Consideration sets and their role in modelling doctor recommendations about contraceptives

Denzil G. Fiebig*[a], Rosalie Viney[b], Stephanie Knox[c], Marion Haas[b], Deborah J. Street[d], Arne R. Hole[e], Edith Weisberg[f,g], and Deborah Bateson[f,g]

[a]School of Economics, University of New South Wales
[b]CHERE, University of Technology Sydney
[c]National Centre for Immunisation Research and Surveillance, The Children's Hospital Westmead
[d]School of Mathematical Sciences, University of Technology Sydney
[e]Department of Economics, University of Sheffield
[f]Family Planning NSW
[g]Department of Obstetrics, Gynaecology and Neonatology, University of Sydney, Australia

Abstract

Decisions about prescribed contraception are typically the result of a consultation between a woman and her doctor. In order to better understand contraceptive choice within this environment, stated preference methods are utilized to ask doctors about what contraceptive options they would discuss with different types of women. The role of doctors is to confine their discussion to a subset of products that best match their patient. This subset of options forms the consideration set from which the ultimate recommendation is made. Given the existence of consideration sets we address the issue of how to model appropriately the ultimate recommendations. The estimated models enable us to characterize doctor recommendations and how they vary with patient attributes and to highlight where recommendations are clear and when they are uncertain. The results also indicate systematic variation in recommendations across different types of doctors and in particular we observe that some doctors are reluctant to embrace new products and instead recommend those that are more familiar. Such effects are one possible explanation for the relatively low uptake of more cost effective longer acting reversible contraceptives and indicates the need for further education and training of doctors may be warranted.

Keywords: contraceptives; doctor recommendations; stated preferences; consideration sets
JEL: I10; J13; C25; C81

* Corresponding author:
School of Economics, UNSW, Sydney, Australia 2052; Ph: +61-2-9385 3958; d.fiebig@unsw.edu.au

1. Introduction

Maintaining good reproductive health requires women to have access to information about safe, effective and affordable methods of contraception. But contraceptive needs differ considerably over the reproductive life cycle and the range of contraceptive products available has expanded rapidly making it more difficult for women to understand the choices available and the trade-offs they present, and for doctors to provide comprehensive information to assist women in making a fully informed contraceptive choice.

Our overarching objective is to provide insights into medical decision-making in the case of doctors making recommendations about contraceptive products. The focus on the role of doctors means we abstract from some of the elements of the agency relationship that characterize this decision-making process. It also means that market data on final product choices is not directly applicable. As such stated preference (SP) methods are used to ask doctors about what contraceptive options they would discuss and ultimately recommend with different types of women. The role of GPs is to use their expertise to match women with the appropriate set of products and then to discuss what is likely to be a subset of all available products. This is in fact what we observe in the data. For the doctors, their situation is similar to the two-stage decision process mapped out by Manski (1977) where an agent first screens all available alternatives, selects a subset of products from the universe of total alternatives and then makes their ultimate choice from a subset of the available options that forms the consideration set. Various explanations have subsequently been provided for this process including exogenous constraints on choices, product awareness, and costs of acquiring and processing product information; see Hauser and Wernerfelt (1990), Roberts and Lattin (1991), Gilbride and Allenby (2004), Belkar et al. (2006) and Goeree (2008).

Doctors are making a recommendation for a patient rather than a choice for themselves and so explanations are different. Doctors are likely to be aware of a wide range of available contraceptive products but face a time constraint that may restrict the number of contraception options that they will discuss during a consultation with a particular woman. Moreover, good clinical practice may routinely preclude some products for certain women. Despite these differences, we maintain the consideration set terminology and draw on the related literature to provide guidance for our modelling.

In the SP choice task we ask a sample of Australian general practitioners (GPs) to complete a hierarchical sequence of three questions. For a given woman, defined by experimentally manipulated socioeconomic and clinical characteristics, GPs were initially asked to decide whether they would confine their discussions of contraceptive options according to three broadly defined categories of contraceptive products. Then within the broad category they chose, they were asked to indicate which specific contraceptive products would form the consideration set to be discussed with the patient. Finally, they chose one product that they would recommend as best suited to the patient from the consideration set that they had specified in the second question of the choice task.

Because the characteristics of the patients are controlled within the SP environment, our data measure how a sample of GPs match recommended products with a patient defined by a mix of

predetermined socio-economic and clinical attributes. These data reveal considerable variation in recommendations associated with patient characteristics and the econometric analysis is used to identify particular types of women where there are extremes in what could be termed clinical ambiguity.

The analysis of GP recommendations provides a different perspective on the phenomenon of medical practice variation (MPV) that is pervasive across the health sector; see for example Wenneberg et al. (1982). Like Davis et al. (2000) and Grytten and Sørensen (2003) the analysis is conducted at the physician level in the area of primary care, but they rely on observational data where concerns about confounding factors are natural while our controlled SP environment avoids these concerns. Our analysis shows that GP characteristics have a substantial role in explaining recommendation patterns, lending support to the possibility that MPV is in part driven by differences in "styles of practice" that are associated with the GP's experience, training, expertise and receptiveness to new products. The issue of willingness to embrace new products is especially relevant given increasing support for the greater use of more effective longer acting reversible contraceptive methods in order to reduce unintended pregnancies and abortion rates; see for example Armstrong and Donaldson (2005) and American College of Obstetricians and Gynecologists Committee on Gynecologic Practice (2009).

The task of modelling the recommendations, given the existence of consideration sets, raises a methodological issue that we need to address. In standard choice modelling situations, the ultimate choice is assumed to be made from an exogenously given set of discrete alternatives. That approach would involve modelling our recommendations data using standard choice modelling techniques assuming that the particular products initially chosen by the GP form the consideration sets; see for example Roberts and Lattin (1991). Carson and Louviere (2006) are warranted in their criticism of such an approach. There is a real threat to econometric analysis because it was the doctors who indicated which products should form the choice set and hence the assumption of exogeneity required by standard choice modelling is potentially problematic.

While there has been considerable research into the motivation of a two-stage decision process: i.e. determine a consideration set and then choose from this restricted set, Horowitz and Louviere (1995) urge caution about the usefulness of formally modelling the ultimate choice as a two-stage process. They argue that a reasonable hypothesis is that the formation of a consideration set is simply a reflection of the underlying preferences and hence if these were known there would be no extra information to be gained from knowledge of consideration sets. Operationally, information on consideration sets will not be useful in predicting final choice given knowledge of preferences.

In our case we proceed along these lines. Our maintained hypothesis is that doctors will choose to restrict the number of contraceptive options they discuss and that choice is made on the basis of the same factors and associated weights that drive their ultimate recommendation. This approach is supported by our econometric results.

2. Background to the choice problem

While market or revealed preference data are available that characterize the contraceptive choices different types of women are currently making (Yusuf and Siedlecky (2007) and Gray and McDonald (2007)), these data provide little or no detailed information about what is driving these choices. In part, this is because the demand for prescribed contraceptive products does not fit the standard case of a single decision-maker making choices for themselves. A single decision-maker framework would only be appropriate in the extreme cases of a perfectly informed patient making the choice or where the doctor makes the choice acting as a perfect agent to maximize the utility of their patient. Instead, choice of a contraceptive product can more realistically be characterized as a doctor-patient interaction where a better informed doctor makes a recommendation and the woman subsequently makes a decision on the basis of her own information and that conveyed by the doctor.

There have been considerable developments in the contraceptive market, leading to the introduction of a wider array of oral contraception formulations with new hormonal constituents and lower dose formulations as well as new long-acting methods such as hormonal intra-uterine devices and implants, injections, dermal patches and vaginal rings. The availability of so many different contraceptive choices poses challenges for GPs in providing the balanced and comprehensive information necessary for women to be able to make an informed choice. In this context it is important to know more about what is driving the recommendations of doctors and hence ultimately the decisions of patients given the doctor's recommendation.

