can provide valuable information on GPs' and patients' preferences. The use of BW3 was chosen over traditional choice experiments to obtain more preference information. This was an important argument as we were constrained in the questionnaire length to the GPs. Furthermore additional insights into individual decision processes were obtained, as BW3 allows us to compare differences between best and worst choices. Having information on which characteristics are important when choosing the best consultation is naturally relevant. In our setting the information gathered from worst choices add some further insights on the utility functions of GPs and patients. This is valuable information in identifying attribute levels that make the good or service - in our case a consultation – (un)acceptable (Flynn and Marley 2012). BW3 thus provides health care decision makers with additional information about which characteristics that should *not* be ignored during a consultation. According to our results we see a slight discrepancy in the relative importance of best and worst consultations most notably for GPs, implying that best and worst preferences are not perfectly inversely correlated. This is most pronounced for the attribute 'health status and life style'. In contrast to worst choices, asking about health status is not deemed particularly important in best choices. Our results thus seem to suggest that GPs acknowledge that a conversation about health status and lifestyle is important and should ideally *not* be ignored. Nevertheless they still choose not to prioritize this as part of their preferred consultation. This discrepancy could be driven by time constraints (public service motivation), lack of financial incentives, paternalism, and/or a (mistaken) belief that patients consider this part of the consultation less relevant (asymmetric information). Our results suggest that the latter bears some explanatory power and that the GPs' agency role can be improved through ensuring greater information dissemination of patients' preferences to GPs. From a policy perspective, this finding is especially relevant in light of increased attention on the future role(s) of the GP in health prevention and health promotion.

[13] This was tested by calculating MRS matrices for all possible combinations of attributes and testing for statistical significant differences in MRS across samples using t-test.

According to our theoretical model, the observation that GPs do not prioritize time for explaining the problem and thus deviate from what they believe is in the individual patient's interest may be explained by considerations for the joint interest of other patients and/or the design of the current remuneration system. By shortening the consultation, there will be more time to see other patients and, at the same time, increase payment per hour. This suggests that GPs – at least in some situations – take the opportunity cost of time into account and actively decide not to meet the individual patient's wishes. According to our findings it thus seems that GPs take both principals' interests into consideration in their choice of consultation trying to optimize the use of their resources by avoiding long and inefficient consultations. However, our results could also be interpreted to support models of supplier-induced demand (SID) arguing that GPs use their superior information to deliberately (and at a cost of the individual patient) deviate from optimality (McGuire 2000). Unfortunately we are not able to distinguish between the different motivational factors included in α and therefore not whether GPs refrain from being perfect agents as to increase own earnings (as assumed in SID) or due to the consideration of other patients, just as we are not able to determine the size of α, never mind the optimal size of α. Although a few papers have addressed this topic (see e.g., Godager and Wiesen 2013; Pedersen et al. 2014), more research on this trade-off is warranted. Nonetheless, the fact that GPs seem to act as agents for both principals (i.e. the individual patient and the collective entity), thereby weighing their interests, is in itself an interesting finding which suggests that primary care providers are responsive to changes in these motivational factors. This implies that multi-pronged strategies targeting different motivational drivers of the GPs could be considered a relevant tool in areas where patient centered health care delivery is warranted.

In our theoretical model, it is the patient's utility - and not health gains more narrowly defined - that enters into the GP's utility function (Mooney and Ryan 1993). This approach was chosen as our experiment focuses on the communication between GP and patient in a consultation more generally. Furthermore, describing a patient's gain in terms of utility is in line with the welfarism approach in neoclassical economic theory (Brouwer et al. 2008), and accords with the European definition of the discipline of general practice[14]. Also, as noted by McKinstry (1992), paternalism is rarely justified when treating patients of sound mind. If GPs take a more paternalistic view and more narrowly consider a patient's health outcome, this could drive a wedge between elicited preferences of GPs and patients. In such a case the divergence in preferences should center on attributes with less clinical relevance. Our result that GPs find the attributes 'problem' and 'treatment' most important indicate that GPs mainly focus on aspects related to the specific

[14] According to the European definition of general practice, a characteristic of the discipline is that GPs should take a holistic view of the patient treating health problems in their physical, psychological, social, cultural and existential dimensions Hence part of the treatment in general practice also includes talking to, comforting and soothing patients.

health problem before concentrating on early detection and prevention of lifestyle diseases, whereas patients themselves focus on general health status and lifestyle before the particular health problem. If GPs consider ‘talking about health status and life style’ of low value in terms of improving patient’s health relative to the opportunity costs, then not prioritising this during the consultation could be based on a legitimate judgement. Furthermore it should be noted that our choice scenario was described as a choice between different types of consultations not involving direct treatment decisions. This allowed us to assume one-sided asymmetric information hereby eliminating asymmetric information on clinical knowledge. Unfortunately we are not able to empirically verify the potential extent to which some consultations were considered more appropriate than others from a clinical perspective (potentially causing non-alignment in preferences). Future studies that investigate whether this type of study design is appropriate in eliminating one side of the asymmetric information are warranted.

To examine whether our results differ across a range of subsamples of GPs and patients, we estimated models for (1) frequent visitors in general practice (patients having three or more consultations per year) to examine differences in preference patterns across patient groups[15]; (2) male vs. female GPs to test for gender differences in motivational factors and information asymmetry, and; (3) old vs. young GPs to see whether years of experience affect the results. Overall, we were not able to detect any differences in preference patterns according to the subgroup analyses. Our results thus seem to be robust within different subsamples of GPs and patients. Moreover, our conclusions are also robust within different model specifications (Model 1-4). That there are no differences in preferences between subgroups of GPs and patients indicates limited (observable) preference heterogeneity between persons. Although we use cluster robust standard errors in the estimation, we do not allow for unobservable preference heterogeneity between persons. To further test the robustness of our results we also ran models allowing for both within- and between-person heterogeneity (we specified these models as mixed logit with all model parameters being normally distributed and with a full covariance matrix being estimated between them, and used Bayesian estimation). Importantly, overall conclusions were not affected by method used. This is in line with previous studies reaching the same conclusions on main effects regardless of model (Lancsar et al. 2013, Louviere et al. 2015). Hence, for the purpose of this paper it was deemed sufficient to use models allowing for within-person heterogeneity (and heteroscedasticity) only.

