HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 15/14

Complexity and doctor choices when discussing contraceptives

Denzil G. Fiebig, Rosalie Viney, Marion Haas, Stephanie Knox,
Deborah Street, Edith Weisberg & Deborah Bateson

September 2015

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

COMPLEXITY AND DOCTOR CHOICES WHEN DISCUSSING CONTRACEPTIVES

Denzil G. Fiebig[1*], Rosalie Viney[2], Marion Haas[2], Stephanie Knox[3], Deborah Street[4], Edith Weisberg[5,6], and Deborah Bateson[5,6]

[1]School of Economics, University of New South Wales
[2]CHERE, University of Technology Sydney
[3]National Centre for Immunisation Research and Surveillance, The Children's Hospital Westmead
[4]Department of Mathematical Sciences, University of Technology Sydney
[5]Family Planning NSW
[6]Department of Obstetrics, Gynaecology and Neonatology, University of Sydney, Australia

PRELIMINARY DRAFT
June 23, 2015

Abstract
In order to better understand choice behaviour, econometric models need to be able to reflect the complexity of decisions that individuals routinely face. We investigate the role of choice complexity in modelling medical decision-making in the case of a doctor choosing which specific contraceptive products to discuss with their patient before ultimately making a recommendation. Clinical vignettes describing patients, developed using stated preference methods, are presented to a sample of Australian general practitioners. An econometric model is developed that captures two salient sources of complexity. The first is associated with patients with particular combinations of clinical and demographic attributes that induce uncertainty around what product to recommend while the second captures variation in the ability of doctors to find appropriate patient-product matches. We are especially interested in the tendencies of doctors to discuss long-acting reversible contraception (LARC) in order to determine whether part of the explanation for the relatively low uptake of LARC in Australia is reluctance on the part of some doctors to even discuss these products.

Key Terms: Choice models; complex decisions; medical decision making; long-acting reversible contraception; clinical vignettes.

JEL: I10; J13; C25; C81

1. Introduction

Individuals routinely face complex decisions and the source of complexity can derive from different aspects of the choice process. There may be a large choice set (Frank and Lamiraud, 2009); alternatives in the choice set may be difficult to evaluate and compare (Sándor and Franses, 2009); the choice context may be unfamiliar (Swait and Adamowicz, 2001). In such situations, decisions are often made with the assistance of a better informed expert. Water heaters are products that are purchased infrequently and often with some urgency and so plumbers routinely provide advice to consumers on the type of water heater to install (Bartels et al. 2006.) Financial advisors are one source of information that has the potential to compensate for poor financial literacy that is a likely cause of households making suboptimal decisions in complex financial choices such as retirement portfolios (Bateman et al. 2014). Our primary objective is to provide insights into medical decision-making in the case of doctors advising patients on the choice of prescribed contraceptive products.

Interactions between a decision maker and an expert can be characterized by an initial discussion stage where the expert narrows the choice set and possibly provides a recommendation as a precursor to the individual making the final choice. Decisions about prescribed contraception fit this stylized version of an individual-expert interaction. In the first stage of a consultation between a woman and her doctor, the doctor will be aware of a wide range of available contraceptive products and faces the task of matching appropriate products to the particular patient characterized by such attributes as their medical history, fertility plans and preferences. This process is likely to lead to a narrowing of the choice set that may be further restricted by costs faced by the doctor such as the time available during a regular consultation. Borrowing from a terminology used extensively in marketing this resultant subset of possible products that the doctor ultimately chooses to discuss with the patient is called a consideration set; see for example Roberts and Lattin (1991).

Our focus is this discussion stage of a clinical encounter where the doctor decides on the consideration set. While the doctor is an expert, elements of complexity are likely to remain. Heiner (1983) and de Palma et al. (1994) emphasize the likely variation in the ability of decision-makers to evaluate alternatives. There will be differences in the expertise of doctors and their familiarity with reproductive health. Moreover, certain combinations of patient characteristics are likely to induce a level of complexity that increases the cognitive burden of matching contraceptive products to particular women. Frank and Zeckhauser (2007) suggest that such complexity is just one of the forces acting on primary care doctors making them less likely to make “custom made” choices and instead more likely to revert to “ready-to-wear” or norm-based choices.

Such discussions of complexity lead naturally to heteroskedastic choice models (Mazzota and Opaluch, 1995, Swait and Adamowicz, 2001, Sándor and Franses, 2009). Here the GMNL of Fiebig et al. (2010) provides the econometric framework for our analysis. GMNL explicitly allows for scale heterogeneity (equivalently individual level heteroskedasticity) that accommodates variation in the ability of the doctor and/or their tendency to revert to practice norms. This form of heterogeneity is also allowed to interact with patient complexity captured through a heteroskedastic random coefficient specification.

Our focus on contraceptive choice is particularly timely because the range of contraceptive products available has expanded rapidly making it more challenging for women to understand the choices available and the trade-offs they present, and more challenging for doctors to provide comprehensive information to assist women in making a fully informed contraceptive choice. 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 (LARC) 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). LARC methods are contraceptives that are administered less frequently than monthly and include hormonal implants, intrauterine contraception (IUC), both hormonal and copper-bearing, and contraceptive injections.

Black et al. (2013) note that despite having relatively high rates of unintended pregnancies and abortions, Australia has relatively low uptake of LARC methods. In a survey of Australian women, 32% of first pregnancies were reported as unplanned and 29% were unwanted (Weisberg et al. 2008). Eeckhaut et al. (2014) provide a cross-country comparison of the use of LARC methods and report rates of use in Australia (7%) that are lower than in the US (10%) and much lower than in a selection of European countries (10-32%). We are especially interested in the tendencies of doctors to discuss LARC methods in order to determine whether part of the explanation for the relatively low uptake of LARC is reluctance on the part of some doctors to even discuss these products. Such an investigation has the potential to provide new evidence to inform policy discussions where, for example, calls to incentivize general practitioners (GPs) to provide LARC information (Black et al. 2013) are predicated on GPs not currently providing such information.

A stated preference (SP) choice task is developed and implemented using a sample of Australian GPs. For a sequence of hypothetical women, each defined by a set of personal characteristics, GPs

were asked to indicate which specific contraceptive products would form the consideration set to be discussed with the patient. Research designs where real doctors evaluate hypothetical patients have been implemented in a number of different contexts and with a range of methods. What we are doing is often called a clinical vignette; see Peabody et al. (2004). Sometimes the vignette is presented in the form of a videotaped patient portrayed by an actor; see Lutfey et al. (2009). Validation of such approaches has been undertaken by Peabody et al. (2004) where the gold standard method was taken to be standardized patients where again trained actors simulate a patient but where the encounter involves face-to-face interaction with the doctor. A variant of this approach has been used by Currie et al. (2011) and Lu (2014) in field experiments where the actor portrays a family member who is consulting the doctor on behalf of a distant relative; this type of interaction is not uncommon in China where these studies were conducted. Our point of difference is to use SP methods common in discrete choice experiments (DCE); see Louviere et al. (2000) and Street and Burgess (2007). This approach delivers advantages in terms of a wider coverage of the type of patients considered and cost effective collection of data by requiring doctors to evaluate multiple patients.

