www.ispor.org

ISPOR

Improving healthcare decisions

IP10: Accounting for Preference Heterogeneity in Stated Preference Studies - Moving from Methods to Practice to Decision Making – the Issue of Developing Useful Guidance

Brought to you by the:

ISPOR Health Preference Research Special Interest Group

ISPOR Virtual 2021

ISPOR

Questions for the speakers?

Submit questions anytime to the Q/A feature.
Use the chat box for comments.

We will address questions & comments in the discussion session *AFTER* the presentation.

What is preference heterogeneity?

- There is not one commonly accepted definition yet, but often understood as *"different individuals make different decisions when faced with the same choice situation"*
- Decision makers are increasingly interested in heterogeneity

EXAMPLE

FDA U.S. FOOD & DRUG ADMINISTRATION

Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and *De Novo* Requests, and Inclusion in Decision Summaries and Device Labeling

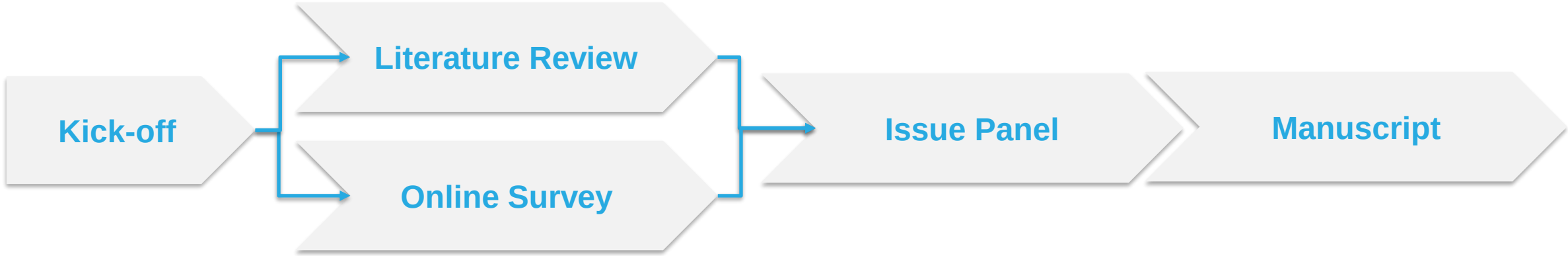
IV. Recommended Qualities of Patient Preference Studies

c) ***Capturing Heterogeneity of Patients' Preferences***: Patients' benefit-risk tradeoff preferences may be heterogeneous even among those with the same disease or condition. Individual circumstances of patients vary. Besides sex, gender, age, race, ethnicity, socioeconomic, cultural background, and other life circumstances, a patient's own experience of his/her disease may influence the patient's personal tolerance for risk. As mentioned in the Benefit-Risk Guidance, patient views may be [...]

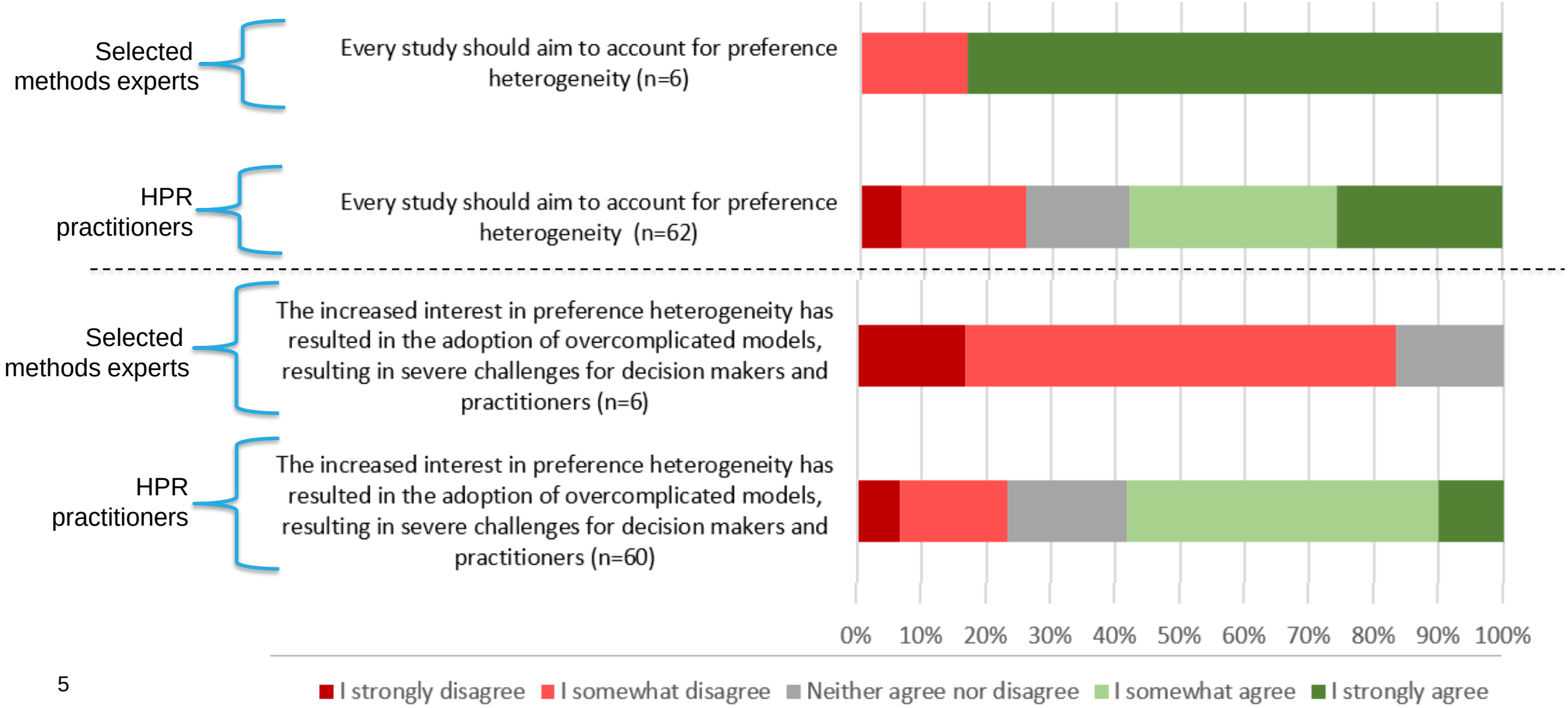
How can we account for preference heterogeneity?