SP methods provide a natural methodology to learn more about this particular choice process. Fiebig et al. (2011) note the limited variability in key product characteristics that is observed in market data for contraceptives and stress that it is this type of information that is required in order to predict how women will respond to new products that constitute different bundles of characteristics. In order to address this issue, they develop a discrete choice experiment to estimate woman's preferences for the attributes of different contraceptive products. In a natural extension of this work, Knox et al. (2012) investigate discrepancies in the evaluations of contraceptive product attributes made by doctors compared to those made by patients.

The SP choice task used here was designed to complement this existing research by focussing on the doctor's recommendation. The task was developed to reflect a typical consultation between a woman and her GP in relation to contraception but with a focus on the GP's decision about what products they would recommend. In doing so we abstract from the product attributes and instead focus on how the patient characteristics impact on the recommendation of the GP. The underlying motivation for this approach is to determine under what circumstances GPs would recommend particular products and to explore whether GP characteristics have any impact on the recommendations.

3. Development of the choice task

The choice task was developed based on findings from a literature review and from focus groups conducted with GPs in Australia. The focus groups were conducted by one of the clinical authors, an expert in reproductive health. Specifically, the participants were asked to consider the issues and options they would discuss with three hypothetical women, who were chosen to cover a range of different life cycle and fertility stages. The three "patients" discussed were a 17 year old who has previously only used condoms but is now in a permanent relationship and has come for advice about contraception, a 34 year old who has just finished breast feeding, is using the progestogen-only (mini-pill) and does not want more children and a 43 year old with 3 children, who wants to avoid pregnancy but who has health concerns about the combined pill. GPs indicated what options they would discuss with the women, the extent to which they would take their previous contraceptive history, likes and dislikes into account, the reasons they would counsel against specific types of contraceptives and what personal characteristics and contraceptive attributes they believed were of most relevance and importance in each specific situation. Focus group discussions were recorded and transcribed, and a thematic analysis was undertaken by two of the authors.

In the choice task, doctors were asked to consider a context where a patient is seeking information, advice and possibly a prescription for contraception. As GPs were considered to be knowledgeable about the attributes of contraceptive products on the Australian market, the choice task did not specify the product attributes apart from in terms of a label. Doctors were asked to consider a series of hypothetical patient encounters described in terms of the characteristics of the woman (her health, her life stage and contraceptive experience and her smoking and socioeconomic status), and then to consider which products they would discuss with the woman, and which specific product they would recommend. Each woman patient is described by a set of attributes that form the experimental design. The final set of attributes and levels are provided in Table 1.

==Table 1 about here==

It is only feasible to show a subset of the $(4^4 x 3^6 x 2^2)$ =746,496 possible "women" to the GPs. As we wanted to allow for potential interactions between age and fertility plans (each with 4 levels) we needed to construct an attribute with 4x4=16 levels. Then the 15 degrees of freedom associated with this attribute would correspond to 3 degrees of freedom for the main effects of each of the attributes "age" and "fertility plans" and 9 degrees of freedom corresponding to the interaction between these two factors. We needed to construct an attribute with 12 levels to estimate the interaction between "periods" and "reason for encounter". Kuhfeld (2006) contains no design with two factors, one with 16 levels and one with 12 levels. A standard construction method in this case (see for example, Construction 2.3.8 in Street and Burgess (2007)) is to choose one of the factors with 4 levels in the design with 64 runs. The whole design is repeated three times but with different names for the levels of that one factor in each of the repetitions of the design. So there are 4 levels in each of 3 designs giving a factor with 3x4=12 levels in total. Thus, 192 "women" in total are divided randomly into 12 versions of 16 women each. The design did not exclude "implausible women" but attribute descriptions made it less likely that such women would be included.

Doctors were asked to answer a sequence of three questions pertaining to each particular patient. A stylized version of the choice task providing the three questions and an example illustrating a possible sequence of responses is given in Figure 1. The predominant method of contraception amongst Australian women is a form of the contraceptive pill. Thus, in Question 1 doctors were asked to decide whether they would confine their discussions of contraceptive options according to three pre-specified and broadly defined sets of products: (i) contraceptive pills only; (ii) methods other than contraceptive pills; or (iii) contraceptive pills and other methods. Then contingent on the broad category they chose, Question 2 required the GP to indicate which specific contraceptive products would form the consideration set to be discussed with the patient. The third and final question required the GP to choose one product that they would recommend as best suited to the patient. Again this third task was contingent on the preceding choice thus potentially constraining the alternatives presented to respondents depending on the outcomes of the previous two questions and making it impossible for respondents to make inconsistent choices in answering the sequence of questions.

===Figure 1 about here===

The products are identified by labels and, as mentioned above, attributes of products are not specified as part of the experiment. The nine products that were considered are the combined pill, the mini-pill, hormonal injection, hormonal implant, hormonal intra-uterine device (hormonal-IUD), hormonal patch, vaginal ring, copper IUD device and condoms. Our emphasis was on prescribed products but it was thought to be more realistic to include condoms as part of the list of products that would likely be discussed, particularly given that doctors may discuss these in addition to a prescribed product. The vaginal ring was relatively new to the Australian market and the hormonal patch was not available at all.

The choice task was completed on-line. The sample frame for the GPs was a list of 14,816 GPs from all states and territories of Australia estimated to be approximately 81% of all recognised GPs currently practicing in Australia. The list was randomised and 1,834 GPs were approached by a phone call and follow-up fax inviting them to participate in the study. 1,512 responded and 177 agreed to participate. Because of the low response rate the sample was augmented through advertising in GP newsletters and forums and a further 44 GPs volunteered to participate. 162 GPs completed the study between December 2008 and June 2009, 22 of whom were volunteers. GP participants were offered $A100 remuneration for their time, paid on completion of the choice tasks.

As has been found in other studies, the response rate for GPs was low (Britt el al., 2008). We therefore compared characteristics of our GP sample to the Australian population of GPs and other national samples of GPs, to gauge the representativeness of the sample (Britt el al., 2008). In terms of age, location and practice characteristics the sample was representative of Australian GPs. The proportion of female GPs however was higher, reflecting the likelihood that this sample was more engaged in reproductive health than other GPs.

4. Econometric framework

The econometric analysis proceeds within the framework of a random utility model where the utility that doctor i derives from recommending alternative j in choice scenario s is given by

(1) $U_{isj} = X'_{is}\beta_j + \varepsilon_{isj}; i = 1, \cdots, N; j = 1, \cdots, J; s = 1, \cdots, S.$

The $K \times 1$ vector of explanatory variables X_{is} contains alternative specific constants (ASCs) that here represent the different contraceptive products, attributes of the hypothetical women and observed characteristics of our sample of GPs. Thus the associated vector of coefficients denoted by β measure GPs relative valuation of alternative products and how these change with different attributes of women and characteristics of GPs. These relative valuations then lead to the resultant recommendations. Note that we are not necessarily assuming that GPs are making recommendations to maximize the utility of their patients in which case only the attributes of the women should matter. The presence of GP effects, conditional on patient characteristics, are required in order to capture the differential abilities of GPs to match women with products and allows for the possibility that prescribing different products may have different financial implications for GPs.

Because of the structure of the data no variables vary over alternatives. The design variables depicting different women vary over doctors and scenarios but are context variables that do not vary over contraceptive products. Similarly, the characteristics of the doctors will only vary over i and not over either j or s. Consequently the associated vector of coefficients does need to vary over j. Using (1) and assuming the disturbance terms ε_{isj} to be identically and independently distributed as extreme value, the standard multinomial logit (MNL) specification results.

In the description of the model thus far we have implicitly assumed an exogenously given set of J alternatives that does not vary over individuals. In fact having different choice sets being considered by different individuals complicates notation but as long as the variation is exogenous it does not substantially change the estimation procedures and is routinely accommodated in standard software packages such as Stata.