The interviewed GPs and the Organisation of General Practitioners in Denmark recommended sending out questionnaires by regular mail as this is more convenient for GPs and in accordance with their usual work

[15] Patients with more visits will be overrepresented in general practice. Thus it is possible that the GPs when answering of behalf of the patients have based their decision on this group of patients.

procedures, and hence most likely to increase response rates. However, the paper based questionnaires allowed the GPs to proceed without completing all choice questions. A web-based survey format was chosen for the patient sample as this provides a cost-effective survey strategy for targeting the general population. In Denmark, 86 percent of the population has Internet access in their own homes (Statistics Denmark 2010), and coverage error is therefore not a major problem, although the prevalence of a panel effect cannot be excluded (Couper 2000). Empirical studies comparing results from paper-based and web-based surveys find no differences in SP results (Banzhaf et al. 2006; Olsen 2009). However, in our study, the data collection procedure caused a number of inconsistent choices, where the same consultation approach was chosen as both best and worst. In an electronic survey, the alternative chosen as best could be removed in the subsequent choice of worst alternative amongst the remaining two alternatives (hereby imposing consistency in all responses), whereas in paper-based surveys this is not possible. Since there is no research yet on how the removal of the best alternative affects preference formation, we chose to use the same survey design across data-collection procedures, albeit this implies permitting inconsistent responses (which were subsequently removed). Marley and Louviere (2005) noted that best–worst choice tasks seem easy for people to complete since they take advantage of people's propensity to identify and respond more consistently to extreme options (Lancsar et al. 2013). However, based on our study, we cannot explain why some respondents make inconsistent choices (choosing the same alternative as the best and worst). Although some research has peripherally engaged in this (see, e.g., Giergiczny et al. 2013), more research is warranted.

For our results to be valid in a policy setting, we need to assume that GPs' and patients' choice patterns are comparable to real-life behavior. Across fields, studies on the predictive value of choice experiments (e.g. Araña and León 2013; Cameron et al. 2002; Carlsson and Martinsson 2001; Mark and Swait 2004; Ryan and Watson 2009) have shown the capability of adequately predicting actual decision behavior. In a recent review, Lanscar and Swait (2014) noted that the external validity of choice experiments is generally good (but under-researched), with some exceptions (e.g. Fifer, Rose and Greaves 2014; Krucien et al, 2015). In this study we do not include a cost attribute and do not estimate absolute values such as willingness to pay. Instead, our focus is on identifying differences in choice patterns (rankings) across samples. We have no reason to believe that these observed differences should not reflect actual differences in preferences and beliefs. Furthermore, we need to assume that GPs understood the task and tried their best when answering on behalf of patients (thus disregarding own interests). As we observe differences across the two GP samples, while randomization was successful, we believe that this is the case.

Because patient and GP data could not be matched, it was not possible to identify which GPs serve which patients, and thus possible nonalignment of preferences may arise because of estimation at an aggregate and unmatched level. However, the 775 GPs included serve a fairly large fraction of approximately 20 percent of the Danish population. Moreover, our subsample analyses show that our results are robust to subsample estimation. Hence, we have no reason to assume that preferences would change had we drawn another random sample of GPs or patients.

In this paper, we have assumed that respondents make sequential choices for best and worst consultations. However, some persons could also have formed preferences for worst consultation before best, or decided on best and worst consultations simultaneously. We tested the decision rule of respondents by also performing analyses on sequential worst–best and simultaneous decisions and compared log likelihood estimates. For the majority of respondents, the sequential best–worst decision rule had the best fit. However, more research is needed on decision behavior in best–worst scaling experiments in the future (see also Giergiczny et al. 2013; Louviere et al. 2015; Rose 2014).

6. Conclusions and implications

Findings from this study extend the existing literature on the agency relationship by providing new insights into the understanding of the complex patient–doctor relationship and the potential underlying causes of nonalignment in preferences. We show that GPs do not act as perfect agents for patients in all aspects of the consultation. This is both due to uncertainty about patients' preferences (asymmetric information) in some aspects, and due to other motivations than user orientation in other aspects. This demonstrates that user orientation is not the sole (or perhaps even primary) motivational factor of the GPs, and that user orientation is likely to conflict with other motivations. Whether the observed deviations are caused by a concern for other patients (public service motivation), a personal financial interest, paternalism, or all of the above, remain unanswered. However, the finding that GPs seem to be responsive to different motivational factors highlights the need for considering other incentive mechanisms than just financial incentives in the endeavor of optimal delivery of primary health care services. Future studies that look further into these motivations and potential interactions between them are warranted. Finally we emphasize that GPs' lack of knowledge on patients' preferences is in itself problematic as it makes GPs unable to act in accordance with patients' preferences even when they intend to do so. This information asymmetry can however be minimized by increasing information dissemination to GPs. This would optimize the agency relationship and increase utility for both GPs and patients. Nevertheless, the degree to which

GPs will act accordingly will ultimately depend on the GPs' trade-offs between different motivational factors.