- Multiple methods have been suggested over the last two decades on how to account for preference heterogeneity
- However, this is matched with absence of guidance on the topic

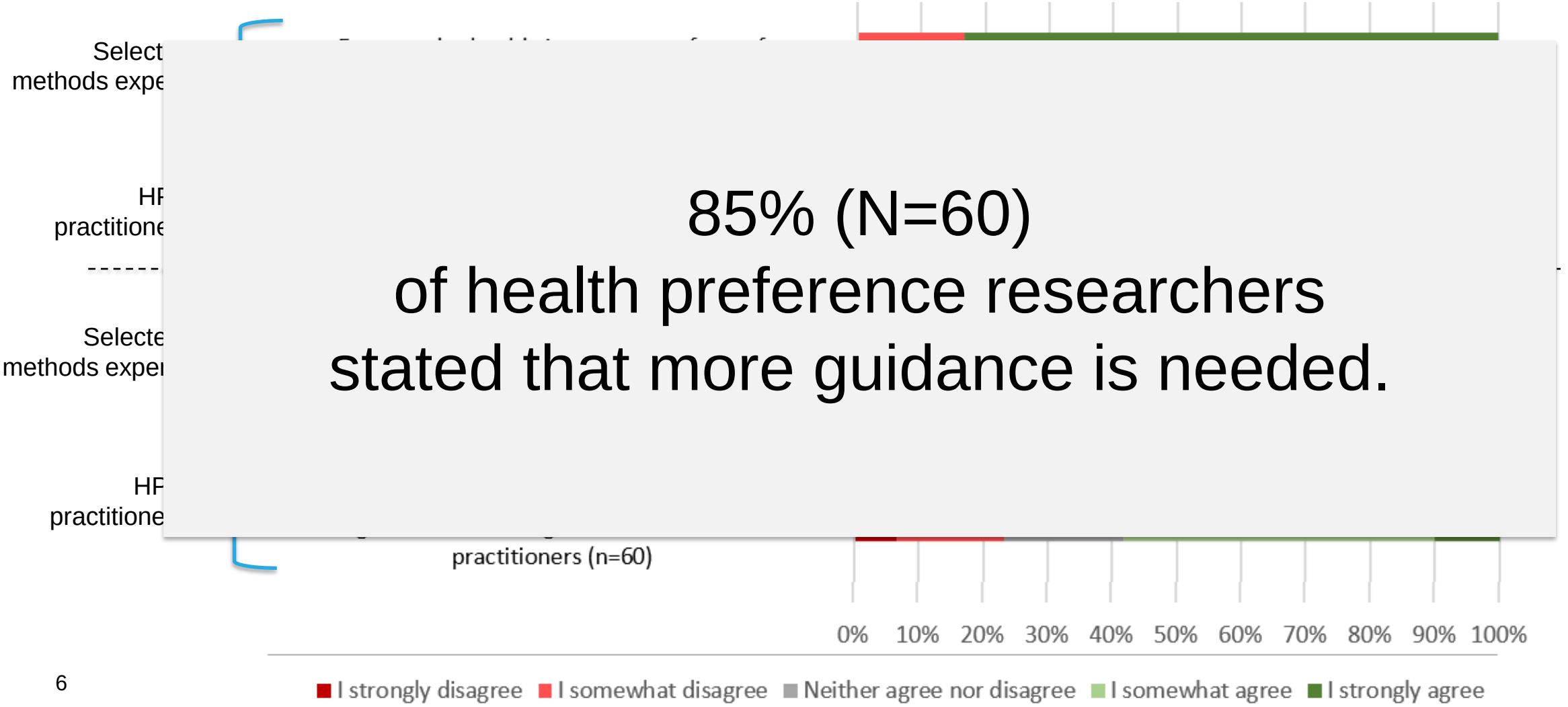
The ISPOR Health Preference SIG therefore conducted a working group to describe current practice



What do preference researchers think?



