What chronic ill patients want from the general practitioners: A stated-preference approach of the Chronic Care Model

Context

Worldwide there is an increasing burden of chronic disease in healthcare (WHO, 2005). In France, the National Health Insurance Fund for Salaried Workers (*Caisse Nationale d'Assurance Maladie des Travailleurs Salariés, CNAMTS*), which covers almost 50 million people, registered 9 million people with long-term disease in 2004 against 3.7 in 1994 (Païta & Weil, 2009) and this number could rise to 12 million in 2015. The increase of chronic diseases reinforces the issue of priority setting in healthcare in the context of resources scarcity. In 2007, in France, the patients identified as suffering from long-term disease by the CNAMTS represented 65% of the reimbursements i.e.80 billion Euros (Vallier et al, 2006, 2007). In the USA the rapid increase in the proportion of chronic patients is an even greater challenge. In 2005 nearly 133 million people had a chronic disease and in 2030 the estimated chronically ill population will be of 171 million (Wu & Green, 2000). Chronic diseases cause about 70% of deaths in the USA and are estimated to represent about 75% of health care costs each year. Ninety percent of seniors have at least one chronic disease, and 77% have two or more chronic conditions (Anderson, 2004).

In response to this change in the healthcare demand several initiatives have emerged since the early 1990s in the form of disease management and case management programs. These programs are based on three main principles (Bras et al, 2006). First, it is to segment the population of chronically ill patients according to their level of clinical risk. Then the management programs are partially (or by segment) tailored to the needs of the patients. Second, to increase the participation of allied health professionals (e.g. nurse) which in some circumstances are able to substitute or complement the

physicians. The third principle is to bring the patient to actively participate in the management of its chronic health problems. Thus these initiatives are essentially focused on and supported by the healthcare demand. Later on, this has been completed by changes at the healthcare supply level in the form of an integrated care model. The most famous initiative of this kind is the Chronic Care Model (CCM) (Wagner et al, 1996, 2001, Bodenheimer et al, 2002) and more recently the Guided Care Model (GCM) (Boult et al, 2008, 2011). As before, these models aim on the one hand to promote the patient participation (e.g. self-management, shared decision making) and on theother hand to develop a proactive care (e.g. medical cooperation, increasing use of computer records). Today these initiatives are used as reference framework for designing overall response to the increasing burden of chronic diseases at the health system level. In France, the CNAMTS has recently implemented an experimental program of disease management (called SOPHIA) focused on 143 000 type 1 & 2 diabetic patients. In United Kingdom, many initiatives have been implemented since the early 2000s to improve health care delivery towards chronically ill patients (e.g. Quality Outcome Framework, Expert Patient Program, and Community Matron).

These relatively new initiatives lead to significant changes in the patient's role in the healthcare and in the doctor-patient relationship. In the international literature some studies show that patients can be reluctant to some of these changes. For example older patients are less likely to engage actively in the medical decision making (Vick & Scott, 1998). Many studies are devoted to patients' preferences for organizational aspects of the general practice (GP) care (Cheraghi-Sohi et al, 2008, Hjelmgren & Anell, 2007, Porteous et al, 2006, Rubin et al, 2006, Gerard & Lattimer, 2005, Ryan & Watson, 2006, Gerard et al, 2006), but few analyze the patients preferences for these "improvements" of the GP care. Longo et al (2006) and Davison et al (2010) study patient preferences for aspects of the doctor-patient relationship relative to shared decision making and information exchange.

In this study, we analyze the patients' preferences for attributes of the CCM directly related to the patients' role in the GP care. To do this we use two stated preferences method that are the Best-Worst Scaling (BWS) method and the Discrete Choice Experiment (DCE) method. Our objective is twofold. First to identify the recommendations of the CCM that are the most likely to encourage chronically ill patients to accept changes in the current GP care supply. We assume that all the CCM attributes are of importance for chronically ill patients (hypothesis n°1) and that these attributes positively influence the utility gained from the general practice care (hypothesis n°2). Second, it is to compare from a methodological point of view the results of both stated preferences methods. We assume that BWS and DCE methods are both stated preferences methods that lead to similar results up to a scale (hypothesis n°3).

The rest of this paper is organized in three parts. First we present the BWS and DCE methods with an emphasis on this latter due to its relatively innovative nature in health studies. We also present the sampling strategy and data collection process. Second we describe the sample of respondents, the results of each method and we finally compare these results. Third we discuss our findings.

Method

1/ Best-Worst Scaling (BWS) method

1.1/ Theoretical presentation

This method has been developed by Louviere & Woodworth (1990) and used for the first time in 1992 (Finn & Louviere, 1992). The "scaling" refers to the process of assigning numerical values that reflect position of item on a scale. The core principle of this method is to select at least two "items" among a set (or subset) of 3 or more "items". An "item" could be defined by the level of a feature, a feature itself, or a collection of levels of features (i.e. a product). Each of these definitions is a variant of the BWS method:

- **Case 1** (i.e. object case) (table 1). People have to choose the best and worst items among subsets of features (or attributes).

Table 1. Case 1 BWS task example.		
GPs' attributes	**Best**	**Worst**
Cost	X	
Waiting time		X
Amiability		
Geographical proximity		

- **Case 2** (i.e. profile case) (table 2). People have to choose the best and worst items' levels among an option.

Table 2. Case 2 BWS task example.		
GPs' attributes levels	**Best**	**Worst**
Cost: **20 euros**	X	
Waiting time: **15 min**		
Amiability: **Yes**		
Geographical proximity: **5 Km**		X

- **Case 3** (i.e. multi-profile case) (table 3). People have to choose the best and worst options/products among a subset of options/products.

GPs' profiles	Best	Worst
Table 3. Case 3 BWS task example.		
Cost: **20 euros**; Waiting time: **15 min**; Amiability: **Yes** Geographical proximity: **5 Km**		X
Cost: **15 euros**; Waiting time: **15 min**; Amiability: **No;** Geographical proximity: **10 Km**	X	
Cost: **10 euros**; Waiting time: **10 min**; Amiability: **Yes** Geographical proximity: **5 Km**		

As originally conceived the BWS tasks require individuals to select the two farthest items on a latent scale which describes a dimension of interest (e.g. importance, utility). These items are then refereed as "best" and "worst" items. This approach of the human choice within a set of items is one among several others (e.g. to choose the best, then the 2nd best and so on).

The psychological model underpinning this choice process can be described in three steps:
- First people "list" all the best-worst (BW) and worst-best (WB) couples (or pairs).
- Second they "value" the magnitude of the distance between the two items on the latent scale of interest (e.g. utility differential).
- Third they choose the pair which maximizes the gap on the latent scale (to select the farthest items on the scale).

At this stage an important assumption is that people choose both the best and worst items simultaneously and independently. Then the probabilities of choosing the best and worst items among the common set of items are independent.

Moreover the "true" utility that people obtain from the selected pair is not fully recoverable. Some information could not be accessed by the researcher and/or people can choose with some uncertainty or inconsistency. This suggests the use of an error component in the modeling of the individuals BW/WB pairs' choices.