7. Acknowledgements

Thanks to Peter Zweifel, Ismo Linnosmaa, Dorte Gyrd-Hansen, Kim Rose Olsen, Ulrich Thy Jensen, Jakob Kragstrup, Thore Eriksen, Adam Moltke, and Jesper Lykkegaard for providing valuable comments and information during different phases of this study. Also thanks to participants at the Stated Preferences in Health Seminar, Utrecht 2015, and at the 36th Nordic Health Economics Study Group Meeting, Uppsala 2015, for their valuable input. The third author acknowledges financial support by the European Research Council through the consolidator grant 615596-DECISIONS. Finally, thanks to three anonymous referees and editor of Journal of Health Economics, Luigi Siciliani, for valuable comments.

8. Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

9. References

Andersen, L.B., T. Pallesen and L.H. Pedersen. 2011. 'Does Ownership Matter? Public Service Motivation among physiotherapists in the Private and Public Sectors in Denmark', *Review of Public Personnel Administration*, 31, 1, 10–27

Araña, J. E., & León, C. J. (2013). Dynamic hypothetical bias in discrete choice experiments: Evidence from measuring the impact of corporate social responsibility on consumers demand. *Ecological Economics*, *87*, 53-61

Ashenfelter, O. & Card, D. 2010. *Handbook Of Labor Economics, Vol 4A*, 4 ed. Elsevier.

Banzhaf, H.S., Burtraw, D., Evans, D., & Krupnick, A. 2006. Valuation of natural resource improvements in the Adirondacks. *Land Economics*, 82, (3) 445-464

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. The Review of Economic Studies, 70(3), 489-520

Blomqvist, A. 1991. The doctor as double agent: information asymmetry, health insurance, and medical care. *Journal ofHealth Economics*, 10, (4) 411-432

Brekke, K.R., Nuscheler, R., & Straume, O.R. 2007. Gatekeeping in health care. *Journal of Health Economics*, 26, (1) 149-170

Brouwer, W., Culyer, A., van Exel, N., & Rutten, F. 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics*, 27, (2) 325-338

Cameron, T. A., Poe, G. L., Ethier, R. G., & Schulze, W. D. (2002). Alternative non-market value-elicitation methods: are the underlying preferences the same?. *Journal of Environmental Economics and Management*, *44*(3), 391-425

Carlsson F, Martinsson P. Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. J Environ Econ Manag. 2001;41(2):179–92

Carlsen, B. & Aakvik, A. 2006. Patient involvement in clinical decision making: the effect of GP attitude on patient satisfaction. *Health Expectations*, 9, (2) 148-157

Cheraghi-Sohi, S., Hole, A.R., Mead, N., McDonald, R., Whalley, D., Bower, P., & Roland, M. 2008. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. *Annals of FamilyMedicine*, 6, (2) 107-115

ChoiceMetrics 2009, *Ngene 1.0 User manual & reference guide. The cutting edge in experimental design*.

Clark, M.D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, EW. 2014. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*, 32, (9) 883-902

Couper, M.P. 2000. Web surveys: A review of issues and approaches. *Public opinion quarterly*, 64, (4) 464-494

Cox, K., Britten, N., Hooper, R., & White, P. 2007. Patients' involvement in decisions about medicines: GPs' perceptions of their preferences. *British Journal of General Practice*, 57, (543) 777-784

Dijk, C.E., Berg, B., Verheij, R.A., Spreeuwenberg, P., Groenewegen, P.P., & Bakker, D.H. 2013. Moral Hazard And Supplier-Induced Demand: Empirical Evidence In General Practice. *Health Economics*, 22, (3) 340-352

Doran, E., Robertson, J., & Henry, D. 2005. Moral hazard and prescription medicine use in Australia�Çöthe patient perspective. *Social Science & Medicine*, 60, (7) 1437-1443

Dusheiko, M., Gravelle, H., Jacobs, R., & Smith, P. 2006. The effect of financial incentives on gatekeeping doctors: evidence from a natural experiment. *Journal of Health Economics*, 25, (3) 449-478

Ellis, R. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17, (5) 537-555

Ellis, R. & McGuire, T. 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5, (2) 129-151

Evans, R. 1984. *Strained mercy: the economics of Canadian health care* Toronto, Butterworths.

Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it?. *Transportation research part A: policy and practice*, *61*, 164-177

Flynn, T. 2010. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert review of pharmacoeconomics & outcomes research*, 10, (3) 259-267

Flynn, T., Louviere, J., Peters, T., & Coast, J. 2007. Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, 26, 171-189

Flynn, T. & Marley, A.J. 2012. 'Best Worst Scaling: Theory and Methods'. *CenSoC Working Paper Series No.12-002*

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. Journal of economic surveys, 15(5), 589-611

Gafni, A., Charles, C., & Whelan, T. 1998. The physician-patient encounter: the physician as a perfect agent for the patient versus the informed treatment decision-making model. *Social Science and Medicine*, 47, (3) 347-354

Giergiczny, M., Hess, S., Dekker, T., & Chintakayala, P. 2013. Testing the consistency (or lack thereof) between choices in best-worst surveys. *third international choice modelling conference, Sydney*

Godager, G. & Wiesen, D. 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32, (6) 1105-1116

Gosden, T., Forland, F., Kristiansen, I.S., Sutton, M., Leese, B., Giuffrida, A., Sergison, M., & Pedersen, L. 2000. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. *Cochrane Database Syst Rev*, 3, (3)

Greenfield, D. & Braithwaite, J. 2008. Health sector accreditation research: a systematic review. *International Journal for Quality in Health Care*, 20, (3) 172-183

Hauber, A. B., Mohamed, A. F., Johnson, F. R., & Falvey, H. (2009). Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents. *Diabetic medicine*, *26*(4), 416-424

Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, *30*(4), 637-646.