For example, situations where there are a large number of possible options from which to choose are easily accommodated in the context of a SP choice task. In related work Fiebig et al. (2011) investigated the preferences of women when choosing a contraceptive product from a similar list of options to those used here. In that case women were presented with choice sets of three products that were chosen as part of the design and so the consideration set is legitimately treated as exogenously given. In the current choice task, doctors have been forced to make their final recommendation in a structured manner. Doctors actually make a recommendation from the alternatives that they have indicated as belonging to the consideration set to be discussed with the patient; again see Figure 1 for a particular example of a doctor choosing to recommend the hormonal IUD in preference to condoms. Now it is much less obvious that the consideration set can be treated as exogenously given.

Because of the potential problems identified by Horowitz and Louviere (1995) and Carson and Louviere (2006) we do not advise conditioning on these consideration sets in estimating models of recommendations. Instead, we follow the argument of Horowitz and Louviere (1995) that a consideration set is simply a reflection of the underlying preferences. Implementation of this maintained hypothesis requires the estimation of choice models using choice sets comprising the full set of contraceptive products ignoring the information on the constrained choice sets available in our data. Geweke et al. (2004) is a prominent example where the same approach is used. Once again using Figure 1 as an illustrative example, for this particular observation we do not impose the consideration set comprising the two products in estimation and instead assume that this doctor is making the recommendation as if choosing from the full set of products.

Simplicity of estimation and interpretation are among the main advantages of MNL but these come at the cost of some restrictive assumptions that are clearly unrealistic in our choice context. Here we adopt the generalized multinomial logit (G-MNL) framework recently developed in Fiebig et al. (2010). This includes as a special case the popular mixed logit (MXL) specification; see for example Revelt and Train (1998), Brownstone and Train (1999), Hall et al. (2006) and Hole (2008). G-MNL provides an appealing and tractable way to extend basic multinomial choice models. In addition to allowing for random parameters which is a natural way to represent preference heterogeneity, G-MNL allows for individual scale heterogeneity. The work of Fiebig et al. (2010) highlights the empirical importance of accommodating this extra dimension of heterogeneity. Further research by Keane and Wasi (2013) demonstrates the flexibility of G-MNL in comparison with latent class models.

The full G-MNL model provides flexibility in how scale and taste heterogeneity are combined but can involve a large number of parameters especially when there is a large number of alternatives as there is here. As such, a relatively parsimonious specification of the G-MNL model is proposed. In the current application the structure of our data means it is unlikely that we would be able to estimate models including substantial preference heterogeneity given there is no attribute variation over alternatives. The specification chosen to form our base model is given by:

$$(2)\ U_{isj} = \left(\alpha_j + z'_i\gamma_j + \eta_{ij}\right) + X'_{is}(\beta_j\sigma_i) + \varepsilon_{isj}; i = 1,\cdots,N; j = 1,\cdots,J; s = 1,\cdots,S;$$

where X_{is} is now interpreted as including only design attributes describing the patients and is exclusive of the ASCs representing the different contraceptive products. Heterogeneity in the GP valuations of these two distinct sets of variables is treated separately. Allowing for scale heterogeneity, denoted by σ_i, indicates that (2) is observationally equivalent to a particular type of preference heterogeneity in the utility weights associated with patient attributes, an observation that has led Louviere et al. (2008) to be critical of the standard MXL model. On the other hand the ASCs are allowed to vary both deterministically with doctor specific characteristics denoted by z_i with corresponding weights γ_j and randomly according to an error η_{ij}. It seems appropriate to have a more flexible heterogeneity structure for how GPs value products than for how they evaluate the attributes of patients. Also, because differential substitution patterns are expected across the contraceptive products the ASCs are assumed to be correlated. Specifically the random errors of the

ASCs are assumed to have a multivariate normal distribution with zero mean and covariance matrix Σ. Thus the α_j parameters are the conditional mean valuations of the contraceptive products when doctor specific characteristics are set to zero. Assuming random effects to be individual specific induces correlation across choice situations, accounting for the dependence structure in unobserved utility among the repeated choices of a doctor which comes from the panel structure of the data.

One attraction of a model of scale heterogeneity, as is specified below, is a considerable amount of flexibility with the addition of only one parameter. Because σ_i is required to be positive we specify an exponential transformation given by:

$$(3)\ \sigma_i = \exp(\bar{\sigma} + \tau \upsilon_i)$$

where $\upsilon_i \sim N(0,1)$ and $\bar{\sigma}$ is a normalizing constant required to ensure identification of β. The additional parameter τ provides a measure of scale heterogeneity. If τ =0, the G-MNL model reduces to a standard MXL specification with random ASCs. Fiebig et al. (2010) argue that ASCs are fundamentally different from most observed attributes and as such recommend that they not be subject to scaling.

This relatively parsimonious representation of heterogeneity seems reasonable given our particular data structure while emphasizing the most important potential sources of heterogeneity. In what follows we simply refer to (2) with the associated assumptions as the G-MNL model. The models were estimated by maximum simulated likelihood using an extended version of the mixed logit procedure developed for Stata by Hole (2007); see Gu et al. (2013) for more detail.

5. Data and summary statistics

The first two questions in the choice task relate to the existence and composition of consideration sets and given our focus in this paper only a very simple descriptive analysis of these data are provided. Answering the first question doctors indicate which of three broad product categories they would discuss with a specific patient. The raw frequencies across all 2592 choice occasions indicate that doctors will sometimes confine their discussions to "pills only" (3%) or "methods other than pills" (22%) but in the vast majority of cases (75%) they consider a mix of "pills and other methods". While this initial stage highlighted the pills/other method distinction, for the second question doctors were in fact asked to indicate which particular products they would actually discuss and so even those who chose "pills and other methods" in their first question subsequently were given the opportunity to indicate a consideration set comprising a subset of products. Over all choice occasions, the median and modal number of products discussed (including the condom) is 4 and in less than 1% of choice occasions did doctors indicate they would discuss all products. These basic results indicate the existence of consideration sets whereby doctors almost always discuss a subset of available contraceptive products with patients.

The first column of Table 2 provides the relative frequencies with which each of the products appeared in a consideration set, the outcome of Question 2. For comparison, this table also provides the relative frequencies of the choices made in the third question, i.e. those products that were

ultimately recommended by the doctors. These recommendations are the outcomes that are the primary focus of our analysis.

Across all 2592 observations the hormonal patch was rarely considered and was only recommended on 0.5% or 13 occasions. This is possibly not surprising given its unavailability in Australia leading to unwillingness on the part of doctors to choose this alternative even in a hypothetical setting where the product is assumed to be available. In any event, it is difficult to estimate choice models with such an infrequently chosen alternative and hence these observations have been dropped in the following analyses. (An alternative approach would be to impute a choice from the consideration set indicated by doctors. This has not been pursued and is unlikely to lead to big differences in results given the small number of observations involved.)

A number of GP characteristics were used in the analyses. They are the doctor's age, gender, whether they were a Fellow of the Royal Australian College of GPs (RACGP), whether they have a Family Planning Certificate, whether they graduated from an Australian medical program, whether their practice is located in an urban area and whether they bulk billed implying patients face no out-of-pocket expenses. Table 3 provides some summary statistics on these key characteristics of the doctors in the sample.

==Tables 2 & 3 about here==

6. Estimation results
6.1 Overview
The base model given by (2) is estimated using the full set of attributes and with the doctor characteristics listed in Table 3 included as interactions with the ASCs. Presenting the estimation results efficiently and informatively presents a major challenge. As previously noted the socio-demographic characteristics of respondents (the doctors), and the attributes of the women do not vary over the alternatives necessitating a specification where their associated coefficients need to vary over the alternatives, here the products.

Thus both MNL and G-MNL models involve the estimation of 34 parameters for each of 7 products (for identification one set of product parameters is normalized to zero). In our specification of the G-MNL model there are 29 additional parameters to be estimated. A majority of these additional parameters associated with the G-MNL model are individually significantly different from zero at the 1% level. In particular the τ parameter that governs the extent of scale heterogeneity is estimated to be 0.164, which is relatively small, but is precisely estimated with an asymptotic standard error of 0.031. Relative to initial MNL estimation there is a dramatic improvement in fit as expected. The associated likelihood ratio test statistic is 644.7 with a $p < 0.0001$. All of this evidence indicates the appropriateness of the more flexible G-MNL model that has been estimated.