The components of the observable difference are the values of the best and worst items (also called "level scale values").

Due to the stochastic nature of the choice process, the individual decision is modeled as a probability. Therefore an additional assumption about the nature of the error/stochastic component is necessary to determine the model of probability. In practice the errors are generally supposed to

be independently and identically (i.e. iiD) distributed as an extreme value of type 1 (i.e. EV1) with a zero mean and a constant variance ($\frac{\pi^2}{6}$). According to this specification the probability to choose a particular pair of items can be predicted with a conditional logistic model (i.e. CLM) (McFadden, 1974). The proportional distance between two items on a latent scale represents the relative choice probability of that pair of items. The larger the distance between the two items, the higher is the probability to select this particular pair.

Among a subset of K=3 items $\{A; B; C\}$, there are K(K-1)=6 possible pairs: $\{AB; AC; BC; BA; CA; CB\}$. The observed choice is the pair "AB".

$$Gap_{ab} = V_{ab} + \varepsilon_{ab}$$

$$V_{ab} = L_a - L_b$$

$$\varepsilon_{ab} \longrightarrow iid\ EV1\ (0, \sigma^2)$$

$$P(pair = AB) = \frac{\exp(V_{AB})}{\exp(V_{AB}) + \cdots + \exp(V_{CB})}$$

Where:
- (Gap_{ab}) represents the "true" difference between the best item (a) and the worst item (b) on the latent scale
- (V_{ab}) represents the systematic/observable part of the "true" difference (a-b)
- (ε_{ab}) represents the stochastic/unobservable part of the "true" difference (a-b)
- (L_a) indicates the value of the best item (a) on the latent scale
- (L_b) indicates the value of the worst item (b) on the latent scale

One of the position parameter (β_K) is omitted from the estimation to avoid saturated model. This omitted item constitutes a reference point for the other estimates and then it can be interpreted as an "origin point" of the latent scale on which the positions of (K-1) items are estimated. Then the BWS method allows estimating the relative position of several items on a common latent scale. According to the ratio scale properties, the BWS estimates of the items can be directly compared and the gaps can be quantified and interpreted.

As in most stated preferences studies, respondents generally provide several observations. Then the data can be correlated within individuals. In practice to take into account this possibility an additional error component is added to the choice model in specifying a random intercept.

$$Lower\ level: U_i = \beta_{0i} + \beta_1.X + \varepsilon_i$$

$$Upper\ level: \beta_{0i} = \beta_{00} + \delta_{0i}$$

$$Multilevel\ model: U_i = \beta_{00} + \beta_1.X + \delta_{0i} + \varepsilon_i$$

In BWS model the intercept has no particular interest. This parameter has no clear theoretical foundations and therefore it is constrained to be zero ($\beta_{00} = 0$). So the final model to estimate is for an individual (i) and a task (t):

$$Gap_{(it)ab} = \beta_a.L_a - \beta_b.L_b + \delta_{0i} + \varepsilon_{(it)ab}$$

The BWS data can also be analyzed via "counts" or "frequencies". According to Marley & Louviere (2005) the simple score of the BW pairs is a closed approximation of the conditional logit estimates. However this relationship is true when people tend to consistently respond. If people are perfectly able to discriminate among items and to rank them in terms of dimension of interest (e.g. importance, utility), then the best and worst frequencies of the items are inversely related with one constant (C) (graphic 1) :

$$Frq(Best) = \frac{C}{Frq(Worst)} => R^2 = 1$$

Graph. Worst vs Best frequencies

Constant = 100

Worst = 100(Best)$^{-1}$

R^2 = 1
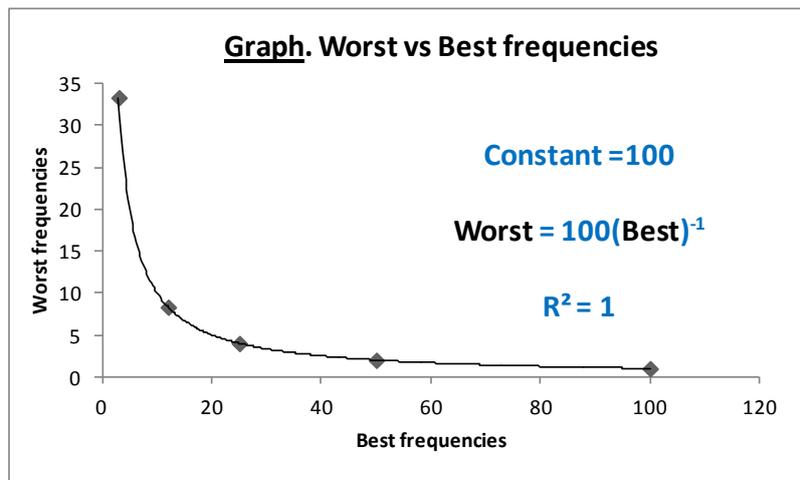
If this relationship is verified then the square root of the BW ratio is a ratio scale of the best frequencies and can be used as a substitute to the logit estimates. In practice previous empirical verifications (Flynn, Louviere & Peters, 2008) showed that these different estimates are very close.

In the count analysis, "Simple score" (= Best-Worst) posits the items on a interval scale and "Relative score" (= $\sqrt{\frac{Best}{Worst}}$) posits the items on a ratio scale.

The BWS method has many comparative advantages over rating, ranking and choice methods. Relative to rating task (e.g. scoring on an X-point Likert scale), the BWS method avoids the "scalar inequivalence problem" (Cohen & Neira, 2003) that occurs when people use rating scale in a

heterogeneous way. Also respondents are constrained to discriminate among items, unlike rating tasks they do not have the ability to specify multiple items as being of utmost importance. Relative to ranking tasks, the BWS method provides cardinal information about the items, it allows to rank the items (as with ranking task) and to interpret the differences between the rows (unlike ranking task). Relative to choosing tasks (e.g. DCE method), the BWS is supposed to be easier for respondents (or less cognitively demanding) because it requires to choose the ends of the utility space. The BWS method provides more information at the individual level. According to the exact nature of the BWS task it allows either a "partial" preference ordering (PPO) (e.g. A ≻ B, C ≻ D) or a "full preference ordering (FPO) (e.g. A ≻ B ≻ C ≻ D). Then this method allows individual-level modeling without questionable assumptions on the distribution of preferences.

1.2/ Use of the method

In this study we analyze the patients' preferences for attributes of the CCM directly related to the patients' role in the GP care (e.g. self-management, shared decision making). According to standard practice in the international stated preferences literature, the pool of candidate attributes have been identified through a literature review and then adjusted in a sequential piloting study with 10 chronically ill patients. This process has reduced the number of attributes from 25, mainly from the Patient Assessment of Care for Chronic Conditions (PACIC) questionnaire (Glasgow et al, 2005), to 10 (table 4).