This is not a standard DCE of the type common in health economics where a fixed context is provided and the attributes of the products in the choice set are varied. For example, Hole et al. (2013) describe a particular patient and then ask doctors to choose between two hypothetical drugs where it is the attributes of the drugs that vary over choice occasions. Instead, we experimentally manipulate the patient characteristics and keep the alternatives fixed. In each scenario the GPs encounter a different hypothetical woman and are asked to match them to a fixed but comprehensive range of contraceptive products identified by generic labels. This leads to another feature that distinguishes our work from standard DCEs and other applications involving hypothetical patients. By focussing on the discussion stage of the clinical encounter rather than the ultimate recommendation that is made, the econometric analysis must accommodate outcomes that involve choices of multiple products rather than the typical situation where a single choice is the outcome of interest.

Analysis of these data confirms that in many situations only a subset of products is in fact discussed. We are also able to identify clinical, life-cycle and socio-demographic characteristics of women that are important in shaping the form of the consultation, as defined by the products discussed, to identify variations associated with GP characteristics and to quantify the impact of the specified sources of complexity. Using our econometric results we simulate predictions of the probability that particular products are discussed to highlight that GPs (of all persuasions) are almost certain to

include the Combined Pill in their consideration sets for a wide class of women where there is no clinical reason to restrict the products to be discussed. While there is consensus amongst GPs about discussing the Combined Pill, no such clear agreement emerges for any of the individual LARC methods. Movement away from the Combined Pill towards LARC products is in part associated with patient complexity. But these tendencies interact with substantial doctor specific heterogeneity associated with all contraceptive products other than the Combined Pill. These estimated effects are consistent with many of our doctors reverting to "ready-to-wear" or norm-based choices.

2. The expert's choice problem and econometric framework

The stylized version of the interaction that underpins our analysis, involves an expert, who can be characterized as being more informed about the alternatives available than the decision-maker who ultimately needs to make the choice. The expert needs to convey information to the decision-maker and does so by choosing to discuss a subset of alternatives, the consideration set, before making a recommendation. The individual subsequently makes a decision on the basis of all the information she has acquired, including that conveyed by the expert. While what follows is relatively general it will be convenient to focus on our application and hence refer to the expert as a doctor, the decision-maker as a patient and the choice set as a range of alternative prescribed products.

Doctors need to evaluate alternative prescribed contraceptive products and determine how they match their patient described by her particular preferences, clinical indicators and personal characteristics. The choice being made by the doctor is whether or not to discuss each of the products in the universe of possible products with a particular woman. Within a random utility framework the benefit of discussing the *j*th product is represented by:

$$(1)\ \ U_j = V_j + u_j; j = 1, \cdots, J.$$

V_j represents the predictable component of the overall utility of discussing product *j.* We refer to V_j as the "index" which will be specified as a function of observable characteristics of the patient, the product and the doctor. Choosing to discuss product *j* requires a comparison with the benefit of not discussing denoted by U_0 where for identification purposes the predictable component is normalized to zero. Under the assumption of independently distributed Type-I extreme value stochastic error terms the probability of discussing the product takes the form:

$$(2)\ \ Prob(y_j = 1) = Prob(U_j - U_0 > 0) = \frac{\exp(\sigma V_j)}{1 + \exp(\sigma V_j)}$$

where σ is the scale parameter which would be set to unity in the case of a standard binary logit model. For our modelling this scale parameter plays a critical role. Heiner (1983) and de Palma et al. (1994) emphasize the likely variation in the ability or willingness of decision-makers to accurately evaluate alternatives and translate this into variance or scale heterogeneity. Complexity adds extra noise to the error term in this random utility framework. If there is variation in scale across the population of doctors then a small scale (large variability) is associated with more random behavior:

$$(3)\ \ \lim_{\sigma \to 0} Prob(y_j = 1) = 0.5.$$

As choice behavior becomes more random the probability of the product being discussed is as likely as not being discussed irrespective of the product specific net benefits identified in V_j. This captures the tension in the choice problem discussed by Frank and Zeckhauser (2007) who distinguish between "custom made" and "ready-to-wear" or norm-based choices. A custom made choice involves the doctor undertaking a careful evaluation of the patient and then matching them to an appropriate product. In terms of the model, the index is accurately assessed and this drives the choice with little role for uncertainty on the part of the decision-maker. Alternatively, as new products are introduced, doctors face considerable costs in the process of gaining the knowledge and expertise required to discuss and prescribe these products. This is particularly the case when more familiar and acceptable products are available even though they may be somewhat inferior to the new products. This is an especially salient cost in our situation. Allowing for scale heterogeneity captures the tendencies of some doctors to adopt norms (here particular products) that work well for a broad class of women and to place less weight on certain patient attributes that would indicate a different product that is potentially a better match. Observing doctor choices across multiple products for different women provides some evidence on the source of doctor heterogeneity.

In order to generate a measure of patient complexity, we follow Swait and Adamowicz (2001) and equate patient complexity with uncertainty about the ultimate recommendation. In cases where the recommendation is clear one alternative will have a probability of being recommended that approaches unity. At the other extreme a complex patient will be one where opinion is divided amongst products and the distribution of probabilities across products is uniform. Entropy captures such uncertainty and is defined as:

$$(4)\ e_t = -\sum_{j=1}^{J} p(j,t) log\big(p(j,t)\big),$$

where $p(j,t)$ is the probability that product j is recommended by the doctor for woman t. This measure achieves a minimum of zero when one product is certainly chosen and a maximum of log *J* when all products are equally likely.

It could be the case that norms are used in response to more complex patients where there are high cognitive costs associated with matching them to appropriate products. Thus patient complexity can have an effect on the evaluation of the woman that is captured in V_j and it is possible for this to interact with doctor heterogeneity.

These considerations lead to our general model specification that is a form of GMNL, Fiebig et al. (2010), where we allow for patient complexity, scale heterogeneity and their interaction. For a representative product this is given by:

$$(5)\ y_{is}^{*} = (\gamma_{0i} + \gamma_{1i} e_{1is} + \gamma_{2i} e_{2is})\sigma_i + \mathrm{a}_{is}'(\delta\sigma_i) + \mathrm{z}_i'(\pi\sigma_i) + \varepsilon_{is}; i = 1, \cdots, N; s = 1, \cdots, S;$$

$$(6)\ \gamma_{ki} = \gamma_k + \eta_{ki}; \eta_{ki} \sim N(0, \omega_k^2)$$

$$(7)\ \sigma_i = \exp(\bar{\sigma} + \tau v_i);\ v_i \sim N(0,1)$$

$$(8)\ y_{is} = 1[y_{is}^{*} > 0]$$

where y_{is}^{*} represents the latent net evaluation that doctor *i* assigns to discussing this contraceptive product when faced with choice scenario *s* representing a particular woman and this is related to the observed binary outcomes, y_{is}, according to equation (8). Note that 1[.] is the indicator function.