The full set of estimation results is not presented but is available on request. Instead a range of graphical summaries are provided in order to depict the key features of the results. The primary approach involves using our estimated G-MNL model to predict product recommendations under a

range of different scenarios and then to display the results as boxplots for all eight products. For example, the boxplots in Figure 2 display the prediction results for all sample observations for each of the contraceptive products. Thus the distribution of predictions for each product is characterized and key summary statistics such as the maximum, minimum, and median predictions can be visualized and compared across products.

==Figure 2 about here==

Using Figure 2, the first thing to note is that the median predictions for each of the products are similar to the raw sample frequencies of product recommendations reported in Table 2. These do not reflect actual contraceptive use across the population of Australian women because our design aims to provide a range of different women and for the purposes of estimation will naturally over-represent women with less common combinations of attributes. But in addition to the central tendencies, these predictions provide some key insights related to variation in choices. The minimum values indicate there are woman/doctor combinations that will lead to predictions for each product being essentially zero. The hormonal implant has the largest minimum prediction of only 1.0%. Turning to the maximum values, the combined pill and hormonal IUD stand out. For these two products there are woman/doctor combinations where the predicted probability of being recommended is approaching unity. Conversely, for the remaining products the predicted probability can be very high but still much less than the probabilities of greater than 85% for the combined pill and hormonal IUD. The hormonal implant is noteworthy amongst these remaining products because it has the highest median predicted probability; as shown in Table 2 it was the most frequently recommended product.

6.2 Impact of doctor characteristics

In an attempt to separate the impact of woman and doctor characteristics, Figure 3 replicates the process of producing Figure 2 but this time summarizes predictions for each of the distinct 192 women produced by the experimental design by averaging over the doctor's characteristics at each of these design points. As expected the range in the predictions has been reduced but in two cases these reductions were especially large. For the hormonal injection the maximum prediction has dropped from 53% to 30% and for the copper IUD the drop was from 42% to 20%. For these two products there is clearly a large doctor effect.

==Figure 3 about here==

The impact of doctor characteristics is further isolated in Figure 4 where the estimated coefficients are summarized. Each row provides the estimated coefficients for each of the eight products with the combined pill being the base and hence always having a coefficient of zero. In this format it is easy to determine contrasts between any pair of products for a given variable by finding the difference between the associated product coefficients for the row associated with the particular variable. For example, the older the GP the more likely they are to recommend a copper IUD compared with the combined pill and even more likely to recommend the IUD over condoms. GPs that bulkbill are also more likely to recommend the IUD compared to any other product. Conversely

GPs who have a Family Planning Certificate are relatively less likely to recommend the copper IUD and relatively more likely to recommend the hormonal IUD. Female GPs, those who are Fellows and to a somewhat lesser extent those with a Family Planning Certificate are all associated with being relatively more receptive to the new product, the vaginal ring. The same three characteristics are also associated with being more inclined to recommend the hormonal IUD. Other large effects are that urban GPs are relatively less likely to recommend an IUD or implant but are relatively more likely to recommend condoms.

The fact that the normalized combined pill coefficients and the estimated mini-pill coefficients tend to be very central for each GP characteristic indicates that these do not play a big role in determining the choice of pill products compared to other products. Another way to see this is to refer back to Figures 2 and 3 where the distributions for the two pills are not overly sensitive to controlling for overall GP effects.

Because the actual estimated coefficients do not easily convert into impacts on probabilities, Figure 5 illustrates an example of the impact of one of the key GP characteristics. This figure summarizes predictions from the estimated G-MNL model where the doctors are separated by whether they have a Family Planning Certificate or not. We see that the coefficients that were large in magnitude do translate into large impacts on predicted probabilities. Those with Family Planning Certificates are much less likely to recommend the injection and copper IUDs and much more likely to recommend the vaginal ring, hormonal implant and hormonal IUD.

==Figures 4 & 5 about here==

6.3 Impact of women's characteristics
In order to summarize the impact of the women's characteristics, we seek to answer the following questions: (1) For what type of women are predictions clear? And; (2) for what type of women is there considerable uncertainty about what to recommend? As a measure of uncertainty, the predicted entropy is calculated for the range of probabilities for each of the distinct 192 women defined by the design. As before, we are controlling for doctor effects and concentrating on the differences related to the characteristics of the women. The entropy is defined as:

$$(4)\ E_i = -\sum_{j=1}^{J} \hat{p}_{ij}\log(\hat{p}_{ij})$$

The entropy reaches its minimum of zero when one of the probabilities equals unity while it attains its maximum value of $\log(J)=\log(8)$ when the predicted probabilities follow a uniform distribution. Thus by finding the design points associated with the smallest and largest entropy estimates it is possible to summarize the characteristics of the associated women.

As expected from Figure 2, the lowest entropy estimates, those that are most certain, are associated with women who our model predicts are very likely to be recommended either the combined pill or the hormonal IUD. Of the 192 within-sample women, the 29 (15%) observations with the lowest

entropy have either of these products with predicted probabilities in excess of 48%. What can we say about these women? Amongst these 29 women there were 19 where the large prediction coincided with the combined pill and 10 where it was the hormonal IUD. Each of these groups is summarized by the boxplots of the predicted probabilities in Figures 6 and 7.

==Figures 6 & 7 about here==

For this first "combined pill" group there are a number of characteristics that matter. The women are younger (15 of the 19 are in the two youngest categories meaning they are aged less than 30), almost half are starting prescribed contraception for the first time, most have normal blood pressure and none have elevated blood pressure, most have children but are not breastfeeding, only one has problems with compliance and most are non-smokers.

For the second "hormonal IUD" group the women are older (there are none in the youngest category), all have problems with periods, only 2 of the 10 have normal blood pressure, 8 of the 10 have children but are not breastfeeding, most don't want to have children in the future, most have problems with compliance and most are smokers. Consider one particular woman within this group whose characteristics include high blood pressure and being a smoker. Good clinical practice dictates that a contractive pill is not recommended for this woman and that is what we find. The average predicted probability for the combined pill is 0.0007 and for the mini-pill 0.0010. For this woman, GPs exhibit a strong preference for the hormonal IUD (average predicted probability of 0.65) after essentially eliminating pills.

At the other extreme, maximum entropy occurs when the predicted probability is spread relatively evenly across products. For example, the maximum average entropy for any particular woman is 1.96 and occurs when the predicted probabilities range from 6% for the copper IUD to 19% for the hormonal IUD and seven of the eight products have predicted probabilities in excess of 10%. The left hand panel of Figure 8 summarizes the predicted probabilities for the 20 women with the highest average estimated entropy values. The relatively high predicted probability for the hormonal implant in this subset of women is consistent with the earlier observation that this is a popular compromise product especially in "difficult" cases.

==Figure 8 about here==

The women associated with the largest entropy values are harder to categorize but of the 20 with the largest entropy, only one had normal blood pressure and most planned to have children in the next 2 years. This suggests that blood pressure problems and shorter term fertility plans present particular challenges for GPs in deciding on a suitable product either because there is no perfect match or because there may be several equally good matches.

Note that these results have been averaged over our sample of GPs. It is especially interesting to explore how our GPs react to this particular group of difficult women. The right hand panel of Figure 8 provides the distributions of predictions over the 268 distinct GP-women observations for this

group of 20 women. As in the previous comparison of Figures 2 and 3 the differences in the two panels of Figure 8 are attributable to GP effects for this subset of women.

Unlike with the full sample comparison, there are marked differences across all products for this group of women. It remains true, that the changes for the hormonal injection and the copper IUD are relatively large. But there are also dramatic changes for the implant, hormonal IUD and ring. For this group of difficult woman there are relatively large GP effects.