Table 4. List of selected attributes		
Abbreviation	**Attribute**	**Detail**
SDM	Shared decision making	The GP asks for my ideas when we made a treatment plan and he gives choices about treatment to think about
INF	Informational continuity	The GP contacts me after visit to see how things were going
VRF	Regular care	The GP verifies regularly that my treatment plan is appropriated
CAL	Planned care	The GP gives to me a copy of my treatment plan
TEL	Possibility to access by phone	The GP gives to me the possibility to call him directly by phone or email
NUR	Cooperation with nurse	The GP works with a nonmedical professional (e.g. nurse) to improve my care
HYG	Health habit	The GP asks questions, either directly or not, about my health habits
AUT	Self-management	The GP helps me to self-manage my condition (e.g. He makes a treatment plan that I could carry out in my daily life)
PSY	Socio-psychological perspective	The GP asks how my chronic condition affects my life
RSP	Coordinated care	The GP is responsible of the organization of my care (e.g. GP "coordinator")
GP: General Physician		

The variant of the BWS method used in this study is the object case approach. The BWS tasks were generated from design desirable properties such as item balance (i.e. each item occurs the same number of times), pair balance (i.e. each BW/WB pair occurs the same number of time) and orthogonality (i.e. the items occur independently of each other through the tasks). The published

literature on the BWS studies is still limited, and then some design choices are partly arbitrary. The evidence on the relationship between task size (i.e. number of items) and responses consistency is sparse. Orme (2006) recommends no more than 6 items per task and by construction the minimum number is 3. In this study we use a 5-items size. On this basis the full factorial design leads to 252 possible tasks. To obtain a number of tasks manageable for respondents, we use a partially balanced incomplete block design (P-BIBD). With an item frequency of 3, it's possible to identify 6 tasks (via OPTEX procedure of SAS) (table 5).

Table 5. Design of BWS tasks					
Task n°1	Task n°2	Task n°3	Task n°4	Task n°5	Task n°6
RSP	SDM	SDM	TEL	CAL	AUT
NUR	VRF	VRF	RSP	PSY	SDM
HYG	RSP	CAL	VRF	AUT	INF
CAL	NUR	INF	PSY	TEL	PSY
INF	AUT	TEL	HYG	NUR	HYG

This design verifies the item balance and partially the property of pair balance. Each BW/WB pair (i.e. 90 possibilities) is available for selection, but 30 pairs have a frequency of 2 whereas 60 pairs appear only once. The orthogonality property is checked with an exam of coefficients correlation matrix which shows no coefficient significantly different from zero.

For each task the respondents are asked to choose both the "most" (i.e. best) and "least" (i.e. worst) important items. Then with this design each item has a bounded score [-3; +3], with (-3) when it is always selected as "least important" and (+3) as "most important".

First a count analysis is performed on the BWS data at both the sample (or aggregated) level and individual level to briefly explore the issue of preferences heterogeneity among respondents. This is appreciated in calculating for each attribute the mean individual score and the standard deviation around this mean (Sackett et al, 2011). Second the BWS data are analyzed with probability model as described in the theoretical presentation above. According to the literature, in the BWS models the attribute "NUR" (i.e. cooperation between general practitioner and allied health professional, especially nurse) is omitted to avoid saturated model. The database is structured as following:
- The unit of analysis is the BW/WB pair
- For each task each respondent chooses one pair among 20
- The chosen pair is coded "1" and the remaining "0"
- For each pair, the "best" attribute is coded "+1" and the "worst" attribute is coded "-1", the remaining are coded "0"

2/ Discrete Choice Experiment (DCE) method

2.1/ Theoretical presentation

The use of this method is constantly increasing with 40 references in the early 2000s to nearly 260 references in 2011. This method is largely described in the books of Ryan et al (2008) and Louviere et al (2000). In this section we detail two methodological points of our study that affect the format of choice task and the design of the tasks.

First we use a "**simple choice format**" (i.e. options are presented to respondents one by one and for each of them there is a binary response possibility: acceptance or rejection). This DCE task format has implications for the modeling.

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

$V_{ij} = \sum_{k=1}^{K} \beta_k . X_{ijk}$ (*)

$$Y_{ij} = 1 \ (or \ acceptance) \ if \ U_{ij} > 0$$

$$Y_{ij} = 0 \ (or \ rejection) \ otherwise$$

Under the hypothesis of iid EV1 errors:

$$P(Y_{ij} = 1) = \frac{\exp(V_{acceptance})}{\exp(V_{acceptance}) + \exp(V_{rejection})}$$

By assumption: $U_{rejection} = 0 \ ==> \ \exp(0) = 1$

Then the probability model becomes (i.e. binary "unconditional" logit model):

$$P(Y_{ij} = 1) = \frac{1}{1 + e^{-(V_{ij})}}$$

(*) Exactly of the form:

$$V_{ij} = \lambda_{ij} \sum_{k=1}^{K} \beta_k . X_{ijk}$$

In limited dependent variables model (LDV) such as discrete choice model, the parameter estimates are not means, they are perfect confound of means and variances on the latent scale (Flynn, 2010). The variance is represented by an error scale parameter (i.e. EV1 scale parameter) which is inversely related to the dispersion (variance or standard deviation) of the error component.

$$\lambda_{ij} = \frac{\pi}{SD(\varepsilon_{ij})\sqrt{6}}$$

Where $(\frac{\pi}{\sqrt{6}})$ corresponds to the variance of standard Gumbel (or EV1) distribution.

In practice the influence of this parameter on the estimates is neutralized in normalizing them to the unit (λ=1). Implicitly this means that the errors are constant within the respondents (i.e. the respondents perform equally the different tasks whatever their number) and constant across the respondents (i.e. all the respondents are equal in terms of capacity to perform the tasks). The implication of this result is that estimates from different LDV model cannot be directly compared due to potential differences in scale parameters.

In the discrete choice part of this study we use the same 10 attributes as in the BWS study. Due to this relatively high number of attributes (K=10) the DCE tasks are designed following an "**attribute blocked design**" strategy (Witt et al, 2009). This kind of design consists in the purposive or random distribution of the attributes among two separate versions of the choice questionnaire (i.e. two DCE are carried out in parallel). However at least one attribute (preferably two) is (are) kept in the two versions to check the possibility of data pooling. This approach is based on a strong assumption that individuals value the attributes in the same way regardless of their presentation (i.e. together or by block). To assist this assumption it's generally asked respondents to imagine that options in comparison in the choice task are "equivalent all things being equal". However according to the choice format used in this study this vignette is not relevant since only one option is included in the choice tasks.

To allow data pooling without missing values on the attributes, we use three imputation *scenarii*:
- "Pessimism": In each choice questionnaire, all the omitted attributes are considered as not realized by the respondents.
- "Optimism": In each choice questionnaire, all the omitted attributes are considered as realized by the respondents.
- "Random": In each choice questionnaire, all the omitted attributes are considered randomly as realized or not realized by the respondents.