Hirth, R.A., Bloom, B.S., Chernew, M.E., & Fendrick, A.M. 2000. Patient, physician, and payer perceptions and misperceptions of willingness to pay for diagnostic certainty. *International journal of technology assessment in health care*, 16, (1) 35-49

Hjelmgren, J. & Anell, A. 2007. Population preferences and choice of primary care models: A discrete choice experiment in Sweden. *Health Policy*, 83, (2-3) 314-322

Krucien, N., Gafni, A., & Pelletier-Fleury, N. (2015). Empirical testing of the external validity of a discrete choice experiment to determine preferred treatment option: the case of sleep apnea. *Health economics*, *24*(8), 951-965

Jacobsen, C. B., Hvitved, J., & Andersen, L. B. (2014). Command and motivation: How the perception of external interventions relates to intrinsic motivation and public service motivation. *Public Administration*, *92*(4), 790-806.

Jensen, U. T. & Andersen, L. B. 2015. Public service motivation, user orientation, and prescription behaviour: Doing good for society or for the individual user? Public Administration, 93: 753–768

Jobe, J. 2003. Cognitive psychology and self-reports: Models and methods. *Quality of Life Research*, 12, (3) 219-227

Labelle, R., Stoddart, G., & Rice, T. 1994. A re-examination of the meaning and importance of supplier-induced demand. *Journal of Health Economics*, 13, (3) 347-368

Lancsar, E. & Louviere, J. 2008. Estimating individual level discrete choice models and welfare measures using best worst choice experiments and sequential best worst MNL. *University of Technology, Centre for the Study of Choice (Censoc)*

Lancsar, E. & Swait, J. 2014. Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics*, *32*(10), 951-965

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., & Burgess, L. 2013. Best worst discrete choice experiments in health: Methods and an application. *Social Science & Medicine*, 76, 74-82

Lancsar, E., Louviere, J., & Flynn, T. 2005. Several methods to investigate relative attribute impact in stated preference experiments. *Newcastle upon Tyne, UK: Health Economics Study Group Meeting*

Lancsar, E., Louviere, J., & Flynn, T. 2007. Several methods to investigate relative attribute impact in stated preference experiments. *Social Science and Medicine*, 64, 1738-1753

Louviere, J., Flynn, T., & Marley, A. 2015. *Best-Worst Scaling. Theory, Methods and Applications* Cambridge, UK, Cambridge University Press

Louviere, J. & Lancsar, E. 2009. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy and Law*, 4, (4) 527-546

Louviere, J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. 2004. Modeling the choices of individual decision-makers by combining efficient choice experiments with extra preference information. *Sydney, CenSoC Working Paper N.04-005*

Louviere, J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. 2008. Modeling the choices of individual decision-makers by combining efficient choice experiments with extra preference information. *Journal of Choice Modelling*, 1, (1) 128-163

Mark, T.L. & Swait, J. 2004. Using stated preference and revealed preference modeling to evaluate prescribing decisions. *Health Economics*, 13, (6) 563-573

Marley, A.A.J. & Louviere, J.J. 2005. Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, 49, (6) 464-480

Marley, A.A.J. & Pihlens, D. 2012. Models of best-worst choice and ranking among multiattribute options (profiles). *Journal of Mathematical Psychology*, 56, (1) 24-34

Marshall, D., Johnson, F., Kulin, N., Õzdemir, S., Walsh, J., Marshall, J., Bebber, S., & Phillips, K. 2009. How do physician assessments of patient preferences for colorectal cancer screening tests differ from actual preferences? A comparison in Canada and The United States using a stated-choice survey. *Health Economics*, 18, (12) 1420-1439

McClellan, M. 1996. "Medicare Reimbursement and Hospital Cost Growth," *In The economics of population ageing*, J. Stephenson & G. Scobie, eds., New Zealand Treasury Wellington: University of Chicago Press, pp. 149-190

McGuire, T. 2000, "Physician agency," *In Handbook of Health Economics Volume 1B*, A. Culyer & J. Newhouse, eds., Elsevier, pp. 461-536

McKinstry, B. (1992). Paternalism and the doctor-patient relationship in general practice. *Br J Gen Pract*, *42*(361), 340-342

Mooney, G. & Ryan, M. 1993. Agency in health care: getting beyond first principles. *Journal of Health Economics*, 12, (2) 125-135

Mühlbacher, A.C. & Nübling, M. 2010. Analysis of physicians' perspectives versus patients' preferences: direct assessment and discrete choice experiments in the therapy of multiple myeloma. *The European Journal of Health Economics*, 12, 193-203

Neuman, E. & Neuman, S. 2009. Agency in health-care: are medical care-givers perfect agents? *Applied Economics Letters*, 16, (13) 1355-1360

Olsen, S.B. 2009. Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and resource economics*, 44, (4) 591-610

Payne, K., Fargher, E.A., Roberts, S.A., Tricker, K., Elliott, R.A., Ratcliffe, J., & Newman, W.G. 2011. Valuing pharmacogenetic testing services: A comparison of patients' and health care professionals' preferences. *Value in Health*, 14, 121-134

Pedersen, L.B., Kjær, T., Kragstrup, J., & Gyrd-Hansen, D. 2012. Do general practitioners know patients' preferences? An empirical survey on the agency relationship at an aggregate level using a discrete choice experiment. *Value in Health*, 15, 514-523

Pedersen, L.B., Riise, J., Hole, A.R., & Gyrd-Hansen, D. 2014. GPs' shifting agencies in choice of treatment. *Applied Economics*, 46, (7) 750-761

Rose, J. M. 2014. Interpreting discrete choice models based on best-worst data: a matter of framing, *Transportation Research Board 93rd Annual Meeting 14-5072 edn.*

Ryan, M. 1994. Agency in health care: lessons for economists from sociologists. *American Journal of Economics and Sociology*, 53, (2) 207-217

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67

Ryan, M. & Watson, V. 2009. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health Economics*, 18, (4) 389-401

Scarpa, R., Notaro, S., Louviere, J., & Raffaelli, R. 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American journal of agricultural economics*, 93, (3) 813-828

Scott, A. 2000, "Economics of general practice," *In Handbook of health economics*, 1 ed. A. Culyer & J. Newhouse, eds., Oxford: Elsevier, pp. 1175-1200.