6.4 Diagnostics

The modelling of recommendations was predicated on an assumption that the formation of a consideration set is simply a reflection of underlying preferences. Before proceeding to a discussion of the implications of these results we provide some regression diagnostics aimed at indicating whether our working hypothesis is reasonable or not.

The working hypothesis implies that no extra information is gained from knowledge of consideration sets. Thus in generating the diagnostics we exploit the fact that while modelling recommendations was undertaken independently of the consideration set data, this information is available for testing purposes and is used in a number of ways.

In addition to the recommendation model that has been estimated ignoring consideration set information, consider a second trinomial choice where the outcomes are the initial choice between the three broadly defined product groups; "Pill", "Not Pill" and "All". Do predictions generated by the recommendation model have predictive power in the product group model and vice versa? Evidence consistent with our maintained hypothesis requires that recommendation predictions have predictive power in the product group regressions but the converse is not true. These diagnostics were generated using MNL. This decision was based primarily on computational convenience and while we acknowledge that the results could be sensitive to the simplification, there is no a priori reason why this should be the case.

In particular, use the recommendation model to generate predictions for all products. Now construct a variable that is the sum of the predictions for the products in the "Not Pill" category and include it as an additional covariate in the product group regression. In this estimation, the Wald test for the null hypothesis that the additional variable has no explanatory power generates a test statistic of 18.23 which when compared to a chi-squared distribution with 2 degrees of freedom produces a p=0.0001. Note that this estimation controls for the same attributes and doctor characteristics used in the recommendation estimation but even after controlling for these, the recommendation predictions have strong explanatory power helping to explain product group outcomes.

Reversing the procedure, the product group estimation was used to generate predicted probabilities for the "Pill" and "Not Pill" groups. These were then added to the recommendation model to produce a Wald test statistic of 9.69 which when compared to a chi-squared distribution with 14 degrees of freedom produces a p=0.78. Consistent with our maintained hypothesis, after controlling

for the attributes of the women and the characteristics of the doctors, predictions of broad product groups have no additional explanatory power in the recommendation model.

For the second type of diagnostic, one might consider formally testing for parameter constancy in the recommendation model across the three subsets of observations formed by splitting the sample according to which of the three broad product sets that were chosen. However, this is problematic because of the very small size of the "Pill" subsample and because our design would potentially be seriously compromised leading to the lack of identification of parameters.

Instead a less formal approach is taken and the coefficient estimates and predictions from the full recommendations model are compared with those generated for the large "All" subsample The simple correlation between the two sets of estimated coefficients is 0.99. Using the two sets of results we generate product predictions and isolate the prediction associated with the actual recommendation. The correlation of the two sets of predictions across the entire sample is 0.97 and even if we restrict attention to the out of sample observations not used in the "All" subsample estimation, the correlation is still 0.96. Both sets of estimation results are generating substantively similar predictions for the actual recommendations.

7. Discussion of results

The results presented in Section 6 reveal considerable variation in recommendations associated with patient characteristics. This is not totally unexpected within the context of a SP choice task where the aim is to provide attributes with levels that encourage respondents to make tradeoffs that result in different choices without compromising the external validity of the results by presenting scenarios that lack credibility. While there seems to be considerable agreement amongst GPs about who should be recommended the combined pill and the hormonal IUD, no such clear agreement emerges for the other products. However, the hormonal implant does represent an interesting case. While it is recommended on more choice occasions than any other product, it does not reach the levels of certainty displayed by the combined pill and the hormonal IUD. One interpretation of this finding is that in addition to representing the appropriate contraceptive option for a range of women, the implant plays the role of a compromise product in many cases, possibly because of greater awareness through marketing; see Windmeijer et al. (2006). Unfortunately we do not have relevant GP characteristics that would enable us to test this hypothesis.

On the other hand there are cases where there is considerable variability as to what product to recommend even after controlling for differences in GPs. It is possible that for certain types of women with complex medical histories there is a range of contraceptive options that GPs view as being equally satisfactory. However, the fact that the maximum entropy situation was associated with predicted probability spread relatively evenly across all products suggests that this is only a partial explanation. We interpret this wide spread of predicted probabilities as evidence of clinical ambiguity that derives from uncertainty relating initially to the process of interpreting the women's medical history and then subsequently to the choice of an appropriate contraceptive method; see Wolff (1989).

While patient characteristics are important in determining what GPs recommend it is also true that GP-specific factors also play a role. After controlling for patient characteristics, there is considerable variation in recommendation patterns attributable to the GP. Such variation has previously been associated with lost opportunities to improve the health of women and to reduce health costs. Sonnenberg et al. (2004) estimate the net impact of contraception on women's health in the US and conclude that:

> "Increasing the use of more effective methods even modestly at the expense of less effective methods will improve health and reduce costs".

Similarly in a review of sexual health in the UK, Armstrong and Donaldson (2005) used consensus methods to derive ideal profiles of contraception provision and compared these with existing profiles to conclude that:

> "Generally, there is support for a change in the profile of methods supplied, mostly from combined hormonal oral methods to longer acting methods and that such a move would produce a very large cost saving to the NHS."

These longer acting methods include the hormonal injection, hormonal implant, hormonal IUD and copper IUD. Wellings et al. (2007) explores why the diffusion of long-acting reversible methods has been slow in the UK and presents evidence suggesting that it is due to a combination of lack of confidence and experience and misconceptions about side-effects. Our results are supportive of the Wellings et al. (2007) findings. It is reasonable to assume that GPs who have a Family Planning Certificate are better informed about and more experienced prescribing contraception. In our results they were more inclined to recommend longer acting methods.

Again if we think in terms of extremes, products rather than patients, the vaginal ring and the copper IUD stand out. The vaginal ring is the newest of the contraceptive products considered, having been introduced into Australia in 2007. In addition to GPs who have a Family Planning Certificate, those who were female and Fellows of the Royal Australian College were very receptive to recommending this new product. This is as expected as these characteristics can be viewed as measures of quality and engagement in relation to reproductive and sexual health. Also, those trained overseas rather than in Australia were more likely to recommend the ring possibly reflecting previous experience with the ring in countries where adoption occurred earlier than in Australia. At the opposite extreme GPs who have a Family Planning Certificate, were female, Fellows or trained overseas were much less likely to recommend the copper IUD which was favoured by older GPs and those not in urban practices. The copper IUD has been available for much longer than the hormonal IUD and whilst it has valuable attributes, including being hormone-free, it has been largely superseded by the newer hormonal device. This suggests reluctance on the part of some GPs to embrace new technologies and instead to rely on more familiar products. Greater investment in the education and training of doctors in general practice about the range and relative attributes of available contraceptive may be warranted, in particular in relation to the highly effective and cost-effective long acting reversible methods.

Economic incentives whereby health practitioners exploit their position to influence the demand for their services has also been suggested as a potential factor that might impact on medical decision-making and explain medical practice variation; see Davis et al. (2000) and Scott and Shiell (1997). In Australia several pill types, including the older combined pill and mini-pill, the hormonal injection, hormonal implant and hormonal IUD are available on the Pharmaceutical Benefits Scheme at subsidized prices and hence are relatively inexpensive for women. Newer combined pills, the vaginal ring and copper IUD are not subsidized. These price differences might impact on GP recommendations but would be accounted for by the alternative specific constants. There are no other obvious price incentives for GPs that would lead to differential prescribing patterns. However, we do observe a large bulkbilling effect for the copper IUD. As the copper and hormonal IUD are the longest lasting of all contraceptives, GPs who bulkbill may be choosing these methods as the most cost-effective in the long run for their patients especially for lower income women. This would explain the large bulkbilling effect for the copper IUD and is consistent with the positive effect for the hormonal IUD. The differential effect between the two IUDs is consistent with the copper IUD lasting for up to 10 years compared to 5 years for the hormonal IUD. It could also be related to the patient mix associated with practices that bulkbill but we do not have data that would enable us to test this. The introduction of financial incentives for GPs could be considered as a means to encourage greater use of more effective and cost-effective new technologies, in particular the long acting reversible implants and intra-uterine devices.