The vector of explanatory variables a_{is} contains attributes of the hypothetical women while z_i represents the vector of observed characteristics of our sample of GPs. The associated vectors of coefficients denoted by δ and π represent how the attributes of women and characteristics of GPs impact the probability of discussing this contraceptive product.

Recall that product attributes are not included and therefore the constant (γ_{0i}) in (5) captures the evaluation of this product conditional on patient attributes. This is likely to be GP specific and even though we control for observable GP characteristics we specify it as a random parameter in order to capture unobservable GP effects and to control for the likely correlation across the multiple choices being made by each GP. Patient complexity is captured by allowing the constant to also vary across

the three levels of entropy where $e_{1is}$and e_{2is} represent medium and high entropy so that low entropy is the base case. These entropy effects are also allowed to vary across doctors and this specification is equivalent to assuming error components that induce heteroskedasticity associated with higher levels of patient complexity. While this particular specification is a priori sensible, sensitivity checks were performed to confirm this specification decision. These random coefficients are all assumed to be normally distributed.

Note that we are not necessarily assuming that GPs are making these decisions to maximize the utility of their patients. Specifically, the presence of GP effects, conditional on patient characteristics, allows the possibility that GPs shade their choices to, in part, reflect their own preferences or expertise. GP-specific affects are also assumed for scale. In specifying the distribution of scale heterogeneity, $\bar{\sigma}$ is a normalizing constant required to ensure identification. This is achieved by setting:

$$(9)\ \ \bar{\sigma} = \frac{\tau^2}{2} \Rightarrow E(\sigma_i) = 1.$$

One attraction of the specification of scale heterogeneity given in (7) is a considerable amount of flexibility with the addition of only one parameter, τ. This parameter provides a measure of scale heterogeneity and if τ=0, the GMNL model reduces to a mixed logit specification with random parameters for the patient complexity effects.

3. Choice task

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 (2010)) and the products GPs are prescribing (Mazza et al. 2012), these data provide little or no detailed information about the interaction between the patient and the doctor. SP methods provide a natural methodology to learn more about this particular choice process.

The choice 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 which products they would discuss and ultimately recommend. In doing so we abstract from the product attributes and instead focus on how the patient characteristics impact on the choices of the GP.

The actual choice task reflected findings from a literature review and was strongly guided by 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. GPs identified 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 in the Australian market, and this was supported by the findings from the focus group discussions, the choice task did not specify the product attributes apart from a label. GPs 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. A benefit of this approach is that, compared with clinical vignettes (Peabody et al., 2004), it allows us to include a broader range of attributes, and facilitates the doctors being presented with a larger number of scenarios than would otherwise be possible.

==Table 1 about here==

Nonetheless, it is only feasible to show a subset of the ($4^4x3^6x2^2$) =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 attribute descriptions were worded such that implausible combinations were rare and unlikely to be included in the design. That is, there were very few combinations of attributes defining a particular woman that were not at least feasible.

Doctors were asked to answer a sequence of three questions pertaining to each particular patient. A stylized version of the entire choice task for one woman is provided in Figure 1. The predominant method of contraception amongst Australian women is a form of the contraceptive pill (Gray and McDonald, 2010). Thus, for the first question 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, the second question required the GP to indicate which specific contraceptive products, constrained by the broad category they chose in the first question, 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 choice was restricted to the products specified in the previous question. Thus at no stage in the sequence of questions were respondents permitted to make inconsistent choices. The focus here is on the outcomes from the second question where the consideration sets are defined. Because of the structure of our choice task some of these decisions about whether to consider a product or not was effectively made in answering the first question about broad product types. The recommendation data are not explicitly modelled here but are pursued in Fiebig et al. (2015).

==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 (progestogen-only pill), Hormonal Injection, Hormonal Implant, Hormonal Intra-uterine Device (Hormonal IUD), Hormonal Patch, vaginal Ring, copper IUD and Condoms. Our emphasis was on prescribed products but expert advice from our clinical authors was that it would 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 for dual protection against Sexually Transmitted Infections (STIs). The vaginal Ring was relatively new to the Australian market and the Patch was not available at all. Doctors were expected to have some knowledge of the attributes of the Ring and Patch because of their existence in other countries.

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. As has been found in other studies, the response rate for GPs was low (Britt et al., 2008). Because of this, 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.

4. Data and summary statistics

In the first question GPs indicate which of three broad product categories they would discuss with a specific patient. The raw frequencies across all 2592 choice occasions indicate that GPs 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”.

In the second question GPs chose what we are calling a consideration set comprising a subset of particular products. It is these outcomes that are our primary data for analysis. Over all choice occasions by all GPs, the median and modal number of products discussed is 4 and in less than 1% of choice occasions did GPs indicate they would discuss all products. These basic results indicate the existence of consideration sets whereby GPs almost always discuss a subset of available contraceptive products with patients. At the other extreme, on 4.3% of choice occasions the consideration set was a solitary product and a majority of these (52%) were when the GP said they would only discuss the Combined Pill.

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

the GPs. The final column labelled “conversion” gives the recommendations expressed as a percentage of the times the product was considered.

The conversion percentages highlight two key features. The Combined Pill, Implant and Hormonal IUD are distinctive because they are very likely to be discussed and conditional on being discussed are quite likely to be recommended. In contrast Condoms, Ring and Injection are quite likely to be discussed but their very low conversion rates indicate they are relatively less likely to be recommended. For Condoms this is not unexpected given that doctors are likely to discuss these in conjunction with prescribed products in the context of STI protection and hence not subsequently recommend them for contraception.

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 which may have led to unwillingness on the part of GPs to choose this alternative even in a hypothetical setting where the product is assumed to be available. Nonetheless the Ring, which is only very recently available in Australia, was much more likely to be considered and even recommended.

In order to derive our measure of patient complexity a multinomial logit model is estimated using the recommendations data. The predicted probabilities for each of the products, except the Patch, were used to predict entropy for each of the 192 distinct women in the design. These are then categorized into low (11%), medium (40%) and high entropy (49%) cases. The high entropy women are more likely to be older, have irregular periods, elevated blood pressure and plans to have children in next 2 years. Low entropy women are almost certainly younger (<29) and have children but are not breastfeeding and are very likely to have normal blood pressure.

A number of GP characteristics were included in the subsequent analyses. They are the GP’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 bulk billed (implying patients face no out-of-pocket expenses) whether they graduated from an Australian medical program and whether their practice is located in an urban area. Table 3 provides some summary statistics on these key characteristics of the doctors in the sample. Comparing characteristics of our GP sample to the Australian population of GPs and other national samples of GPs, (Britt el al., 2008), we find our sample to be representative in terms of age, location, practice characteristics and bulk-billing rates. The

proportions of female GPs and those who were Fellows of the RACGP however were higher, reflecting the likelihood that this sample was more engaged in reproductive health than other GPs.

==Tables 2 & 3 about here==

5. Estimation results

The outcomes of interest are a set of nine binary indicators corresponding to each of the contraceptive products and which denote whether or not the GP said they would discuss that particular product given the hypothetical patient. While this effectively represents a system of binary choice equations where there is likely to be cross-equation correlations, the advantages of joint estimation of such a complete specification are confined to increased efficiency of estimates. Because each of the equations contain the same set of regressors, efficiency gains are likely to be minimal. In the classical seemingly unrelated regression case there are no gains from joint estimation; see Fiebig (2001). Thus models described by (5) – (8) were estimated separately for each product by maximum simulated likelihood; see Gu et al. (2013) for more detail.