Under these *scenarii* we are able to pool the data and to verify the validity of the attribute block design approach in testing the null hypothesis of no difference between estimates of separate model and estimates of pooled model. The validity of this approach is also checked by comparative analysis (i.e. Chi-2 test) of some descriptive statistics (e.g. easiness, success to [apparent] rationality test).

2.2/ Use of the method

In this study, the DCE is based on the 10 CCM attributes previously identified. Each attribute has two extreme levels (i.e. realized versus not realized). This format facilitates the reading (i.e. information acquiring) of tasks by respondents, but has potential to reinforce lexicographic preferences (Howard & Salkeld, 2009, Kjaer et al, 2006, Scheibehenne et al, 2007). The distribution of the attributes across the two choice questionnaires is random with 2 common and 4 specific attributes (table 6).

Table 6. Distribution of CCM attributes across the two choice questionnaires.		
Version n°1	**Version n°2**	
SDM	SDM	*Common*
INF	INF	
CAL	HYG	*Specific*
VRF	PSY	
TEL	AUT	
NUR	RSP	

In each version, the full factorial combination of attributes' levels leads to 64 options. To obtain a more manageable number of options for respondents we use an orthogonal main effects plan (OMEP) generated with NGENE software. The final DCE design contains 8 tasks/options (table 7).

Table 7. DCE design						
Task n°	**SDM**	**INF**	**CAL or HYG**	**VRF or PSY**	**TEL or AUT**	**NUR or RSP**
1	No	No	No	No	No	No
2	Yes	No	Yes	No	Yes	No
3	Yes	Yes	No	Yes	No	No
4	No	Yes	Yes	Yes	Yes	No
5	Yes	Yes	Yes	No	No	Yes
6	No	Yes	No	No	Yes	Yes
7	No	No	Yes	Yes	No	Yes
8	Yes	No	No	Yes	Yes	Yes

This design verifies the properties of orthogonality and level balance (attribute frequency = 4). According to the task format (i.e. simple choice), the condition of minimal overlap does not appear to be appropriate.

Finally this design is marginally revised by including an additional task to test the response stability (i.e. the task n°2 is repeated at the end of the questionnaire) and the task n°1 is posited at the 2nd place in the questionnaire because its particular nature (i.e. all the attributes are not realized) seems to be too confusing for respondents when placed on 1st position. Moreover this task is used to test the reversed hypothesis of monotonicity in preferences (i.e. individuals always prefer more to less). Then a "consistent" respondent is supposed to reject the task n°1 and to respond in the same way at the tasks n°2 and n°9.

The DCE data are analyzed with a multilevel approach (i.e. random intercept model) to account for potential correlation of observations within respondents.

3/ Sample

The participants have been consecutively recruited between October 2010 and June 2011 in a hospital department of respiratory medicine (Saint Antoine Hospital, Paris). Patients are deemed eligible when they satisfy four criteria:
- To come for a sleep apnea syndrome (SAS) follow-up visit
- To be treated for at least 1 year and to be compliant
- To have consulted at least once a GP in the last 6 months
- To suffer from at least two chronic diseases

The SAS has been selected to recruit chronically ill patients because it is frequently associated with a lot of comorbidities, such as diabetes and high blood pressure (Chung et al, 2002). Moreover the effect of the SAS treatment on the GP-patient relationship is in large part controlled by the high level of treatment efficacy (near than 90%) when patients are compliant. That's why we recruit patients with a treatment length no shorter than 1 year. Moreover we assume that only patients who frequently consult their GP are able to value improvements in the GP care, because they have sufficient care experience. Moreover we focus more specifically on patients with multiple long-term diseases because they are patients who may benefit the most from improvements in the GP care.

Given the complexity of the stated preferences (i.e. BWS and DCE) tasks for the respondents we use a face-to-face assisted process of questionnaire distribution and completion. To define the sample size we use the same formula as Aspinall et al (2008) based on the parameters of the discrete choice task (i.e. number of tasks [T], choice format [F], maximum number of levels per attribute [L]).

$$Sample\ size = \frac{500.L}{T.F}$$

According to this formula a minimal sample size of 125 individuals is required. Next, we anticipate a net response rate (i.e. response rate after deletion of responses with a lot of missing values or which are inconsistent) of 83%[1]. Then 150 individuals have to be recruited for this study. It was necessary to contact 159 patients to obtain 150 respondents (response rate of 94%).

---

[1] Statistic calculated from 39 health DCE using assisted completion of the questionnaire.

Results

1/ Sample

The 150 chronically ill patients recruited suffer an average of 3.3 chronic diseases (SD: 1.02, Min: 2, Max: 7) that are mainly high blood pressure (27.7%), diabetes (17.5%), and cardiovascular disorder (12.5%). These patients are mostly men (72%) and are 63.4 years old on average (SD: 9.25). Their SAS appear to be controlled for several years (table 8).

Table 8. Individual characteristics of respondents (n=150).		
Characteristic		**%**
Age	*Less than 60 years*	30
	Between 60 and 65 years	32,7
	More than 65 years	37,3
Number of chronic disease	*2 disease*	22
	3 diseases	39,3
	4 or more diseases	38,7
Perceived health	*Excellent, Very good, Good*	47,3
	Acceptable, Bad	52,7
Gender	*Man*	72
	Woman	28
Length of SAS treatment	*1 to 5 years*	36,7
	6 to 10 years	24
	More than 10 years	39,3

2/ BWS

Therefore with 150 respondents, 6 BWS tasks, and 20 pairs per task, the database for econometric analysis contains 18 000 rows.

The attribute that is selected most often as the best is the continuity of medical information among health professionals (i.e. INF). It was designed as such in 51.1% of the cases (230/450). At the opposite, the attribute related to the cooperation between GP and allied health professional (i.e. NUR) has been selected in 69.6% of the cases (313/450) as the least important (i.e. worst).

The count analysis identifies a ranking of the attributes in three distinct groups (table 9):

- Informational continuity (INF) and coordinated care (RSP) that are clearly identified as the most important attributes of the GP care for chronically ill patients. The second important attribute (RSP) is about 3.5 times higher than the attribute in third position (PSY).

- Planned care (CAL) and professional cooperation (NUR) that are clearly identified as the least important attributes of the GP care for chronically ill patients. The "best" attribute (INF) is about 33 times higher than the "worst" attribute (NUR).

- Remaining attributes. Despite a ranking of these 6 attributes at different positions of the cardinal scale of importance, they appear to be broadly of similar importance. The attribute in third position (PSY) is 2.1 times higher than the attribute in seventh position (SDM).

Table 9. BWS results - Sample-level count analysis.

Attribute	Best	Worst	Best-Worst	Sqrt(Best/Worst)
INF	230	8	222	5,36
RSP	184	8	176	4,80
PSY	119	65	54	1,35
AUT	70	46	24	1,23
HYG	88	60	28	1,21
VRF	85	86	-1	0,99
TEL	66	99	-33	0,82
SDM	40	102	-62	0,63
CAL	10	113	-103	0,30
NUR	8	313	-305	0,16

The relationship between the BW ratio and the best counts appear to be linear with a R² close to 90% (graphic 2). This result confirms the possibility of using the BW ratio as an estimate of the position attributes on the scale of importance. This is consistent with the international literature on the BWS method.