The identification of the low entropy, hormonal IUD group of women indicates some support for the use of long acting reversible methods. However, these are cases where clear clinical indicators, including problems with periods and blood pressure, necessitate a move away from recommending contraceptive pills. In the other low entropy case where the clear recommendation was the combined pill, there are no clinical reasons why long acting reversible methods shouldn't be recommended. In fact, it could be argued that greater use of longer acting reversible methods will be more beneficial for young women in good health and starting prescribed contraception for the first time who are potentially at most risk of unintended pregnancy.

The group of women associated with high entropy is especially interesting. For these women, particular combinations of factors make recommendations difficult and GP characteristics drive a considerable amount of the variation across and within products. As we have noted there is a tendency for GPs who are female, have a Family Planning Certificate and who are Fellows to have similar recommendation patterns. In particular, these are typically the most important factors determining the recommendation of longer acting reversible methods and these are amongst the products that are most impacted by GP characteristics. As we are associating these characteristics with GPs being better qualified and engaged with contraception we are again led to suggest the need for greater support for education and training in general practice with regard to contraception. Notice that the impact on recommendations of a GP being Australian trained does not closely align with the estimated impacts of these characteristics. For example, Australian trained GPs were less likely to recommend the ring and more likely to recommend the copper IUD. This suggests a gap between general Australian training and more specialized experience and engagement with contraception issues.

Despite the importance of identifying the impact of contraceptive knowledge and experience of GPs on their recommendation choices, we are somewhat limited by our sample size and by the broad proxies we are using. It would be especially useful to have more indicators specific to contraception and family planning knowledge and to be able to determine how much promotional activities of drug companies are influencing Australian GPs.

Evidence has been presented that is broadly supportive of our main hypothesis that when faced with a complex choice problem, the GPs in our data have formed consideration sets on the basis of their underlying evaluations of all of the contraceptive products being considered for a particular patient. The implications of this result are important from a methodological perspective. Even when consideration set information is available, choice need not be modelled as a multi-stage process with consideration preceding choice. While there may be an efficiency loss associated with this approach, under our maintained hypothesis it will not introduce any biases.

In contrast, inappropriate use of consideration set data, ignoring the likely endogeneity of choice sets chosen by respondents, will likely lead to biased inferences. Comparing our results with those generated when the stated products considered were used to define the choice sets for each choice occasion, confirms clear discrepancies between the two sets of estimates. These differences are interpreted as a reflection of selection bias as anticipated by Horowitz and Louviere (1995) and Carson and Louviere (2006).

For those designing SP choice tasks the current analysis suggests that there may be little to be gained by adding extra questions regarding consideration sets if it is the final choice that is of most interest. This is not to say that consideration sets are not formed or are unimportant, simply that they convey limited additional information about preferences over and above that derived from a standard task which simply asks for the preferred choice.

Considerable care needs to be exercised in claiming wide generalisability for our methodological results because of the specific nature of the choice problem being considered. In particular, GPs can be expected to be relatively well informed about the full range of contraceptive products. Hence all contraceptive products presented in the current work can reasonably be expected to belong to the universal set of alternatives considered by GPs. In situations where this is not the case there is clearly a role for modelling the initial awareness stage as emphasized in Belkar et al. (2006). Also non-experts may be more prone to use simplifying heuristics and modelling consideration sets becomes more relevant.

8. Conclusions

By exploiting SP methods we have been able to provide new insights into medical decision-making in the particular case of GP recommendations for contraceptives. These results indicate considerable variation in recommendations. In part this is due to patient characteristics which allow us to identify particular types of women where there is uncertainty or clinical ambiguity about appropriate contraceptives to recommend. But the variation in recommendations is also partly explained by GP specific effects or "styles of practice". In the literature on MPV such effects have been linked to

unnecessarily high surgical risks and expenditure levels. Here they provide one credible explanation for the relatively low uptake of longer acting reversible contraceptives and hence the source for large potential cost savings from reductions in unplanned and unwanted pregnancies. In order to capture these savings it may not be enough to incentivise GPs to prescribe longer acting reversible contraceptives. Instead it is likely that the first policy response should be to encourage and facilitate further education and training of doctors.

Because the contraceptive pill is so entrenched as the default option for many GPs (and women) shifting preferences towards long acting reversibles may also require GPs being instructed in how this product information is presented to women. Campaigns to educate women on the benefits of long acting reversibles may also be warranted. Contraceptive choice is one area where the patient is likely to play a considerable role in the prescription choice and given Australian women can freely choose their GP, more informed women may gravitate to doctors willing to embrace the new technologies.

The results also indicate that in their interactions with patients GPs restrict discussions of contraceptive options to form consideration sets. While it is sensible to contemplate other decision-making models when dealing with complex choice problems, our conclusion is that the assumption that both the formation of consideration sets and ultimate choices are governed by the same preferences is a reasonable working hypothesis. Moreover, our application to the recommendations of contraceptive products by GPs illustrates a viable compromise approach where both considerations sets and final choices can be analysed using SP methods.

In further research it would be extremely useful to extend the analysis with a larger sample of GPs with more contraceptive-specific detail in order to undertake a more nuanced exploration of GP variations. A complementary analysis of women in order to explore their primary sources of information on contraception and, when this is sourced from GPs, what they say GPs discuss with them would also provide a clearer picture of the medical decision-making process in this case.

References

American College of Obstetricians and Gynecologists Committee on Gynecologic Practice Long-Acting Reversible Contraception Working Group. (2009), "ACOG Committee Opinion no. 450: increasing use of contraceptive implants and intrauterine devices to reduce unintended pregnancy", *Obstetrics and Gynecology*, 114, 1434–1438.

Armstrong, N. and Donaldson, C. (2005), *The economics of sexual health*, Family Planning Association.

Belkar, R., Fiebig, D.G., Haas, M. and Viney, R. (2006), "Why worry about awareness in choice problems? Econometric analysis of screening for cervical cancer", *Health Economics*, 15, 33-47.

Britt, H., Miller G.C., Charles, J., Henderson, J., Bayram, C., Harrison, C., Valenti, L., Fahridin, S., Pan, Y. and O'Halloran, J. (2008), *General Practice Activity in Australia 2007-2008*, Australian Institute of Health and Welfare.

Brownstone, D. and Train, K. (1999), "Forecasting new product penetration with flexible substitution patterns", *Journal of Econometrics*, 89, 109-129.

Carson, R.T. and Louviere, J.J. (2006), "Statistical properties of consideration sets", UCSD, Department of Economics, Discussion Paper, 2006-07.

Davis, P., Gribben, B., Scott, A. and Lay-Yee, R. (2000), "The 'supply hypothesis' and medical practice variation in primary care: Testing economic and clinical models of inter-practitioner variation", *Social Science and Medicine* 50, 407-418.

Fiebig, D.G., Keane, M.P., Louviere, J.J. and Wasi, N. (2010), "The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity", *Marketing Science*, 29, 393-421.

Fiebig, D.G., Knox, S.A., Viney, R., Haas, M., and Street, D.J. (2011), "Preferences for new and existing contraceptive products" *Health Economics* 20 (S1), 35-52.

Geweke, J., Gowrisankaran, G. And Town, R. (2004), "Bayesian inference for hospital quality in a selection model", *Econometrica*, 71, 1215-1238.

Gilbride, T.J., and Allenby, G.M. (2004), "A choice model with conjunctive, disjunctive, and compensatory decision rules", *Marketing Science*, 23, 391-406.

Goeree, M.S. (2008), "Limited information and advertising in the U.S. computer industry", *Econometrica*, 76, 1017-1076.

Grytten, J. and Sørensen, R. (2003), "Practice variation and physician-specific effects" *Journal of Health Economics* 22, 403-418.

Gu, Y., Hole, A.R. and Knox, S. (2013), "Estimating the generalized multinomial logit model in Stata", *Stata Journal*, 13, 382-397.