5.1 Comparison of model fit

In order to provide some indication of the relative contribution of the two sources of complexity a sequence of alternative specifications are estimated and their relative fit compared. As a baseline model consider a random effects logit specification (M1) where $\gamma_{1i} = \gamma_{2i} = 0$ and τ=0 implying both sources of complexity are ignored. Then patient complexity (M2) and GP scale heterogeneity (M3) can be added separately. Finally our general GMNL model (M4) is estimated where both features are included. In order to compare the improvement in fit McFadden's R-squareds are reported with the simulated log-likelihood of the random effects logit (M1) specification as the base.

==Figure 2 about here==

The improvement in the M4 fit is marked. M4 nests the other models and LR tests comparing M4 to M1 confirms significant improvement in fit for all products. But a comparison across products indicates considerable variation in the relative impact of sources of complexity on fit. Introducing patient complexity in M2 typically yields modest improvements compared to accounting for scale heterogeneity in M3 despite M3 being the more parsimonious model. Moving from M1 to M2 involves 4 additional parameters while M1 to M3 requires only a single additional parameter.

The Combined Pill is a notable exception to this superiority of M3. Here patient complexity has a relatively large impact on fit and GP scale effects are relatively small. The Condom is the only other product where M2 fits better than M3 although for the Mini-pill both forms of complexity have comparable impacts on fit. For the other products, M3 typically fits dramatically better than M2; see especially the IUD and Ring. For the Ring it is clear that the M4 improvement is almost all due to GP scale effects. In several cases there seems to be an interaction effect between the two sources of complexity, this is especially evident for the Implant.

The fit statistics for the Patch have been included but here M3 and M4 are somewhat different variants of the model estimated for the other products. Because of the rarity of this product being considered (more than 80% of GPs never chose this product in any of their 16 scenarios) there were convergence problems with the GMNL specification in (5) associated with τ becoming excessively large to accommodate the choice patterns. Fiebig et al. (2010) made note of this extreme form of variability and suggested not scaling the alternative specific constant as a solution. This is the variant of the M3/M4 models reported here for the Patch. Because this represents a special case making comparison with other products difficult and because of the rarity of it being considered, we will not pursue further discussion of this product.

While these comparisons confirm improvements in fit, it is important to check how improved fit impacts substantive findings. Presenting these estimation results efficiently and informatively poses a major challenge. Thus a full set of estimation results for each of the contraceptive products is not presented but is available on request. Instead, selected results are provided relying heavily on a range of graphical summaries. The selection of results is driven by our focus on the impact of complexity and in particular on LARC methods and the main aim of obtaining a better understanding of their role in the discussion of contraceptive options between GPs and patients.

5.2 Estimated impact of complexity on the probability of being discussed

First consider scale heterogeneity and recall that τ=0 would imply no scale effects and when there is a non-zero τ, scale is normalized to have a mean of unity. The τ estimates together with their 95% confidence intervals are provided in the left hand panel of Figure 3. The uniformly significant τ estimates were anticipated from the considerable improvement in fit provided by M3. There is also considerable logic to the relative magnitudes across products. The most common and best known contraceptive is the Combined Pill and it is associated with the least amount of GP variability. Conversely the products with the 3 largest amounts of GP variability are the IUD which is being

superseded by the Hormonal IUD; the Ring which is the newest and least known product; and the Implant where there seems to be extremely divergent views across GPs.

Significant GP effects, which are present here for all products, manifest themselves in a range of scale values across GPs that imply the estimated index parameters get scaled up or down depending and whether the scale is greater or less than unity. Moreover, as τ increases the scale distribution becomes more heavily skewed to the right while keeping the mean equal to 1. To assist in interpretation and to better understand the implications of alternative τ values consider the estimated median of the scale distribution together with their 95% confidence intervals that are presented in the right hand panel of Figure 3. The relatively low estimated τ for the Combined Pill implies a modest amount of scale heterogeneity and a median scale effect that is not significantly different from unity.

==Figure 3 about here==

At the other extreme the large τ estimate for the Implant implies considerable GP scale heterogeneity and the estimated median is now only 0.13. Even though the distribution has a mean of unity, with large amounts of heterogeneity the distribution of scale is highly skewed. For the Implant the choices being made by the GPs are subject to huge scale effects and thus for a given index comprising particular GP/woman combinations the probabilities can be very different depending where we are in the scale distribution. The same is true to a somewhat lesser extent for the ring and IUD. The HIUD, Condom, Injection and Mini-pill are an intermediate cluster.

The results for patient complexity are presented separately for the mean (Figure 4) and variance (Figure 5) effects. Our specification has low entropy as the base and the estimated effects of medium and high entropy patients are relative to this base. Choice of the base when dummy coding in random coefficient models imposes constraints of the implied form of heteroskedasticity but we checked for such coding sensitivity and this specification proved appropriate. This is in part due to patient variance effects not being overly significant and hence not important in terms of improved fit.

==Figures 4 & 5 about here==

Amongst the LARC all except the IUD exhibit significant patient complexity mean effects. The Implant, HIUD and Injection are all more likely to be discussed in higher entropy patients. Because of the differential scaling effects we need to be careful in comparing magnitudes of effects across products. For example, the large Implant effects will tend to be scaled down. Amongst the non-LARC products there are significant shifts in the mean associated with both pills. For the Mini-pill again there is a tendency for higher complexity to have a positive impact but the reverse is true for the Combined Pill. The variance effects are modest. Only for the Implant are there significant variance effects at both levels.

5.3 Comparison of predicted probabilities across models

The estimated models can be used to generate predictions of whether a product is discussed and these can be informative about the impact of accounting for complexity. Figure 6 displays a comparison of density plots of the predicted probabilities of discussing the Combined Pill generated by the baseline M1 model and the full M4 model. These are provided for all within sample GP-women combinations and then separately for the women characterized by the three levels of entropy.

For the Combined Pill, the impact of accounting for complexity is most pronounced in the low entropy regime. Even without modelling complexity, it is clear that predicted M1 probabilities are large with little variation. These two key features are reinforced after including complexity in M4. In the high entropy regime there is a noticeable shift to the left in the density.

Figure 7 provides a comparable set of plots for the Implant. Again explicitly modelling complexity matters for predictions although here there are similar effects over entropy regimes. Recall that this was a product where interaction between the sources of complexity mattered. On one hand this is the product chosen most often to be discussed and was a product where patient complexity was significant. But despite many GPs indicating they would discuss the Implant, there are some GPs who never discuss it. In order to capture this considerable GP diversity the variability in scale is large. The appearance of a tendency to "Bimodality" in part reflects tension between willingness to discuss that suggests high predicted probabilities and large scale effects which imply some of these large probabilities get muted.

==Figures 6 & 7 about here==

Similar graphs for each of the products have not been presented but they do indicate an impact albeit not as stark as the two extreme cases that have been provided.