Scott, A. & Shiell, A. 1997. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics*, 16, 323-342

Scott, A. & Vick, S. 1999. Patients, Doctors and Contracts: An Application of Principal-Agent Theory to the Doctor-Patient Relationship. *Scottish Journal of Political Economy*, 46, (2) 111-134

Statistics Denmark. Internetadgang i hjemmet (Internet access at home). 2010. http://www.statistikbanken.dk/BEBRIT01, April 5, 2011.

Statistics Denmark. SYGP: Number of persons with contacts covered by public health insurance by region, type of benefits, age, sex and socioeconomic status. 2011. http://www.statistikbanken.dk/statbank5a/default.asp?w=1280, May 9, 2011.

Steijn, B. (2008). Person-environment fit and public service motivation. *International public management journal*, *11*(1), 13-27

Street, D. & Burgess, L. 2007. *The Construction of Optimal Stated Choice Experiments. Theory and Methods* John Wiley & Sons.

Swait, J. & Louviere, J. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30, (3) 305-314

Train, K. 2003. *Discrete Choice Methods with Simulation* Cambridge University Press.

Vandenabeele, W. 2008. Development of a Public Service Motivation Measurement Scale: Corroborating and Extending Perry's Measurement Instrument, *International Public Management Journal*, 11, 1, 143–67
van den Berg, M.J., de Bakker, D.H., Westert, G.P., van der Zee, J., & Groenewegen, P.P. 2009. Do list size and remuneration affect GPs' decisions about how they provide consultations? *BMC Health Services Research*, 9, (1) 39

Van den Hombergh, P., Engels, Y., van den Hoogen, H., Van Doremalen, J., van den Bosch, W., & Grol, R. 2005. Saying 'goodbye' to single-handed practices; what do patients and staff lose or gain? *Family Practice*, 22, (1) 20-27

Vedsted, P., Mainz, J., Lauritzen, T., & Olesen, F. 2002. Patient and GP agreement on aspects of general practice care. *Family Practice*, 19, (4) 339-343

Wijnen, B. F., van der Putten, I. M., Groothuis, S., de Kinderen, R. J., Noben, C. Y., Paulus, A. T., ... & Hiligsmann, M. (2015). Discrete-choice experiments versus rating scale exercises to evaluate the importance of attributes. *Expert review of pharmacoeconomics & outcomes research*, *15*(4), 721-728.

Yoo, H. & Doiron, D. 2013. The use of alternative preference elicitation methods in complex discrete choice experiments. *Journal of Health Economics*, 32, (6) 1166-1179

Table 1. Attributes and levels for patients and GPs

Attributes	Levels (patients)[a]	Levels (GPs)[a]
Problem	(1) The GP asks me to explain my problem in detail. I am given plenty of time to describe all symptoms and the discomforts I suffer in everyday life (0) The GP asks me to briefly explain what the problem is	(1) I ask the patient to explain his/hers problem in detail. The patient gets plenty of time to describe all symptoms and the discomforts he/she suffers in everyday life (0) I ask the patient to briefly explain what the problem is
Treatment	(1) The GP discusses with me what the treatment options are, and which one would suit me best (0) The GP tells me what treatment I need	(1) I discuss with the patient what the treatment options are, and which one would suit him/her best (0) I tell the patient what treatment he/she needs
Follow-up	(1) We set a new date for me to come back (0) The GP tells me to contact him/her if the treatment doesn't work	(1) We set a new date for the patient to come back (0) I tell the patient to contact me if the treatment doesn't work
Care	(1) The GP asks me whether there are other things that I want to talk about (0)	(1) I ask the patient whether there are other things that he/she wants to talk about (0)
Health status	(1) The GP asks about my general health status and lifestyle (0)	(1) I ask about the patient's general health status and lifestyle (0)

[a] Coding in parentheses.

Table 2. Comparison between samples: Stage specific best-worst model

	Patients			GPs on behalf of patients			GPs		
	Coeff.	Robust std. error		Coeff.	Robust std. error		Coeff.	Robust std. error	
Problem_best	0.465	0.032	***	0.869	0.070	***	0.332	0.064	***
Treatment_best	0.411	0.039	***	0.641	0.070	***	0.634	0.066	***
Follow-up_best	-0.176	0.039	***	-0.145	0.070	**	-0.255	0.065	***
Care_best	0.288	0.023	***	0.486	0.048	***	0.091	0.048	**
Health status_best	0.598	0.042	***	0.365	0.082	***	0.031	0.083	
ASC A_best	-0.539	0.029	***	-0.469	0.060	***	-0.308	0.052	***
ASC B_best	-0.370	0.028	***	-0.320	0.054	***	-0.207	0.049	***
Problem_worst	0.292	0.043	***	0.411	0.085	***	0.180	0.072	**
Treatment_worst	0.455	0.046	***	0.378	0.081	***	0.490	0.073	***
Follow-up_worst	-0.074	0.045		0.074	0.079		0.043	0.072	
Care_worst	0.380	0.041	***	0.315	0.090	***	0.086	0.072	
Health status_worst	0.999	0.047	***	0.865	0.088	***	0.515	0.078	***
ASC A_worst	0.845	0.049	***	0.813	0.095	***	0.823	0.085	***
ASC B_worst	0.723	0.058	***	0.595	0.106	***	0.541	0.099	***
LL(0)	-9883.35			-2802.31			-2752.14		
LL(Model)	-8642.70			-2364.31			-2515.25		
Adjusted R^2	0.1241			0.1513			0.0810		
n (choices)	5516			1564			1536		
N (respondents)	1379			391			384		