Graphic 2. Sqrt(BW ratio) versus Best counts

$$Y = 0,0244.X - 0,5125$$
$$R^2 = 0,902$$
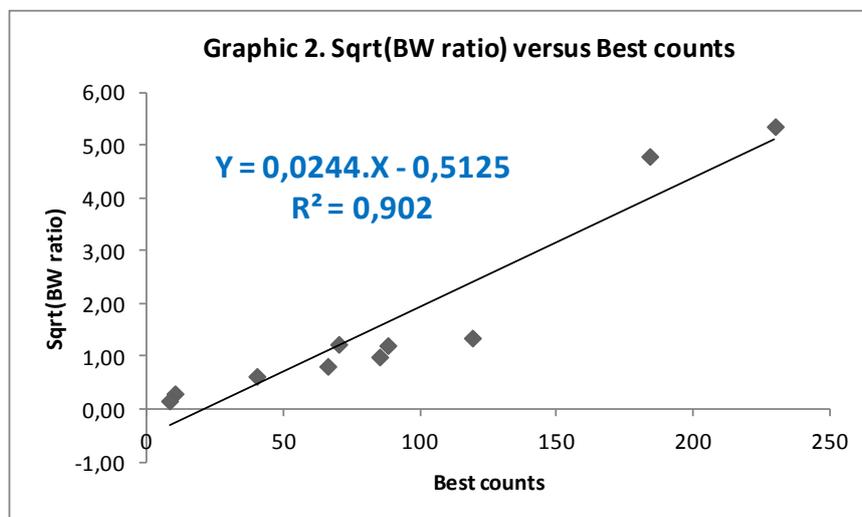
A similar count analysis at the individual level leads to the same results in terms of attributes importance (table 10). The mean score of importance for the informational continuity (i.e. INF), which is bounded to [-3; +3], is around 1.5. Moreover this disaggregated analysis highlights a complementary result about the level of preferences heterogeneity in these mean scores. The attributes with a moderate score of importance (i.e. score close to zero) show larger standard deviation. This result indicates the coexistence of different groups of patients with different preferences among the respondents.

| Table 10. Individual-level count analysis (n=150).				
Item	Mean	SD	CI 95%	
INF	1,48	1,19	1,29	1,67
RSP	1,17	1,30	0,96	1,38
PSY	0,36	1,56	0,11	0,61
HYG	0,19	1,45	-0,05	0,42
AUT	0,16	1,14	-0,02	0,34
VRF	-0,01	1,56	-0,26	0,24
TEL	-0,22	1,49	-0,46	0,02
SDM	-0,41	1,30	-0,62	-0,20
CAL	-0,69	1,12	-0,87	-0,51
NUR	-2,03	1,29	-2,24	-1,83

The conditional logit estimates (table 11) indicate the importance of each attribute relative to the origin point of this ratio that is defined by the omitted attribute (i.e. NUR). To make the interpretation easier these estimates can be converted in a probability scale[2] (Jaeger & Cardello, 2009). Thus the likelihood of the attribute "informational continuity" (INF) being chosen in preference to the attribute "professional cooperation" (NUR) is approximately 19:1. These odds (versus NUR) are of 14:1 for RSP, 8:1 for PSY and HYG, 5:1 for AUT and VRF, 4:1 for TEL and SDM, and 3:1 for CAL.

---

[2] Probability= $100.\frac{\exp{(U_i)}}{\sum \exp{(U_j)}}$

Table 11. BWS results - Conditional Logit estimates *					
Attribute	**Estimate**	**SE**	**P-value**	**95% CI**	
INF	2,94	0,11	0,000	2,71	3,16
RSP	2,61	0,11	0,000	2,40	2,82
PSY	2,03	0,11	0,000	1,81	2,26
HYG	2,02	0,11	0,000	1,80	2,24
AUT	1,67	0,10	0,000	1,47	1,88
VRF	1,66	0,11	0,000	1,45	1,88
TEL	1,38	0,11	0,000	1,17	1,59
SDM	1,31	0,10	0,000	1,11	1,52
CAL	0,94	0,10	0,000	0,75	1,13
NUR	*(omitted)*				
* Log Likelihood = -3430.4; Nbr observations = 18 000					

The strong linear relationship between the conditional Logit (CL) estimates and the (Best-Worst) score confirms that the count analysis is a valid approach to posit all the attributes on a same ratio scale of importance (graphic 3).

Graphic 3. (Best-Worst) versus CL estimates

$R^2 = 0,97$
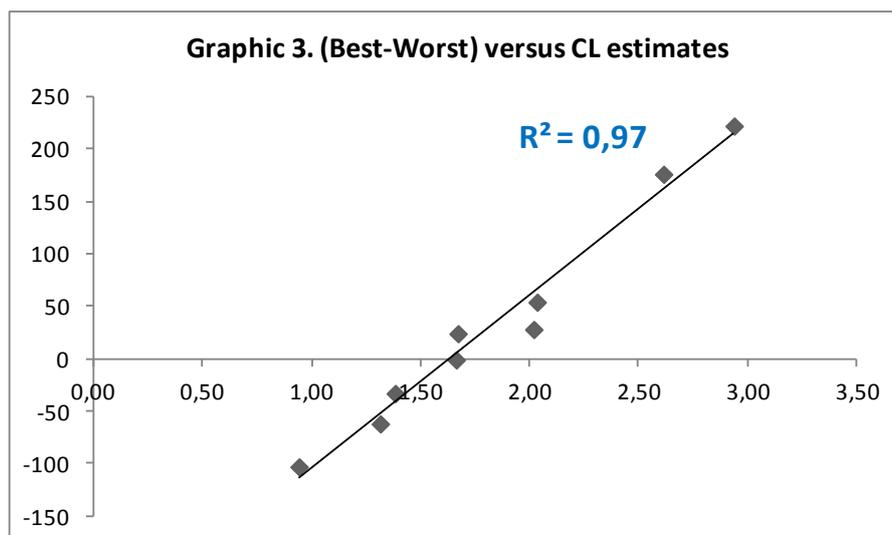
3/ DCE

Among the 150 respondents only 7 (4.7%) have not accepted any GP care option. However the objective of this analysis is to obtain results at the population (or sample) level, then these respondents were maintained in the choice data analysis. On average the chronically ill patients have accepted 3 options among 8.

The responses largely satisfy the monotonicity and stability assumptions with respectively 0.7% and 12% of respondents who fail the tests. In addition, nearly 70% of respondents felt the DCE questionnaire as easy or very easy and 94% as interesting or very interesting.