Hall, J.P., King, M.T., Fiebig, D.G., Hossain, I., and Louviere, J.J. (2006), "What influences participation in genetic carrier testing? Results from a discrete choice experiment", *Journal of Health Economics*, 25, 520-537.

Hauser, J.R. and Wernerfelt, B. (1990), "An evaluation cost model of consideration sets", *Journal of Consumer Research*, 16, 393-408.

Hole, A.R. (2007), "Fitting mixed logit models by using maximum simulated likelihood", *Stata Journal*, 7, 388-401.

Hole, A.R. (2008), "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment", *Journal of Health Economics*, 27, 1078-1094.

Horowitz, J.L. and Louviere, J.J. (1995), "What is the role of consideration sets in choice modelling", *International Journal of Research in Marketing*, 12, 39-54.

Keane, M.P. and Wasi, N. (2013), "Comparing alternative models of heterogeneity in consumer choice behaviour", *Journal of Applied Econometrics*, 28, 1018-1045.

Knox, S.A., Viney, R., Street, D., Haas, M., Fiebig, D.G., Weisberg, E. and Bateson, D. (2012), "What's good and bad about contraception? A comparison of the values placed on features of

contraceptive products by women consumers and GP providers", *PharmacoEconomics* 30, 1187-1202.

Kuhfeld, W.F. (2006), *Orthogonal Arrays*, SAS Institute, http://support.sas.com/techsup/technote/ts723.html.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000), *Stated choice methods: Analysis and application*, Cambridge University Press.

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

Manski, C. (1977), "The structure of random utility models", *Theory and Decisions*, 8, 229-254.

Roberts, J.H. and Lattin, J.M. (1991), "Development and testing of a model of consideration set composition", *Journal of Marketing Research*, 28, 429-440.

Scott, A. and Shiell, A. (1997), "Analysing the effect of competition on general practitioners' behaviour using a multilevel modelling framework", *Health Economics* 6, 577-588.

Sonnenberg, F.A., Burkman, R.T., Hagerty, C.G., Speroff, L. and Speroff, T. (2004), "Costs and net health benefits of contraceptive methods", *Contraception* 69, 447-459.

Street, D.J. and Burgess, L. (2007), *The construction of optimal stated choice experiments: theory and methods*, Wiley.

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

Wellings, K., Zhihong, Z., Krentel, A., Barrett, G. and Glasier, A. (2007), "Attitudes towards long-acting reversible methods of contraception in general practice in the UK", *Contraception* 76, 209-214.

Wenneberg, J.E., Barnes, B.A and Zubkoff, M. (1982), "Professional uncertainty and the problem of supplier-induced demand', *Social Science and Medicine* 16, 811-824.

Windmeijer, F., de Laat, E., Douven R., and Mot, E. (2006), "Pharmaceutical promotion and GP prescription behaviour", *Health Economics* 15, 5-18.

Wolff, N. (1989), "Professional uncertainty and physician medical decision-making in a multiple treatment framework" *Social Science and Medicine* 28, 99-107.

Figure 1

Stylised example of a choice task completed by doctors

FIRST SCREEN		
Woman	**Attribute**	**Level**
	Age	Aged 30-39 years
	Reason for encounter	Starting prescribed contraception for first time
	Periods	No problem with periods
	Blood pressure	Has low blood pressure
	Relationship	In new relationship
	Children	Has no children
	Fertility plans	Unsure about future fertility plans
	Pill preference	Has no strong opinion about pill
	Weight concern	Is concerned about gaining weight
	Compliance	Has no difficulty with compliance
	Income	Has a high income
	Smoking	Is a non-smoker
Choice task	**Question 1**	**Options**
	What would you discuss?	**(Check one)**
		○ Pills only
		○ Pills and other methods
		● Other methods but not pills

SECOND SCREEN		
Choice task	**Question 2**	**Options**
	Please select the appropriate set of products you would discuss?	**(Check as many as appropriate)**
		○ Hormonal injection
		○ Hormonal implant ○ Vaginal ring
		● Hormonal IUD ○ Copper IUD
		○ Hormonal patch ● Condoms

THIRD SCREEN		
Choice task	**Question 3**	**Options**
	Which ONE of these would you most likely recommend?	**(Check one)**
		● Hormonal IUD ○ Condoms

