

## Technical Appendix

Random utility theory acknowledges that there is a ‘random’ component of utility that is unobserved by the analyst, because it is due to unobservable factors or just psychological impulse [52]. Therefore an individual’s,  $n$ , utility,  $U$ , for alternative,  $j$ , can be defined as:

$$U_{n,j} = V_{n,j} + \varepsilon_{n,j} \quad [1]$$

$$V_{n,j} = f(\beta_k, X_{n,j}) \quad [2]$$

Where  $V_{n,j}$  is the observed component of utility which is a function of  $\beta_k$  (the utility associated with the  $K$  attributes and the level of that attribute ( $X$ )).

In this case, the utility function to be estimated is:

$$\begin{aligned} U_{nj} = & \alpha None_{nj} + \beta_1 RUNI_{nj} + \beta_2 RUNHS_{nj} + \beta_3 RUNHSG_{nj} + \beta_4 INHS_{nj} + \beta_5 INHSE_{nj} + \\ & \beta_6 INHSEB_{nj} + \beta_7 BENEFITD_{nj} + \beta_8 BENEFITG_{nj} + \beta_9 PROFITN_{nj} + \beta_{10} PROFITPUB_{nj} + \\ & \beta_{11} PROFITNV_{nj} + \beta_{12} NONGOV_{nj} + \beta_{13} PUB_{nj} + \beta_{14} GOV_{nj} + \varepsilon_{nj} \quad [3] \end{aligned}$$

In equation 3,  $\beta_{1-14}$  are preference weights associated with each attribute level (as defined in Table TA1), relative to the base case (dropped). Note,  $\beta_k$  assumes homogenous preferences as it does not vary by  $n$ . The constant,  $\alpha$ , reflects the utility associated with ‘opting-out’ with no data linkage. The random component,  $\varepsilon_{n,j}$ , means  $U$  cannot be perfectly observed.

In the heteroskedastic conditional logit model the utility function is estimated as

$$\begin{aligned} U_{n,j} = & \lambda_n \alpha_1 None_{nj} + \lambda_n \beta_1 RUNI_{nj} + \lambda_n \beta_2 RUNHS_{nj} + \lambda_n \beta_3 RUNHSG_{nj} + \lambda_n \beta_4 INHS_{nj} + \\ & \lambda_n \beta_5 INHSE_{nj} + \lambda_n \beta_6 INHSEB_{nj} + \lambda_n \beta_7 BENEFITD_{nj} + \lambda_n \beta_8 BENEFITG_{nj} + \\ & \lambda_n \beta_9 PROFITN_{nj} + \lambda_n \beta_{10} PROFITPUB_{nj} + \lambda_n \beta_{11} PROFITNV_{nj} + \lambda_n \beta_{12} NONGOV_{nj} + \\ & \lambda_n \beta_{13} PUB_{nj} Soct_n + \lambda_n \beta_{14} GOV_{nj} + \lambda_n \alpha_2 None_{nj} Soct_n + \lambda_n \beta_{15} RUNI_{nj} Soct_n + \\ & \lambda_n \beta_{16} RUNHS_{nj} Soct_n + \lambda_n \beta_{17} RUNHSG_{nj} Soct_n + \lambda_n \beta_{18} INHS_{nj} Soct_n + \lambda_n \beta_{19} INHSE_{nj} Soct_n + \end{aligned}$$

$$\begin{aligned} &\lambda_n \beta_{20} INHSEB_{nj} Soct_n + \lambda_n \beta_{21} BENEFITD_{nj} Soct_n + \lambda_n \beta_{22} BENEFITG_{nj} Soct_n + \\ &\lambda_n \beta_{23} PROFITN_{nj} Soct_n + \lambda_n \beta_{24} PROFITPUB_{nj} Soct_n + \lambda_n \beta_{25} PROFITNV_{nj} Soct_n + \\ &\lambda_n \beta_{26} NONGOV_{nj} Soct_n + \lambda_n \beta_{27} PUB_{nj} Soct_n + \lambda_n \beta_{28} GOV_{nj} Soct_n + \varepsilon_{nj} \quad [4] \end{aligned}$$

In equation 4,  $Soct_n$  is a dummy variable equal to 1 when the respondent was Scottish and 0 when they were Swedish. The dummy was interacted with each of the attribute levels.

Therefore  $\beta_{15-28}$  are the difference in preference weights associated with each attribute level due to being Scottish. Where  $\lambda_n$  is the relative scale parameter for the Scottish sample relative to the Swedish sample. HCL model parameterises  $\lambda_n$  as  $\exp(Scot \gamma)$ , and therefore testing the significance of  $\gamma$  is testing if individuals' nationality had a statistically significant effect on the scale parameter.

Table TA1: Attribute level coding

Attributes and levels	Labels	
The researchers are:	Only university researchers	RUNI
	Only university researchers or NHS staff (researchers employed by a county council)	RUNHS
	Only university researchers, NHS staff or government researchers (researchers employed by a county council or researchers employed by one of the authorities)	RUNHSG
	University researchers, NHS staff, government researchers (researchers employed by a county council or researchers employed by one of the authorities) and commercial researchers such as market research organisations or pharmaceutical companies <sup>a</sup>	RUNHSGP
The type of data being linked:	Information from your GP (primary care) records being linked with information from your other NHS (county council) health records e.g. hospital records	INHS
	Information from your NHS (county council) health records being linked with information from your social care or education records.	INHSE
	Information from your NHS (county council) health records being linked with information from your social care or education records, or from your employment and benefits (national health insurance) records.	INHSEB
	Information from your NHS (county council) health records being linked with information from your social care, education, employment, and benefits (national health insurance) records, as well as information collected about you in the private sector e.g. through online shopping accounts <sup>a</sup>	INHSEBX
The purpose of the research:	Research using linked information should only be conducted if it will have direct benefits for the people whose information is being used.	BENEFITD
	Research using linked information should only be conducted if it will have general public benefits.	BENEFITG
	Research using linked information should be allowed for any reason. <sup>a</sup>	BENEFITX
Profit-Making:	Nobody should be allowed to profit from research carried out using linked information.	PROFITN
	Any profit made from research carried out using linked information should be shared with the public.	PROFITPUB
	Any profit made from research carried out using linked information should be invested into public services.	PROFITINV
	Any profit made from research carried out using linked information should be kept by those carrying out the research. <sup>a</sup>	PROFITR
Oversight:	The process should be overseen by the Scottish (Swedish) Government.	NONGOV
	The process should be overseen by a non-governmental independent body (an independent body that is not part of the Swedish Government).	PUB
	The process should be overseen by the relevant public service(s); for example, research that uses information from people's health records should be overseen by the NHS (county council).	GOV
	The process should be overseen by the organisations undertaking the research <sup>a</sup> .	ORG

<sup>a</sup>Defined as the base level for dummy coding of the categorical levels