5.4 Comparison of predicted probabilities across women

Table 4 provides the mean and the range of predicted probabilities for all within sample GP-women combinations for all products. These are the same predictions used to generate Figures 6 and 7. It is important to note that these results are not expected to reflect market shares for actual product use because: (i) they are not choices made by women (or even recommendations made by GPs); (ii) the design is chosen to induce trading and not to reflect actual behaviour; and (iii) we included a product (Patch) that is not available on the Australian market.

The means of the predictions essentially mimics the summary statistics presented in Table 3. What is new relative to the summary statistics is the extent of variability in the predictions. Across the board there is considerable variability reflecting both the range of women presented for evaluation and GP effects. In part, this is to be expected. Within the controlled environment of the SP choice task, a primary aim is to choose attribute levels in order to encourage trading which, in turn, will lead to variability in outcomes. So in some sense the choice task would be considered a failure if we did not observe such variability.

==Table 4 about here==

Because the focus is on LARC and the fact that clinical guidelines supports their use as first line contraceptives for women of all reproductive ages, we are interested in what GPs say they would discuss with "low risk" women. Such a woman would have no clinical, life cycle or preference related attribute that should *a priori* deter GPs from discussing LARC. Our "Baseline" woman defined in Table 1 by the set of omitted categories in our dummy coded model specification is such a case. She is 20-29, starting prescribed contraception for the first time, has no problems with periods, has normal blood pressure, is in a new relationship, has no children but plans to have children but not in the next 2 years, has no strong opinion about the pill, is not concerned about gaining weight, has no difficulty with compliance, has low to middle household income and is a non-smoker. There is nothing here that should deter GPs from saying they would discuss LARC. The middle section of Table 4 provides predictions for our low risk woman. The overall compression of the range of predictions is to be expected as one of the possible distinct women in our design has been isolated.

Any remaining variation in predictions is attributable to the combined impact of GP effects and the mix of associated GP characteristics in our sample.

Possibly the most dramatic feature of these results is the change in the Combined Pill. Our model predicts that GPs (of all persuasions) are almost certain to include the Combined Pill in their consideration sets. Amongst the LARC products, the Implant and Injection have small increases in popularity and exhibit somewhat less variability but neither effect is as dramatic as that observed for the Combined Pill. The other two LARC products move in the opposite direction with the mean predictions for the Hormonal IUD and IUD both decreasing. The results indicate that these two products together with the Mini-pill are unlikely to feature in the consideration set for the low risk woman.

There are a number of dimensions in which we could change the baseline characteristics to define alternative "low-risk" women without seriously disrupting the overall pattern. For example, ageing our Baseline woman (holding all else constant) would lead to large increases in the predicted probabilities of the Hormonal IUD being discussed but other changes would be relatively minor. This indicates that the key features of the results are representative for a relatively broad class of woman and are not confined to the baseline characteristics.

The variation in the LARC predictions is associated with both GP scale effects and observable GP characteristics. For the Implant, being trained in Australia and having a Family Planning Certificate were both positive and significant effects in explaining the consideration choices. If we simulate a situation where all of our GPs have these two characteristics then the mean prediction of considering the Implant increases from 0.727 to 0.796 and so even here the Implant does not reach the levels of the Combined Pill. What's more, considerable variation still remains with the predictions ranging from 0.683 to 0.898.

From the range of predictions across the entire sample, we know that there are situations where the probability of discussing the Combined Pill can be relatively low. An interesting scenario to investigate is what happens when GPs are faced with a woman with risk factors that deter discussing the Combined Pill. An example of such a "some risk" woman would be one who is breastfeeding and has high blood pressure but has the baseline level for all other attributes. The predictions for this woman are presented in the right hand section of Table 4. Essentially there seems to be a substitution to the Mini-pill. For this woman the mean predicted probability for the Combined Pill

declines from 0.979 to 0.454 while the Mini-pill increases from 0.168 to 0.567. Amongst the LARC products there is no comparable change although the Hormonal IUD mean prediction increases from 0.092 to 0.285 and the IUD from 0.166 to 0.288.

6. Discussion of results

Our analysis predicts that GPs (of all persuasions) are almost certain to include the Combined Pill in their consideration sets for a wide class of woman not too different from our low-risk (Baseline) woman. While there seems to be considerable consensus amongst GPs about discussing the Combined Pill, no such clear agreement emerges for any of the individual LARC methods. In part, the Combined Pill result reflects the actual, current choices of Australian women and hence represents an entirely rational decision on the part of GPs. Given the current evidence supporting the use of LARC for these women, the question is why there isn't a similar result for a LARC option. The Implant comes close. While it is included in the consideration sets of GPs on more choice occasions than any other product, for the low risk woman it does not reach the levels of certainty of being discussed displayed by the Combined Pill. In particular, this product exhibits the highest levels of GP heterogeneity reflecting divided reactions amongst GPs.

This comparison needs some qualification because we are taking a snapshot at one point in time and comparing the Combined Pill with LARC products, where the former has been available and popular for an extended period while awareness and product knowledge is still developing for the latter. It could be that LARC acceptance amongst doctors is increasing. Some support for this scenario is provided when we model more highly qualified GPs who were trained in Australia. This increases the rate at which the Implant was considered for low risk women, even though the Combined Pill remains the dominant single product.

It is highly likely that the unwillingness of GPs to discuss LARC relative to the Combined Pill is associated with their experience with the effectiveness of the Combined Pill that has grown over the past 50 years. Both GPs and women are so comfortable with the idea of the Combined Pill as a relatively cheap, safe and effective form of contraception that everything else is considered in relation to it. Conversely, GPs were willing to customize their discussions of the Combined Pill with particular women. For this product, more than any other, the decision whether to discuss or not was associated more with the characteristics of the women than heterogeneity of the GPs. It is true though that our comparison of the women with no risk factors versus those with some risk factors

suggests that when GPs more away from the Combined Pill, the Mini-pill is a likely replacement in the discussion rather than a LARC.

There are a number of costs that form impediments to LARC products reaching the level of consensus associated with the Combined Pill. As new products are introduced, doctors face considerable costs to gain the knowledge to be able to discuss and prescribe some of these products. This is particularly the case when the more familiar pills are available even though they may be somewhat less effective than the newer LARC products.

Even if a GP has acquired knowledge about the efficacy of LARC and which LARC products are suitable for a particular type of woman, it is unlikely that s/he will discuss a product with a woman unless s/he can also be confident about prescribing and providing it. This is more difficult in relation to LARC products than for the Combined Pill as the GP has to either acquire the knowledge and expertise to insert the Implant or IUC methods or be willing to refer their patients elsewhere.

There is also the issue of the economic incentives facing the woman; while LARC products except for the copper IUD are subsidised by the Pharmaceutical Benefits Scheme (PBS), so are many brands of the Combined Pill and the upfront cost of the latter is lower than for an implant or IUD methods which generally requires an additional insertion cost. It is also the case that prescription of a LARC often necessitates a second consultation (either with a GP, specialist or Family Planning Clinic), more inconvenient than filling a script for the Combined Pill and likely to be associated with an out-of-pocket cost to the woman. The Medicare rebate system in Australia may also discourage GPs from inserting LARC, particularly the Hormonal or copper IUD, as the rebate is low relative to the time, equipment and skills involved.