* Options displayed on the second screen are conditional on the answer to Question 1. The display is as it would appear given the answer was "Other methods but not pills". Thus the Combined Pill and Mini-pill do not appear on the second screen. If instead the GP answered "Pills only" then only the Combined Pill and Mini-pill would appear. If the GP answered "Pills and other methods" then the full list of products would appear. Similarly the choices for Question 3 answered on the third screen were restricted to options checked in Question 2.
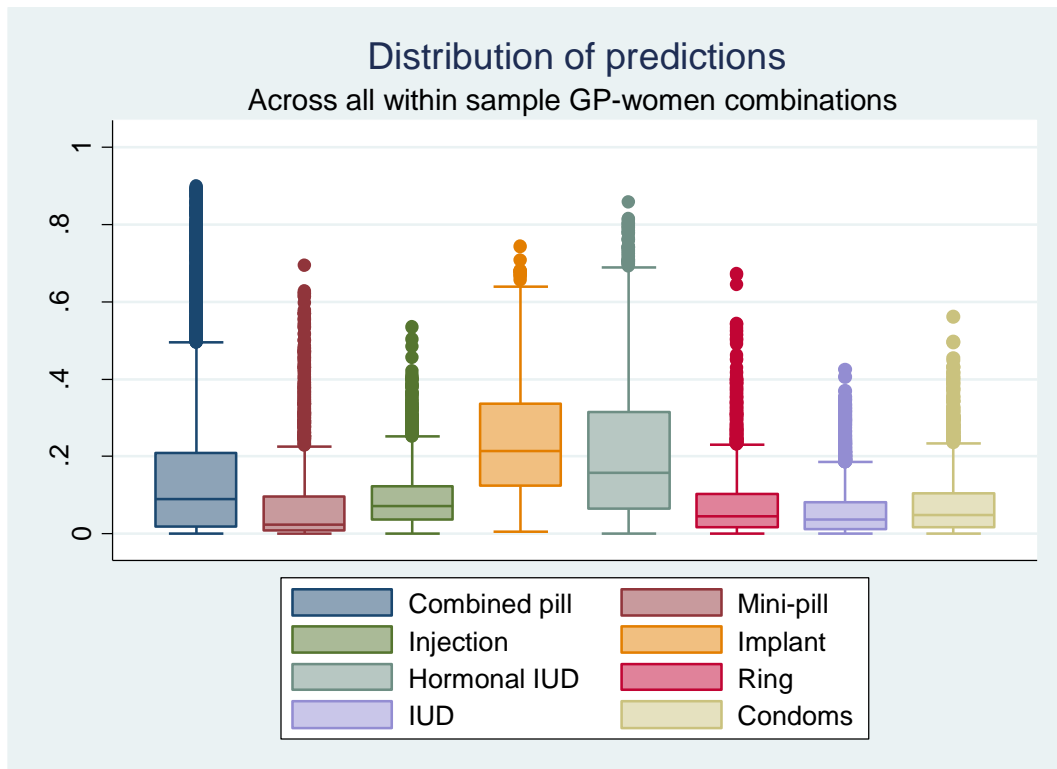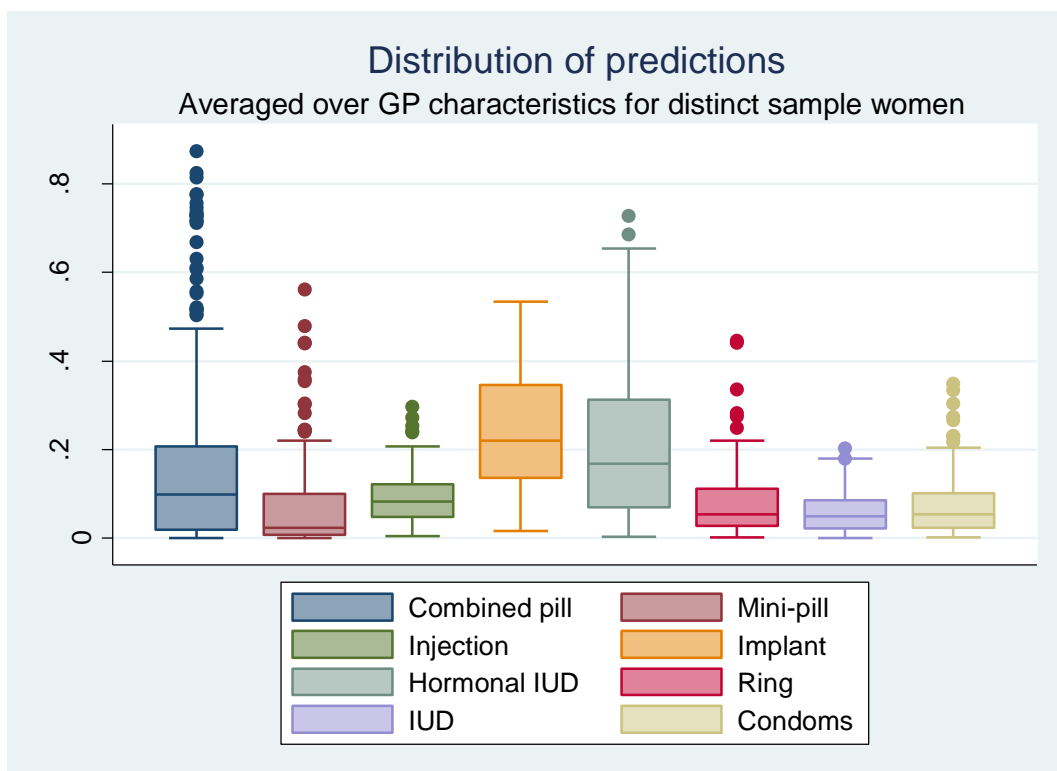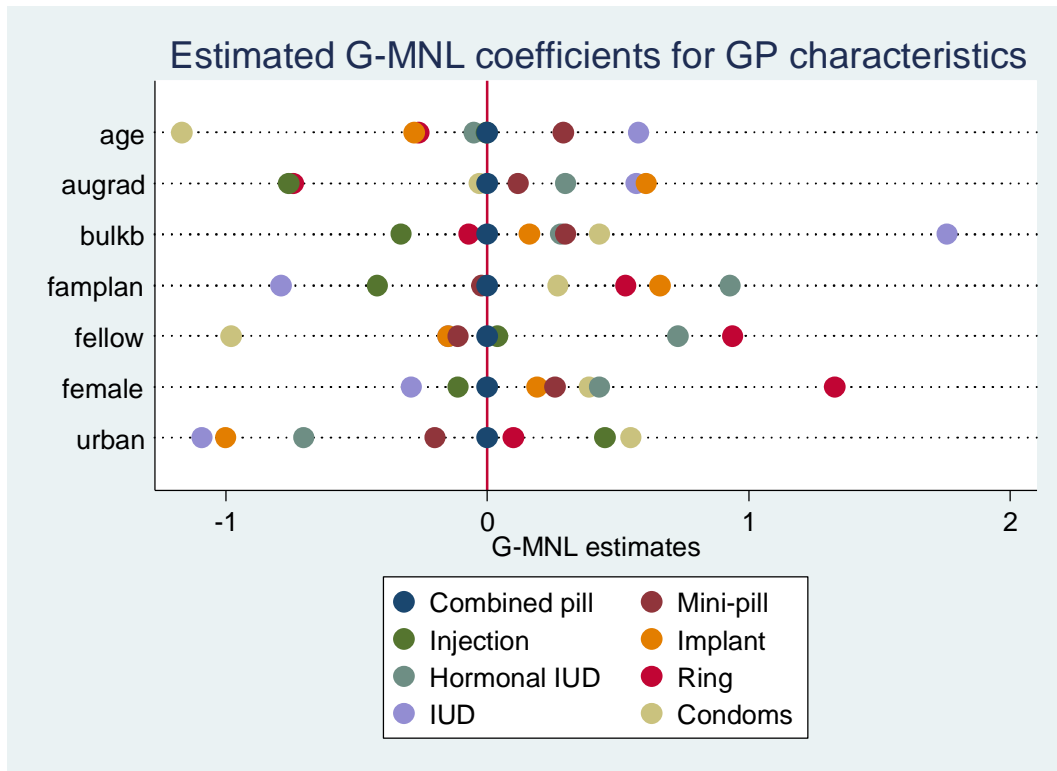
Figure 2

Distribution of predictions
Across all within sample GP-women combinations

Legend: Combined pill, Mini-pill, Injection, Implant, Hormonal IUD, Ring, IUD, Condoms

Figure 3

Distribution of predictions
Averaged over GP characteristics for distinct sample women

Legend: Combined pill, Mini-pill, Injection, Implant, Hormonal IUD, Ring, IUD, Condoms

Figure 4

Estimated G-MNL coefficients for GP characteristics

Figure 5

Family planning effect on predicted probability

Figure 6

Low entropy high Combined Pill women
Averaged over GP characteristics

Figure 7

Low entropy high Hormonal IUD women
Averaged over GP characteristics

Figure 8

Table 1: Attributes and their levels

Age	
dagegp 1	Aged 16-19 years
dagegp2	Aged 20-29 years
dagegp3	Aged 30-39 years
dagegp4	Aged 40 years or more
Reason for encounter	
drfe1	Starting prescribed contraception for first time
drfe2	Recommencing prescribed contraception
drfe3	On pill but dissatisfied
drfe4	Using non-pill method but dissatisfied
Periods	
dbleed1	Heavy and/or painful periods
dbleed 2	Irregular periods
dbleed 3	No problems with periods
Blood pressure	
dbp1	Has low blood pressure
dbp2	Has normal blood pressure
dbp3x	Elevated blood pressure
Relationship	
drel1	In long-standing relationship
drel2	In new relationship
drel3	Has no steady relationship
drel4	No information about relationship
Children	
dchild1	Is currently breastfeeding
dchild2	Has children but is not breast-feeding
dchild3	Has no children
Fertility plans	
dfut1	Does not want to have children in future
dfut2	Plans to have children in next 2 years
dfut3	Plans to have children but not in next 2 years
dfut4	Unsure about future fertility plans
Pill preference	
dpil1	Prefers pill to other methods
dpil2	Has no strong opinion about pill
dpil3	Prefers methods other than pill
Weight concern	
dwt1	Is concerned about gaining weight
dwt2	Is not concerned about gaining weight
Compliance	
dcomp1	Has no difficulty with compliance
dcomp2	Has difficulty with compliance
Income	
dpay1	Has a low to middle household income
dpay2	Has a health care card
dpay3	Has a high household income
Smoking	
dsmk1	Is a non-smoker
dsmk2	Smokes less than 10 cigarettes per day
dsmk3	Smokes 10 or more cigarettes per day

Key: The table shows the dummy variables associated with each level of each attribute. Shaded levels represent those that are omitted for estimation and so define a particular base woman.

Table 2: Relative frequencies of contraceptive product consideration and recommendation

Product	Considered	Recommended
Combined pill	0.532	0.192
Mini-pill	0.365	0.076
Hormonal injection	0.549	0.093
Hormonal implant	0.746	0.240
Hormonal IUD	0.564	0.216
Hormonal patch	0.071	0.005
Vaginal ring	0.439	0.070
Copper IUD	0.251	0.043
Condom	0.646	0.064

Table 3: GP Characteristics*

Characteristic	Mean
Age (years)	47.0
Female	0.642
Fellow RACGP	0.605
Family Planning Certificate	0.327
Bulk bill all patients	0.278
Australian graduate	0.759
Urban practice	0.778

* Apart from age, all other variables are binary dummy variables indicating the presence of the characteristic. In the econometric models age is standardized by subtracting the mean of 47 and dividing by two standard deviations.