In the first choice questionnaire, the acceptance rate of the GP care options is around 37.7%. The results confirm the theoretical validity of the model, all but one of the estimates are significantly different from zero and of positive sign (table). The only not significant attribute is the professional cooperation (i.e. NUR). This indicates that achieving the "NUR" attribute does not improve the utility gained from the GP care offer. These logit estimates can be exponentiated to obtain results in terms of probability[3]. For example when a GP is continuously informed of the changes in the medical condition of his patients (i.e. INF) then he is 15 times more likely to be accepted by chronically ill patients than a GP without informational continuity, all things being equal.

According to the effect coding of the DCE data, the model intercept can be interpreted as a general tendency (i.e. inclination) to accept or reject the GP care options. In both DCE questionnaires the intercept estimates are of negative sign thus indicating a tendency to reject the choice options.

In the second version of the DCE questionnaire, the acceptance rate of the GP care options is around 36.8%. As in the first version the results confirm the theoretical validity of the model, all the attributes are significant and of expected signs (table 12). All things being equal the informational continuity (i.e. INF) is still the attribute with the larger effect on the acceptance probability of the GP care options. In this version of the questionnaire this attribute increases the chances of option selection by 10. The second attributes with the larger effect are the care coordination (i.e. RSP) and the socio-psychological perspective (i.e. PSY) that increases the chances by 4. The model intercept is also of negative sign.

In both discrete choice models the variance of the random effect around the model intercept is significant[4] thus justifying the multi-level approach to account for intra-individual correlation.

---

[3] Odds Ratio = $\exp(2.\beta_k)$
[4] Version n°1: β=0.37 (SE: 0.22) and in version n°2: β=1.15 (SE: 0.39).

Table 12. DCE results - Binary Logit estimates

Attribute	Version n°1 *			Version n°2 **		
	Estimate	SE	P-value	Estimate	SE	P-value
Intercept	-0,97	0,15	<0,0001	-1,01	0,18	<0,0001
SDM	0,76	0,14	<0,0001	0,44	0,13	<0,0001
INF	1,35	0,15	<0,0001	1,16	0,14	<0,0001
VRF	0,85	0,13	<0,0001	-	-	-
CAL	0,48	0,13	<0,0001	-	-	-
IDE	0,03	0,12	0,797	-	-	-
TEL	0,62	0,14	<0,0001	-	-	-
HYG	-	-	-	0,6	0,12	<0,0001
PSY	-	-	-	0,7	0,13	<0,0001
AUT	-	-	-	0,31	0,12	0,012
RSP	-	-	-	0,74	0,13	<0,0001

Log Likelihood = -293,5; Pseudo-R²=26,2%

** *Log Likelihood = -311,2; Pseudo-R²=20%*

To check the validity of the attribute blocked design approach we first compare statistics between the 2 choice questionnaires (table 13). This analysis shows no significant difference at the 5% level and indicates similar respondents' profiles in the two versions.

Table 13. Comparisons of individual characteristics between the two versions of the DCE questionnaire.

| Characteristic	DF	Chi-2	P-value
Age	2	0,51	0,776
Nbr of chronic disease	2	3,43	0,180
Nbr of annual medical visit	1	0,67	0,413
Length of doctor-patient relationship	1	3,15	0,076
Medical facility	1	0,03	0,870
Perceived health	1	1,31	0,252
Length of SAS treatment	2	5,23	0,073
Gender	1	1,19	0,275

The estimates obtained from the different imputation *scenarii* show 2 main results (table 14). First, the "pessimism" and "optimism" *scenarii* lead to similar estimates for the CCM attributes, unlike the "random" scenario that causes important changes in the estimates values and in the attributes order. Thus this last scenario was abandoned. Second, the "pessimism" and "optimism" *scenarii* show opposite estimates for the model intercept. In the first scenario the intercept value indicates a positive inclination of chronically ill patients to accept GP care options, whereas in the "optimism" scenario the estimated intercept is of negative sign and therefore indicates a global trend to reject the GP care options. According to these results and those obtained previously on the separate models the "optimism" scenario was used to test the validity of choice data pooling.

Attribute	"Pessimism"		"Optimism"		"Random"	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Intercept	1,17	< 0,0001	-3,14	< 0,0001	-0,74	< 0,0001
SDM	0,6	< 0,0001	0,6	< 0,0001	0,32	< 0,0001
INF	1,25	< 0,0001	1,26	< 0,0001	1	< 0,0001
VRF	0,91	< 0,0001	0,89	< 0,0001	0,45	< 0,0001
CAL	0,53	< 0,0001	0,51	< 0,0001	-0,05	0,53
IDE	0,04	0,66	0,03	0,80	-0,46	< 0,0001
TEL	0,61	< 0,0001	0,59	< 0,0001	0,13	0,15
HYG	0,55	< 0,0001	0,57	< 0,0001	0,11	0,20
PSY	0,68	< 0,0001	0,71	< 0,0001	0,27	0,00
AUT	0,27	0,01	0,3	0,01	-0,17	0,04
RSP	0,68	< 0,0001	0,72	< 0,0001	0,21	0,01

Table 14. DCE results - Binary Logit estimates according imputation *scenarii*

The likelihood ratio tests between the pooled model and the separated model fail to reject the null hypothesis of no difference between parameter estimates ($H_0: \beta^{pooled} - \beta^{separated} = 0$) at the 5% level. So the final CCM results indicate that only the professional cooperation (i.e. NUR) is not valued by chronically ill patients to improve the GP care.

4/ BWS versus DCE

With the BWS approach the conditional logit estimates indicate the importance of each CCM attribute relative to an omitted attributed that is the professional cooperation (i.e. NUR). With the DCE "pooled" approach the binary logit estimates indicate the impact of each CCM attribute on the choice decision. According to the format of the attributes' levels (i.e. two extremes levels), we can suppose that the marginal estimate from moving to "not realized" level to "realized" level is a fair proxy of the attribute "importance" for respondents. Then a simple linear regression of the DCE estimates against the BWS estimates (graphic 4) shows a moderate linear relationship (R²=44%). This result suggests a lack of correspondence between these two stated preferences methods to study the importance of attribute (i.e. case 1 BWS) in the choice decision.