Our results pose something of a challenge for enhancing LARC uptake in Australia, given increasing evidence of their greater effectiveness in preventing unwanted pregnancy. They suggest the need for greater education of GPs of their benefits and suitability. Even for the Implant, which was a popular choice for discussion, there was considerable variation across GPs possibly reflecting the impact of some of the costs mentioned above on some doctors. A possibly more worrying result is evident in the simulation for the low risk woman where there was little support for discussing the Hormonal IUD. This suggests that the suitability of intrauterine methods for younger women as well as those in new relationships has not yet become widely accepted. Instead they seem to be viewed as one of the products that get considered when the Combined Pill becomes problematic.

Our results indicate support for LARC amongst GPs but this support is not uniform nor is it at the level enjoyed by the Combined Pill. It should be acknowledged that the analysis is based on a potentially biased sample in terms of the GP participants being more interested and informed about contraception. The context of our choice task also encourages the idea of discussion of different products as it involves a woman explicitly consulting the GP for "information, advice and potentially a prescription". If these two aspects have introduced any biases it is likely to manifest itself in a greater propensity to discuss LARC which would only reinforce the tendencies we have documented.

7. Conclusions

Maintaining good reproductive health requires women to have access to information about safe, effective and affordable methods of contraception. By exploiting SP methods we have been able to provide new insights into medical decision-making in the particular case of GPs providing such information in a discussion of potential contraceptives with women. Our analysis for Australia points to less than universal support for discussing LARC in cases where best practice guidelines suggest this is appropriate. This provides evidence that part of the explanation for relatively low uptake of LARC methods is a reluctance on the part of GPs to recommend or even discuss them. Evidence of significant GP specific effects reflecting norm-based practices associated with whether LARC methods are discussed suggests greater investment in the education and training of GPs may be warranted and particularly for those GPs who were not educated in Australia. Australia currently has no national policy promoting LARC awareness and provision at either the health care professional or community level unlike in the UK and US.

Our primary focus has been on the GP providing information to women. It could be true that low LARC uptake partly reflects a general lack of community awareness so that women simply aren't asking for fuller discussions of contraceptive options and GPs are not being proactive in initiating such discussions. Importantly, our results condition on the interaction between the GP and the woman. While such an interaction is required for prescribed contraceptives, other popular contraceptive products (especially Condoms) do not require a prescription. For both these reasons, policies that incentivize GPs to provide LARC information (Black et al. 2013) can only be part of the response. It is necessary to ensure such information flows to women, especially young women, from additional sources other than GPs and so that ultimately women are making a fully informed contraceptive choice.

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.

Bartels, R., Fiebig, D.G. and van Soest, A. (2006), "Consumers and experts: An econometric analysis of the demand for water heaters", *Empirical Economics*, 31, 369-391.

Bateman, H., Eckert, C., Geweke, J., Louviere, J., Satchell, S. and Thorp, S. (2014), "Financial competence, risk presentation and retirement portfolio preferences", *Journal of Pension Economics and Finance*, 13, 27-61.

Black, K., Bateson, D. and Harvey, C. (2013), "Australian women need increased access to long-acting reversible contraception", *Medical Journal of Australia*, 199(5), 317-318.

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.

Currie, J., Lin, W. and Zhang, W. (2011), "Patient knowledge and antibiotic abuse: Evidence from an audit study in China", *Journal of Health Economics*, 30, 933-949.

de Palma, A., Myers, G.M. and Papageorgiou, Y.Y. (1994), "Rational choice under an imperfect ability to choose", *American Economic Review*, 84, 419-440.

Eeckhaut, M.C.W., Sweeney, M.M. and Gipson, J.D. (2014), "Who is using long-acting reversible contraceptive methods? Findings from nine low-fertility countries", *Perspectives on Sexual Reproductive Health*, 46, 149-155.

Fiebig, D.G. (2001), "Seemingly unrelated regression", in *Companion in Theoretical Econometrics*, B.H. Baltagi ed., Basil Blackwell: Oxford, 101-121.

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.

Fiebig, D.G., Viney, R., Knox, S.A., Haas, M., Street, D., Hole, A.R., Weisberg, E. and Bateson, D. (2015), "Consideration sets and their role in modelling doctor recommendations about contraceptives", mimeo.

Frank, R.G. and Lamiraud, K. (2009), "Choice, price competition and complexity in markets for health insurance", *Journal of Economic Behavior and Organization* 71, 550-562.

Frank, R.G. and Zeckhauser, R.J. (2007), "Custom-made versus ready-to-wear treatments: Behavioral propensities in physicians' choices", *Journal of Health Economics*, 26, 1101-1127.

Gray, E. and McDonald, P. (2010) "Using a reproductive life course approach to understand contraceptive method use in Australia", *Journal of Biological Science*, 42, 43-57.

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

Heiner, R.A. (1983) "The origin of predictable behavior", *American Economic Review*, 73, 560-595.

Hole, A.R., Kolstad, J.L. and Gryd-Hansen, D. (2013), "Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour", *Journal of Economic Behavior and Organization* 96, 21-31.

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.

Lu, F. (2014), "Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in China", *Journal of Development Economics*, 106, 156-167.

Lutfey, K.E., Link, C.L., Grant, R.W., Marceau, L.D. and McKinlay, J.B. (2009), "Is certainty more important than diagnosis for understanding race and gender disparities?: An experiment using coronary heart disease and depression case vignettes", *Health Policy*, 89, 279-287.

Mazza, D., Harrison, C., Taft, A., Brijnath, B. Britt, H. Hobbs, M., Stewart, K. and Hussainy, S. (2012), "Current contraceptive management in Australian general practice: an analysis of BEACH data", *Medical Journal of Australia*, 197(2), 110-114.

Mazzota, M.J. and Opaluch, J.T. (1995), "Decision making when choices are complex: A test of Heiner's hypothesis", *Land Economics*, 71, 500-515.

Peabody, J.W., Luck, J., Glassman, P., Jain, S., Hansen, J., Spell, M. and Lee, M. (2004), "Measuring the quality of physician practice by using clinical vignettes: A prospective validation study", *Annals of Internal Medicine*, 141, 771-780.

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.

Sándor, Z. and Franses, P.H. (2009), "Consumer price evaluations through choice experiments", *Journal of Applied Econometrics*, 24, 517-535.

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

Swait, J. and Adamowicz, W. (2001), "Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice", *Organizational Behavior and Human Decision Processes*, 86, 141-167.

Weisberg, E., Bateson, D., Read, C., Estoesta, J. and Lee, C. (2008), "Fertility control? Middle-aged Australian women's retrospective reports of their pregnancies", *Australian and New Zealand Journal of Public Health*, 32, 390-392.

Yusuf, F. and Siedlecky, S. (2007), "Patterns of contraceptive use in Australia: Analysis of the 2001 National Health Survey", *Journal of Biological Science*, 39, 735-744.

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
Current status	
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 baseline woman.