*** Statistically significant at the 0.001 level, ** Statistically significant at the 0.05 level

Figure 1. H1: Perfect agent hypothesis

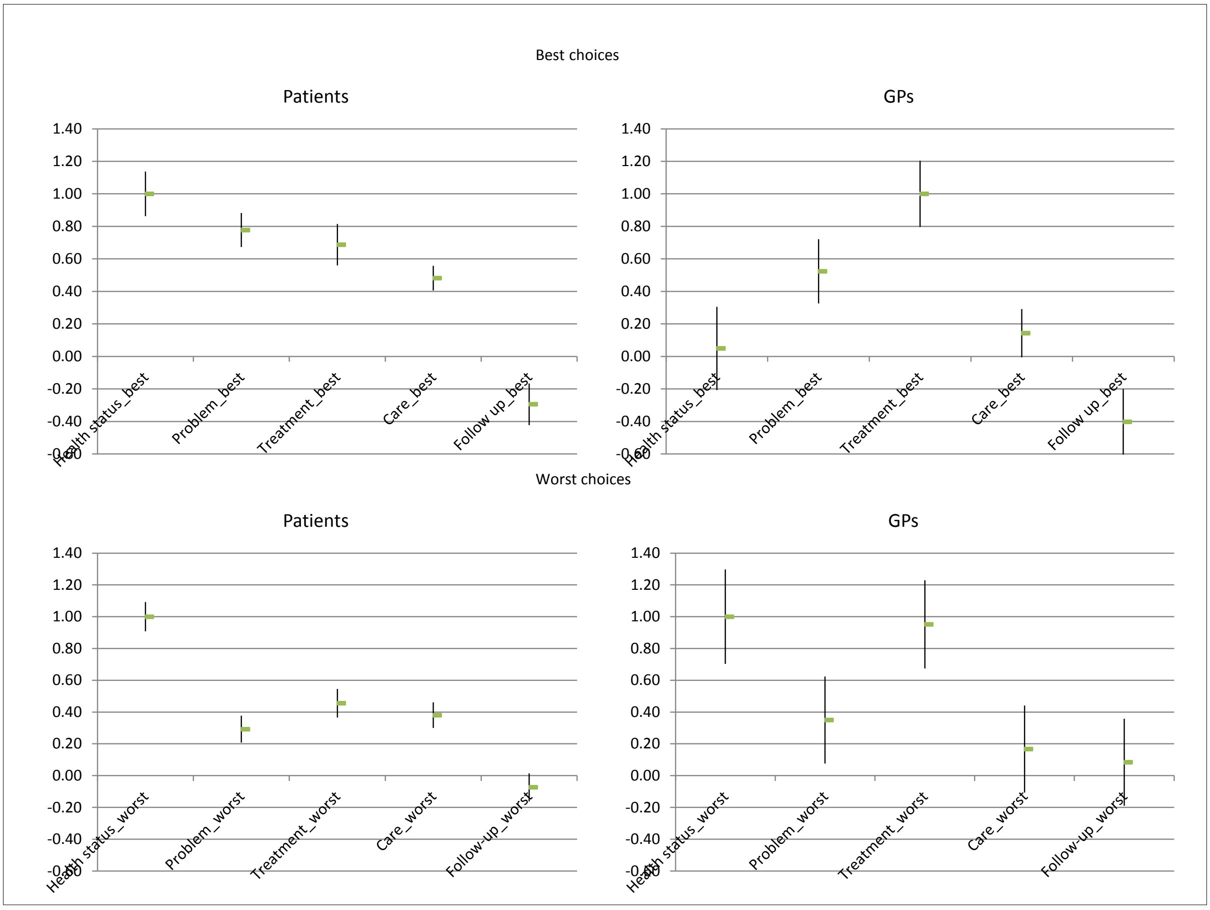


Figure 2. H2: Complete information hypothesis

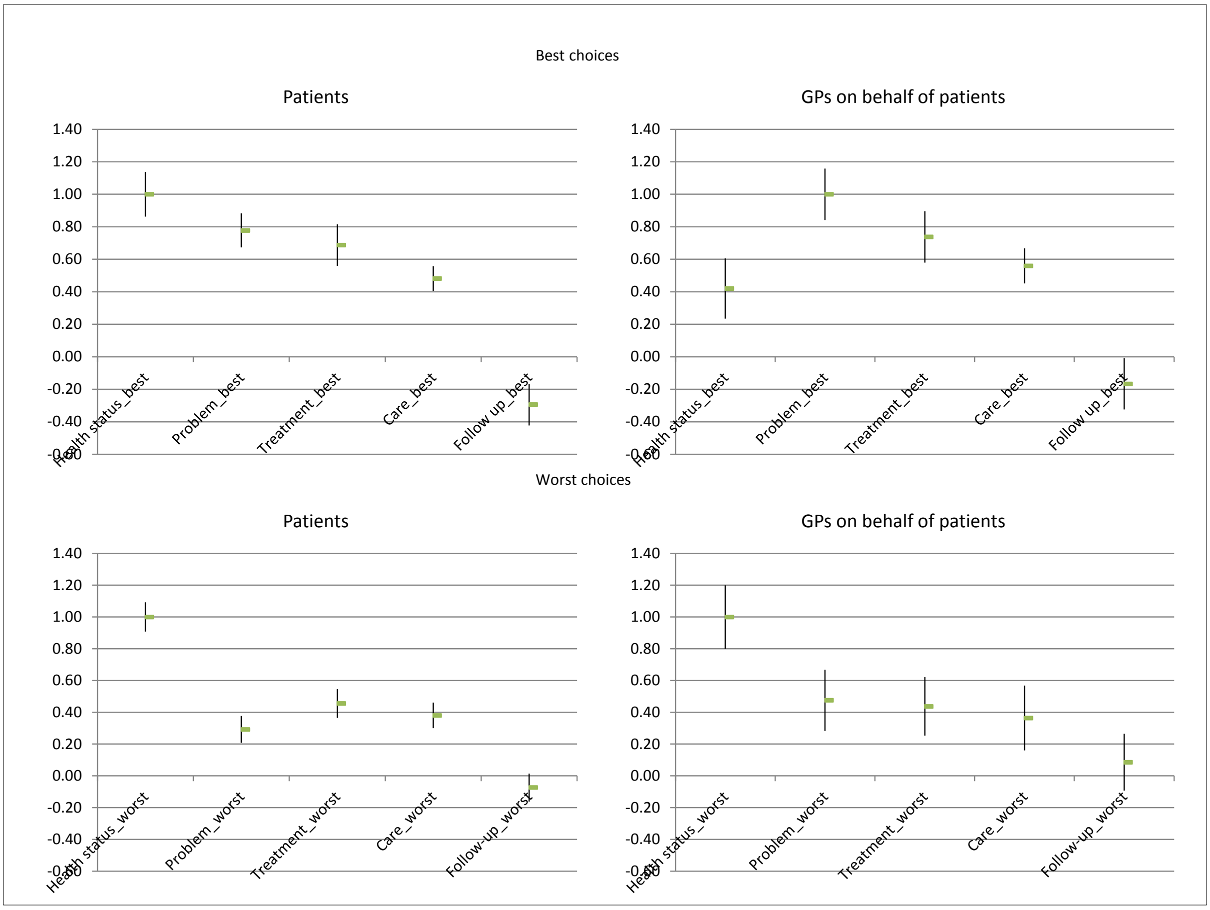
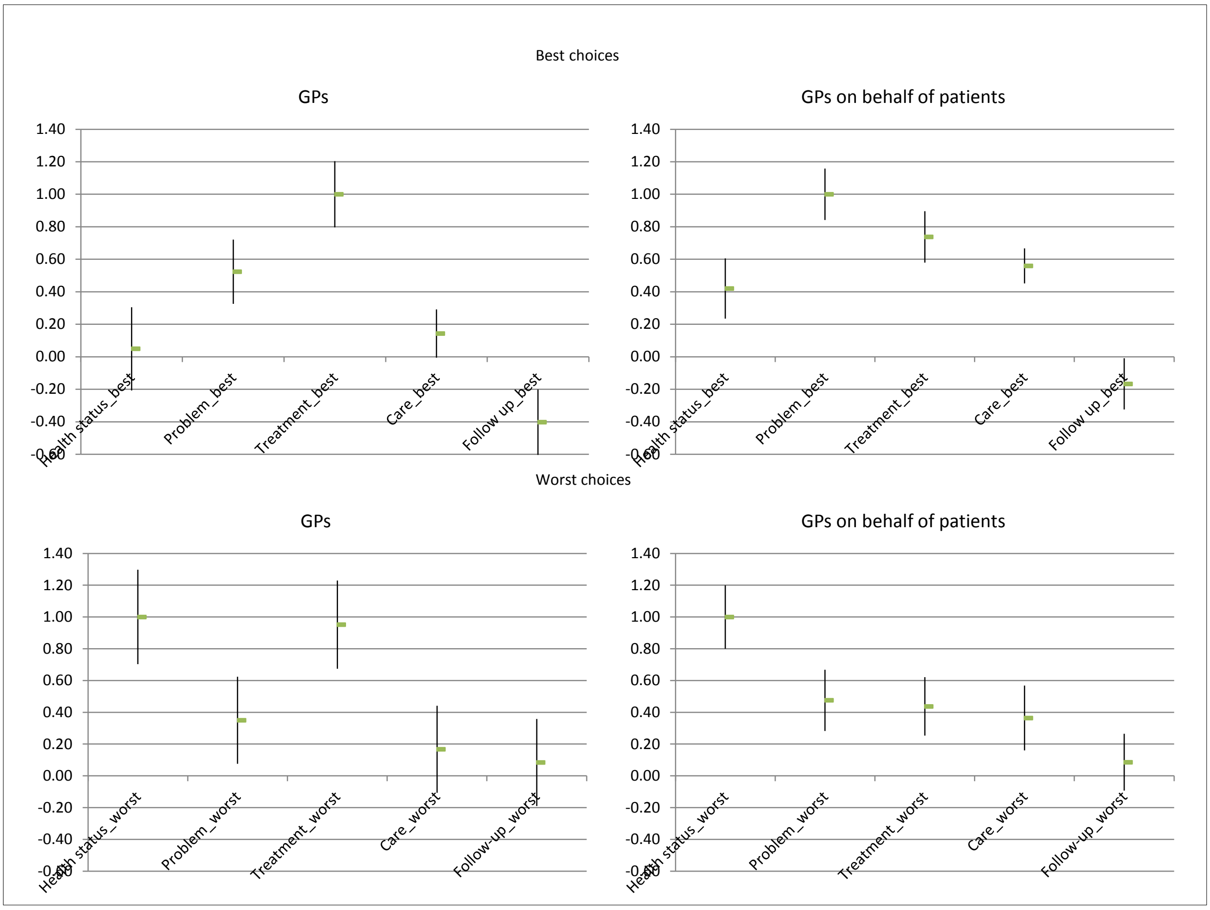


Figure 3. H3: User orientation hypothesis



10. Appendix A

Patient version:

Imagine that you have an appointment with your GP due to a sore throat and headache. You have had pain for nearly 14 days and are not getting better. You have been very tired and unwell but haven't been off work sick/ill in bed.