Graphic 4. DCE Pooled estimates versus BWS CL estimates

$\beta_{DCE} = 0,28.\beta_{BWS} + 0,16$

$R^2 = 0,44$
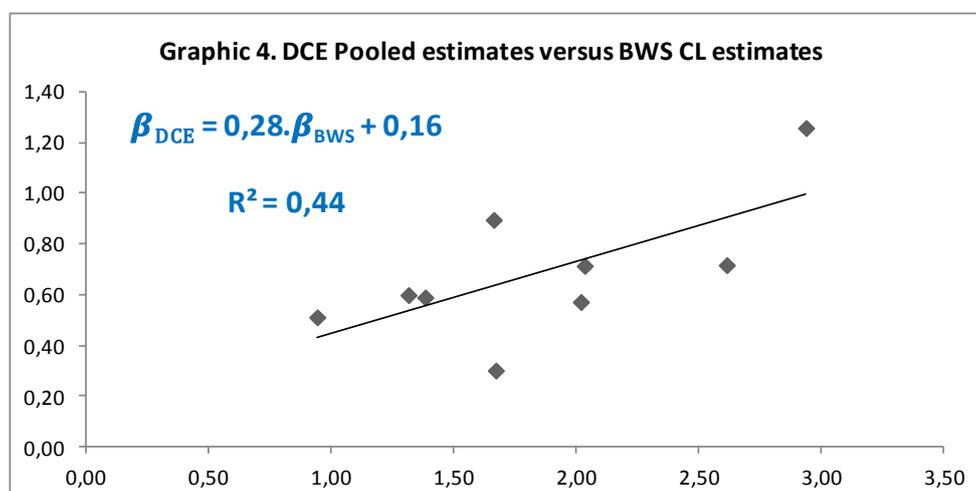
Discussion

This study shows that patients value the different recommendations of the Chronic Care Model to improve the GP care towards chronic disease management. Thus our first hypothesis seems to be verified and the CCM appears to be an appropriate framework to rethink the GP care delivery. Moreover we show that patients value highly changes that do not require much involvement in their care. The CCM attributes relative to shared decision making (SDM) or self-management (AUT) appear to be relatively less important both in the DCE and BWS results. At the opposite, the informational continuity of the physician (INF) and the coordination of care by the physician (RSP) are of particular importance in the care options assessment. This result seems consistent with the international literature that identifies "coordination" and "patient engagement" as the two main areas to improve the quality of care for chronically ill patients (WHO, 2005, Rijken et al, 2008).

Our 2nd hypothesis is also in large part verified. The chronically ill patients perceive all but one CCM attributes as a useful way to improve the GP care delivery. The only no significant attribute concerns the professional cooperation (NUR) between the general practitioner and allied health professionals, mainly nurse. The literature on patients' preferences for GP care highlights the effect of experience and severity of ailments (Turner et al, 2007, Porteous et al, 2006, Caldow et al, 2007). Initially patients prefer to consult a physician rather than a nurse, but acquiring medical experience of nurse consultation or a large improvement of GP care attributes (e.g. strong reduction of waiting times) allows changes in patients' preferences. In this study the respondents do not usually experience the intervention of a nurse in their GP care, then they may have difficulties to perceive the benefits of such professional cooperation in the management of their chronic diseases. Further research needs to take into account the current medical situation of patients to better understand their preferences for care improvements as identified by stated preferences methods. Also, he DCE results show a general tendency of chronically ill patients to reject the GP care options. In the literature this pattern of response is generally identified as a "*statu quo*" bias (i.e. conservative response) (Salkeld et al, 2000). According to this psychological process we can posit that individuals value the benefit or utility of a choice option according to an "anchor" (or reference comparator) which can be explicitly defined in the choice task (i.e. "constant comparator task format") or internally identified in the respondent experience.

From a methodological point of view we use BWS and DCE results to compare the usefulness of theses stated preferences methods in identifying the "importance" of attributes in the decision making process. The third hypothesis can be rejected with confidence, the two sets of estimates are not similar (up to scale). In a recent paper (not published), Flynn et al (2011) also investigate the relationship between BWS and DCE results in a case 2 (or profile case) BWS study and show a strong linear relationship (R^2=81%). This result suggests that BWS approach can be used as an alternative to traditional DCE to identify preferences. Moreover the estimated slope (i.e. scale factor) of this linear relationship seems to indicate that BWS method has potential to identify similar results to the DCE method but with smaller sample size. In this study we are not able to confirm this result. This could

be explained first by differences in the modeling framework. Unlike Flynn et al (2010), our BWS and DCE logit models do not have the same structure and consequently the estimated effects are not strictly the same. However the DCE format of the CCM attributes in this study allows estimates closed to the (unobservable) attribute "importance". Potoglu et al (2011) in a similar attempts show also differences between (rescaled) BWS and DCE results. A second potential explanation of this discrepancy between DCE and BWS results could be that these two methods are inherently not equivalent. With a case 2 BW approach, the BWS and DCE tasks are associated within a same task, and then we can assume that the two sets of estimates are not fully independent of each other. The completion of the BWS task in first could encourage respondents to adopt particular heuristic when completing the DCE task. For example a choice option with a large distance between the best and worst attributes may have a higher probability of acceptance than an option for which the extremes attributes are closed. Thus this process tends to align the BWS and DCE results.

This study leaves some unexplored issues. First as in usual discrete choice modeling we do not take into account a potential effect of heteroskedastic error among respondents. This could be problematic in comparing BWS and DCE results while the first method is supposed easier for respondents than the second stated preferences method. Then a track for future research could be to analyze if the BWS method is easier for all respondents or only for a part of them. Moreover it is unclear in which circumstances the BWS method is easier than the DCE method. The respondents seem to be very sensitive to the framing of the BWS tasks (e.g. wording of "best"/"worst" scale). Initially we sought to use the case 2 BWS approach, but it appeared too confusing for respondents to identify for example the "best" item when all the attributes' levels were of negative value. Further research on this topic is needed. The second unexplored issue is related to the heterogeneity of patients' preferences. In this study we used a basic approach to quantify the magnitude of this phenomenon, but more sophisticated methods are available such as latent class analysis (LCA). The aggregated (or sample-level) analysis could be misleading while a same importance score may combine individuals with opposing preferences.

Conclusion

This study shows that the Best-Worst Scaling (BWS) method is an appropriate method to identify the importance of GP care attributes for chronically ill patients. Among the many recommendations of the CCM, those relative to the general practitioners' tasks seem to take precedence over a higher engagement of patients in the care management.

Bibliography

ANDERSON G. The Growing Burden of Chronic Disease in American. Public Health Reports, 2004, vol 119.

ASPINALL PA, HILL AR, DHILLON B, et al. Quality of life and relative importance: a comparison of time trade-off and conjoint analysis methods in patients with age-related macular degeneration. The British journal of ophthalmology, 2007, 91, 6, p.766-772

BODENHEIMER T, WAGNER EH, GRUMBACH K. Improving primary care for patients with chronic illness. JAMA, 2002, 288, 14, p.1775-1779

BOULT C, KARM L, GROVES C. Improving chronic care: the "guided care" model. Perm J, 2008, 12, 1, p.50-54

BOULT C, REIDER L, LEFF B, et al. The effect of guided care teams on the use of health services: results from a cluster-randomized controlled trial. Arch Intern Med, 2011, 171, 5, p.460-466

BRAS PL, DUHAMEL G, GRASS E. Improving the care of the chronically ill: lessons from foreign disease management experience. Prat Organ Soins, 2006, 37, 4, p.341-356

CALDOW J, BOND C, RYAN M, et al. Treatment of minor illness in primary care: a national survey of patient satisfaction, attitudes and preferences regarding a wider nursing role. Health expectations, 2007, 10, 1, p.30-45

CHERAGHI-SOHI S, HOLE AR, MEAD N, et al. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann Fam Med, 2008, 6, 2, p.107-115

CHUNG S, JAIRAM S, HUSSAIN MR, et al. How, What, and Why of Sleep Apnea - Perspectives for Primary Care Physicians. Canadian Family Physician, 2002, 48, p.1073-1080

COHEN SH, NEIRA L. Measuring preference for product benefits across countries: Overcoming scale usage bias with Maximum Difference Scaling. ESOMAR 2003 Latin America Conference Proceedings, 2003.