Table 2: Relative frequencies of contraceptive product consideration and recommendation

Product	Considered	Recommended	Conversion(%)
Long acting reversible			
Implant	0.746	0.240	32
Hormonal IUD	0.564	0.216	38
Injection	0.549	0.093	17
IUD	0.251	0.043	17
Other			
Condom	0.646	0.064	10
Combined Pill	0.532	0.192	36
Ring	0.439	0.070	16
Mini-pill	0.365	0.076	21
Patch	0.071	0.005	7

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.

Table 4: Comparison of distribution of product predictions of being considered: Means and range*

Product	Entire Sample	Low risk woman	Some risk woman
Long acting reversible			
Implant	0.723 (0.299, 0.922)	0.727 (0.396, 0.898)	0.705 (0.357, 0.884)
Hormonal IUD	0.555 (0.017, 0.967)	0.092 (0.009, 0.261)	0.285 (0.045, 0.585)
Injection	0.527 (0.153, 0.905)	0.548 (0.295, 0.815)	0.510 (0.266, 0.791)
IUD	0.276 (0.029, 0.708)	0.166 (0.069, 0.382)	0.288 (0.144, 0.545)
Condom	0.647 (0.183, 0.896)	0.668 (0.308, 0.800)	0.646 (0.268, 0.766)
Combined Pill	0.538 (0.266, 0.640)	0.979 (0.957, 0.993)	0.454 (0.266, 0.640)
Ring	0.437 (0.095, 0.755)	0.493 (0.280, 0.682)	0.197 (0.083, 0.323)
Mini-pill	0.382 (0.038, 0.858)	0.168 (0.055, 0.265)	0.567 (0.381, 0.715)

* Means are provided with (minimum, maximum) underneath. "Entire sample" refers to predictions generated over all within sample GP-women combinations. "Low risk woman" considers predictions over all GPs but for a single woman defined as the base case attribute levels in Table 1. "Some risk woman" is the baseline women but with the *Blood pressure* attribute set to "Elevated blood pressure" and the *Children attribute* set to "Breastfeeding".

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: Comparison of model fit: McFadden's R-squared with a random effects logit as the base