Which consultation would you, in the situation described, be most satisfied with, and which consultation would you be least satisfied with?

Consultation 1	Consultation 2	Consultation 3
The GP asks me to briefly explain what the problem is	The GP asks me to briefly explain what the problem is	The GP asks me to explain my problem in detail. I am given plenty of time to describe all symptoms and the discomforts I suffer in everyday life.
The GP discusses with me what the treatment options are, and which one would suit me best	The GP discusses with me what the treatment options are, and which one would suit me best	The GP tells me what treatment I need
We set a new date for me to come back	We set a new date for me to come back	The GP tells me to contact him/her if the treatment doesn't work
The GP asks me whether there are other things that I want to talk about	The GP asks me whether there are other things that I want to talk about	
	The GP asks about my general health status and lifestyle	

a. Which of the three consultations would satisfy you **most**? (Tick one box).

Consultation 1	Consultation 2	Consultation 3
☐	☐	☐

b. Which of the three consultations would satisfy you **least**? (Tick one box).

Consultation 1	Consultation 2	Consultation 3
☐	☐	☐

GP version:

Imagine that one of your patients has a non-acute appointment for a consultation at your clinic due to a sore throat and headache. The patient has had pain for nearly 14 days and is not getting better. The patient has been very tired and unwell but hasn't been off work sick/ill in bed.

GPs on behalf of patient: Which of the three consultations do you think that the patient, in the situation described, would be most satisfied with, and which consultation would the patient be least satisfied with?

GPs: Which of the three consultations would you, in the described situation, be most satisfied with, and which consultation would you be least satisfied with?

Consultation 1	Consultation 2	Consultation 3
I ask the patient to briefly explain what the problem is	I ask the patient to briefly explain what the problem is	I ask the patient to explain his/her problem in detail. The patient gets plenty of time to describe all symptoms and the discomforts he/she suffers in everyday life.
I discuss with the patient what the treatment options are, and which one would suit him/her best	I discuss with the patient what the treatment options are, and which one would suit him/her best	I tell the patient what treatment he/she needs
We set a new date for the patient to come back	We set a new date for the patient to come back	I tell the patient to contact me if the treatment doesn't work
I ask the patient whether there are other things that he/she wants to talk about	I ask the patient whether there are other things that he/she wants to talk about	
	I ask about the patient's general health status and lifestyle	

a. GPs on behalf of patient: Which of the three consultations do you think would satisfy the patient **most**?
GPs: Which of the three consultations would satisfy you **most**? (Tick one box).

Consultation 1	Consultation 2	Consultation 3
☐	☐	☐

b. GPs on behalf of patient: Which of the three consultations do you think would satisfy the patient **least**?
GPs: Which of the three consultations would satisfy you **least**? (Tick one box).

Consultation 1	Consultation 2	Consultation 3
☐	☐	☐

11. Appendix B

Table 3. Characteristics of the study populations and tests for representativeness

	Patients		Danish population		p-value
	n	%	n	%	
Gender					
Male	669	48.51%	2748185	49.57%	0.4317
Female	710	51.49%	2795634	50.43%	0.4317
Geography					
Capital Region of Denmark	398	28.88%	1687071	30.43%	0.2047
Region Zealand	219	15.89%	820609	14.80%	0.2685
Region of Southern Denmark	302	21.92%	1200841	21.66%	0.8191
Central Denmark Region	305	22.13%	1255876	22.65%	0.6421
Region of Northern Denmark	154	11.18%	579422	10.45%	0.3939
Age					
18-29	437	31.85%	777655	17.96%	0.0000
30-39	236	17.20%	737011	17.02%	0.8566
40-49	228	16.62%	813591	18.79%	0.0311
50-59	220	16.03%	714595	16.50%	0.6394
60+	251	18.29%	1288190	29.74%	0.0000
	GPs		**GP population**		**p-value**
	n	%	n	%	
Gender					
Male	458	59.40%	2163	59.90%	0.7987
Female	313	40.60%	1448	40.10%	0.7987
Geography					
Capital Region of Denmark	231	29.96%	1094	30.30%	0.8538
Region Zealand	88	11.41%	518	14.35%	0.0227
Region of Southern Denmark	180	23.35%	810	22.43%	0.5850
Central Denmark Region	209	27.11%	837	23.18%	0.0248
Region of Northern Denmark	63	8.17%	352	9.75%	0.1532
Age					
-34	1	0.13%	3	0.08%	0.7329
35-39	34	4.44%	177	4.90%	0.5755
40-44	81	10.57%	390	10.80%	0.8538
45-49	114	14.88%	517	14.32%	0.6891
50-54	139	18.15%	580	16.06%	0.1708
55-59	180	23.50%	827	22.90%	0.7234
60-64	159	20.76%	773	21.41%	0.6880
65-69	55	7.18%	311	8.61%	0.1700
70+	3	0.39%	33	0.91%	0.0584

12. Appendic C

Table 4. Incomplete and inconsistent respondents

	Patients		GPs on behalf		GPs	
	n	%	n	%	n	%
Sample	1435		478		491	
Only incomplete	0	0.00%	33	6.90%	45	9.16%
Only inconsistent	56	3.90%	5	1.05%	9	1.83%
Both incomplete and inconsistent	0	0.00%	49	10.25%	53	10.79%
Final sample	1379		391		384	