DAVISON SN, KROMM SK, CURRIE GR. Patient and health professional preferences for organ allocation and procurement, end-of-life care and organization of care for patients with chronic kidney disease using a discrete choice experiment. Nephrology, dialysis, transplantation, 2010, 25, 7, p.2334-2341

FINN A, LOUVIERE JJ. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. Journal of Public Policy & Marketing, 1992, 11, p.12-25

FLYNN TN, LOUVIERE JJ, PETERS TJ, COAST J. Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. BMC Medical Research Methodology, 2008, 8, 76, 12 p.

FLYNN TN, PETERS TJ, COAST J. Quantifying response shift or adaptation effects in quality of life by synthesizing best-worst scaling and discrete choice data. International Choice Modelling Conference, 2011, Leeds, UK. 26 p.

FLYNN TN. Valuing citizen and patient preferences in health: Recent developments in three types of best-worst scaling. Expert Review of Pharmacoeconomics & Outcomes Research, 2010, 10, 3, p.259-267

GERARD K, LATTIMER V, SURRIDGE H, et al. The introduction of integrated out-of-hours arrangements in England: a discrete choice experiment of public preferences for alternative models of care. Health expectations, 2006, 9, 1, p.60-69

GERARD K, LATTIMER V. Preferences of patients for emergency services available during usual GP surgery hours: a discrete choice experiment. Family practice, 2005, 22, 1, p.28-36

GLASGOW RE, WAGNER EH, SCHAEFER J, et al. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). Med Care, 2005, 43, 5, p.436-444

HJELMGREN J, ANELL A. Population preferences and choice of primary care models: a discrete choice experiment in Sweden. Health Policy, 2007, 83, 2, p.314-322

HOWARD K, SALKELD G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. Value Health, 2009, 12, 2, p.354-363

JAEGER SR, CARDELLO AV. Direct and indirect hedonic scaling methods: A comparison of the labeled affective magnitude (LAM) scale and best-worst scaling. Food Quality and Preference, 2009, 20, p.249-258

KJAER T, BECH M, GYRD-HANSEN D, HART-HANSEN K. Ordering effect and price sensitivity in discrete choice experiments: need we worry? Health Econ, 2006, 15, 11, p.1217-1228

LONGO MF, COHEN DR, HOOD K, et al. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. The British journal of general practice, 2006, 56, 522, p.35-42

LOUVIERE JJ, HENSHER DA, SWAIT JD. Stated choice methods: Analysis and application. 1ère éd. Cambridge University Press, 2000, 402 p.

LOUVIERE JJ, WOODWORTH. Best-worst scaling: A model for largest difference judgments. Working Paper, 1990.

MARLEY AAJ, LOUVIERE JJ. Some probabilistic models of Best, Worst, and Best-Worst choices. Journal of Mathematical Psychology, 2005, 49, p.464-480

MCFADDEN D. Conditional logit analysis of qualitative choice behavior. In: Frontiers in econometrics. ZAREMBKA P. New York: Academic Press, 1974, p.105-142

ORME B. Adaptive Maximum Difference Scaling. Sawtooth Software technical paper series, 2006.

PAITA M, WEILL A. Les personnes en affection de longue durée au 31 décembre 2008. Points de repère, 2009, 27, 12 p.

PORTEOUS T, RYAN M, BOND CM, HANNAFORD P. Preferences for self-care or professional advice for minor illness: a discrete choice experiment. The British journal of general practice, 2006, 56, 533, p.911-917

POTOGLU D, BURGE P, FLYNN TN, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Social science & medicine, 2011, 72, 10, p. 1717-1727
RIJKEN M, JONES M, HEIJMANS M, DIXON A. Supporting self-management. In: Caring for people with chronic conditions: a health system perspective. NOLTE E, MCKEE M. Press: Open University Press, 2008, p.116-142

RUBIN G, BATE A, GEORGE A, et al. Preferences for access to the GP: a discrete choice experiment. The British journal of general practice, 2006, 56, 531, p.743-748

RYAN M, GERARD K, AMAYA-AMAYA M. Using discrete choice experiments to value health and health care. Springer: Dordrecht, The Netherlands, 2008, 254 p.

RYAN M, WATSON V. Counting the cost of fast access: using discrete choice experiments to elicit preferences in general practice. The British journal of general practice, 2006, 56, 522, p.4-5

SACKETT H, SHUPP R, TONSOR G. Consumer Perceptions of Sustainable Farming Practices: A Best-Worst Scenario. Working paper, 2011.

SALKELD G, RYAN M, SHORT L. The veil of experience: do consumers prefer what they know best? Health Econ, 2000, 9, 3, p.267-270

SCHEIBEHENNE B, MIESLER L, TODD PM. Fast and frugal food choices: uncovering individual decision heuristics. Appetite, 2007, 49, 3, p.578-589

SWAIT J, LOUVIERE J. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research, 1993, 30, 3, p.305-314

TURNER D, TARRANT C, WINDRIDGE K, et al. Do patients value continuity of care in general practice? An investigation using stated preference discrete choice experiments. Journal of Health Services Research and Policy, 2007, 12, p.132-137

VALLIER N, SALAVANE B, WEILL A. Les aspects économiques des affections de longue durée. ADSP, 2007, 59, p.30-37

VALLIER N, WEILL A, SALAVANE B, et al. Coût des trente affections de longue durée pour les bénéficiaires du régime général de l'assurance maladie en 2004. Pratiques et Organisation des Soins, 2006, 37, 4, p.267-283

VICK S, SCOTT A. Agency in health care. Examining patients' preferences for attributes of the doctor-patient relationship. J Health Econ, 1998, 17, 5, p.587-605

WAGNER EH, AUSTIN BT, DAVIS C, et al. Improving Chronic Illness Care: translating evidence into action. Health Affairs, 2001, 20, 6, p.64-78

WAGNER EH, AUSTIN BT, VON KORFF M. Organizing Care for Patients with Chronic Illness. Milbank Q, 1996, 74, 4, p.511-544

WITT J, SCOTT A, OSBORNE RH. Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. Health Econ, 2009, 18, 6, p.681-696

WORLD HEALTH ORGANIZATION (WHO). Preparing a health care workforce for the 21st century. The challenge of chronic conditions. Rapport, 2005, 65 p.

WORLD HEALTH ORGANIZATION (WHO). Preventing chronic diseases: a vital investment. The World Health Report, 2005, 203 p.

WU SY, GREEN A. Projection of chronic illness prevalence and cost inflation. Santa Monica, CA: RAND Health, 2000.