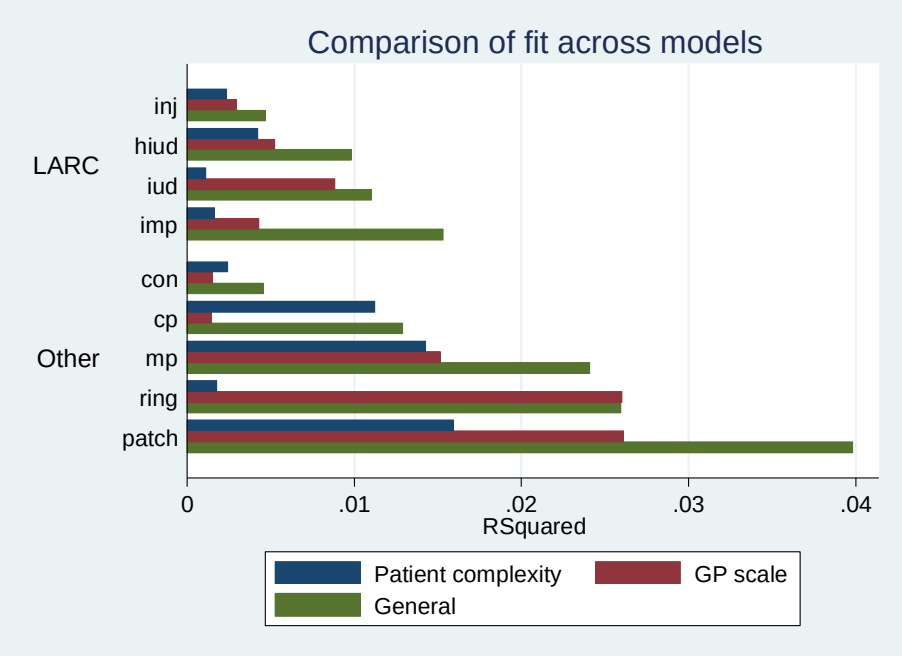

Figure 3: Estimates of GP scale variability

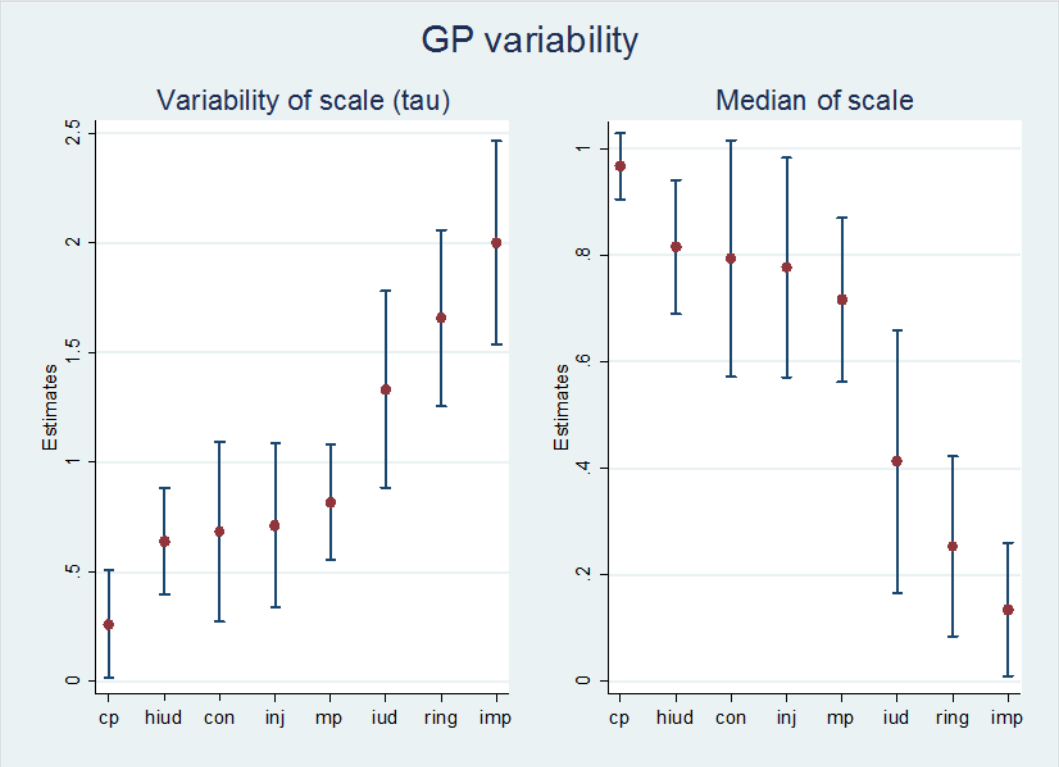

Figure 4: Estimates of patient complexity: Mean effects

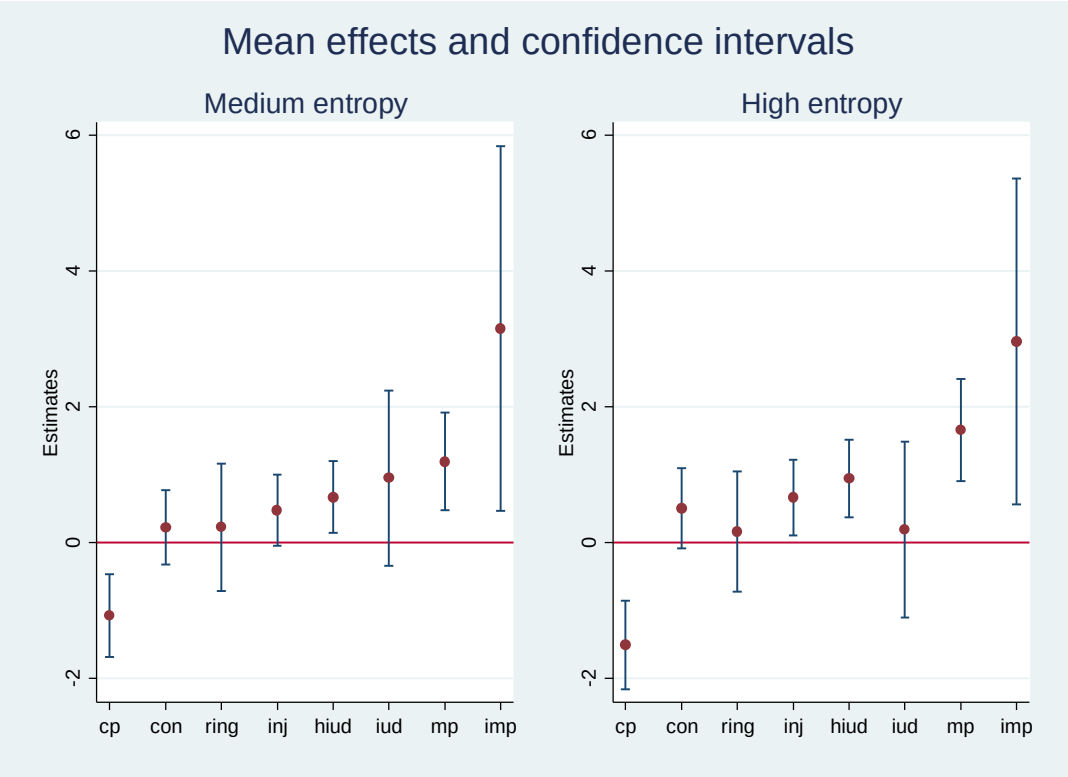

Figure 5: Estimates of patient complexity: Variance effects

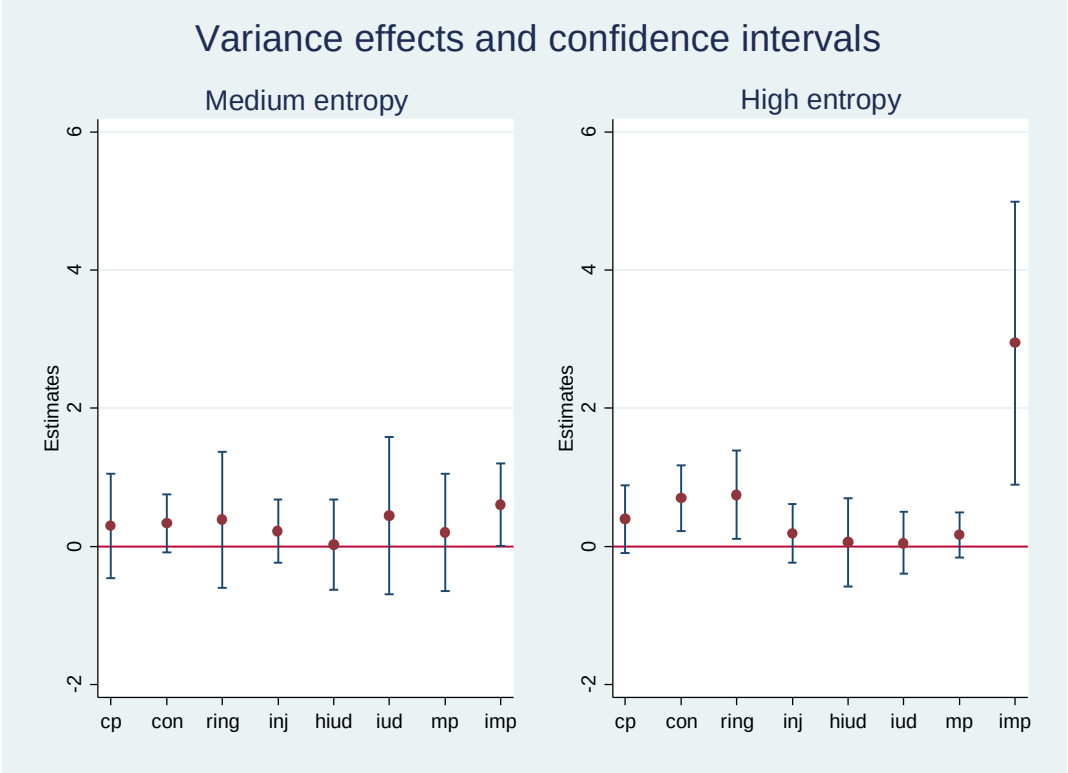

Figure 6: The impact of accounting for complexity:
Comparing Combined Pill predictions from the baseline and general models

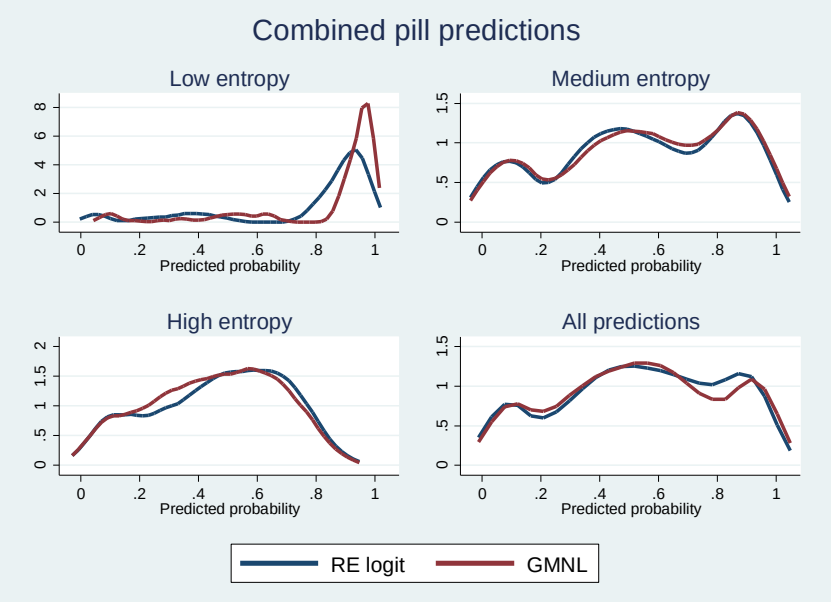

Figure 7: The impact of accounting for complexity:
Comparing of Implant predictions from the baseline and general models