What do preference researchers think?



85% (N=60)
of health preference researchers
stated that more guidance is needed.

Introducing the Panel…

Marco Boeri, PhD, Health Preference Assessment, RTI Health Solutions, UK will discuss the current state of practice, by drawing on findings from the SIG working group.

Ellen Janssen, PhD, Benefit-Risk Assessment, The Janssen Pharmaceutical Companies of J&J, USA will outline the relevance of understanding preference heterogeneity for decision making and outline gaps in current practice.

Prof. Stephane Hess, PhD, Choice Modelling Centre, University of Leeds, Leeds, UK will discuss best-practice in accounting for preference heterogeneity and how to overcome practical challenges.

Marco Boeri

1

The current state-of practice

Considerations

- Choice models with no preference heterogeneity ➔ Assume everyone is the same
  - Strong assumption
  - It is possible to account for preference heterogeneity and explain preference heterogeneity (either / both)
- What is currently going on in practice?
  - Type of heterogeneity, model specification and distribution used, tests used…
  - How do we use the results in conclusions and policy implications?

- In the SIG working group we explored:
  - The common practice and how it developed over time (lit. review)
  - The views of “experts” and “users” of DCE data about preference heterogeneity (survey)

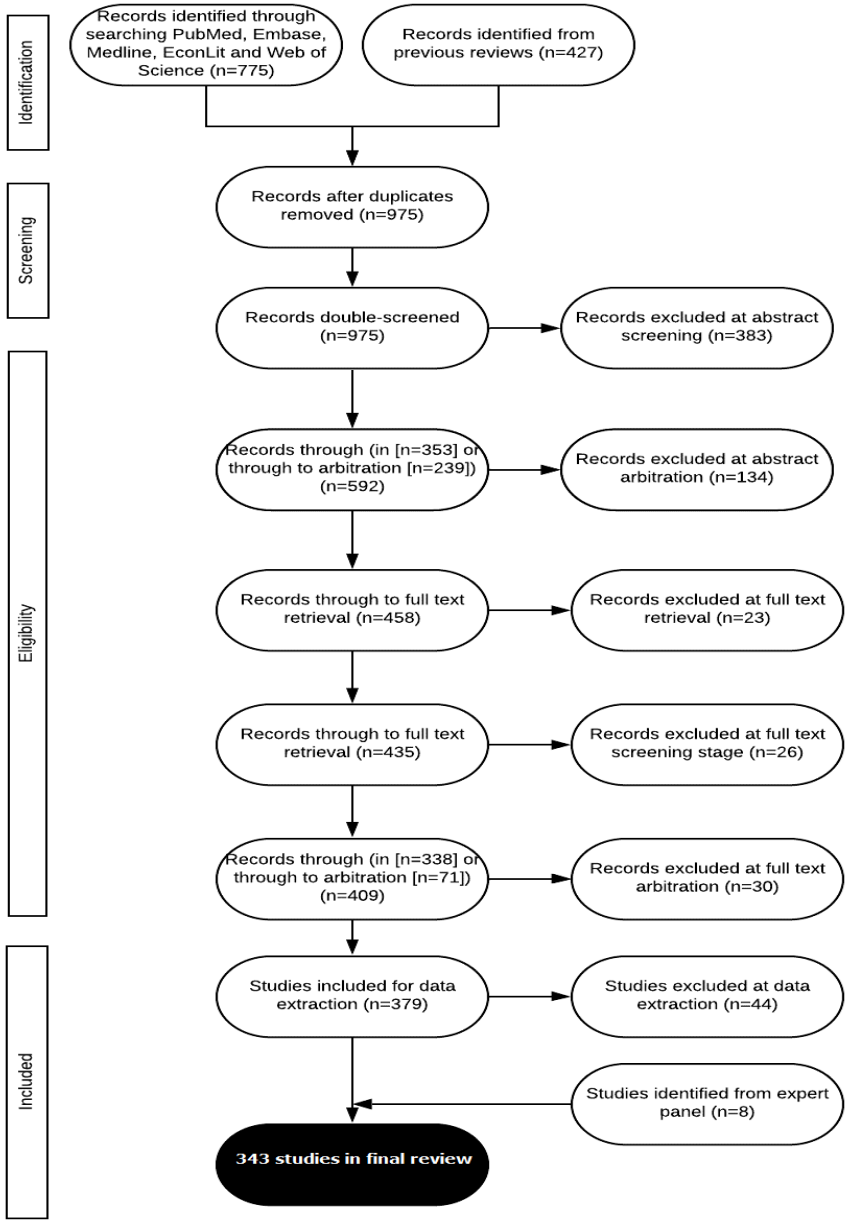
Literature Review – Inclusion / Exclusion Criteria

Inclusion	Exclusion

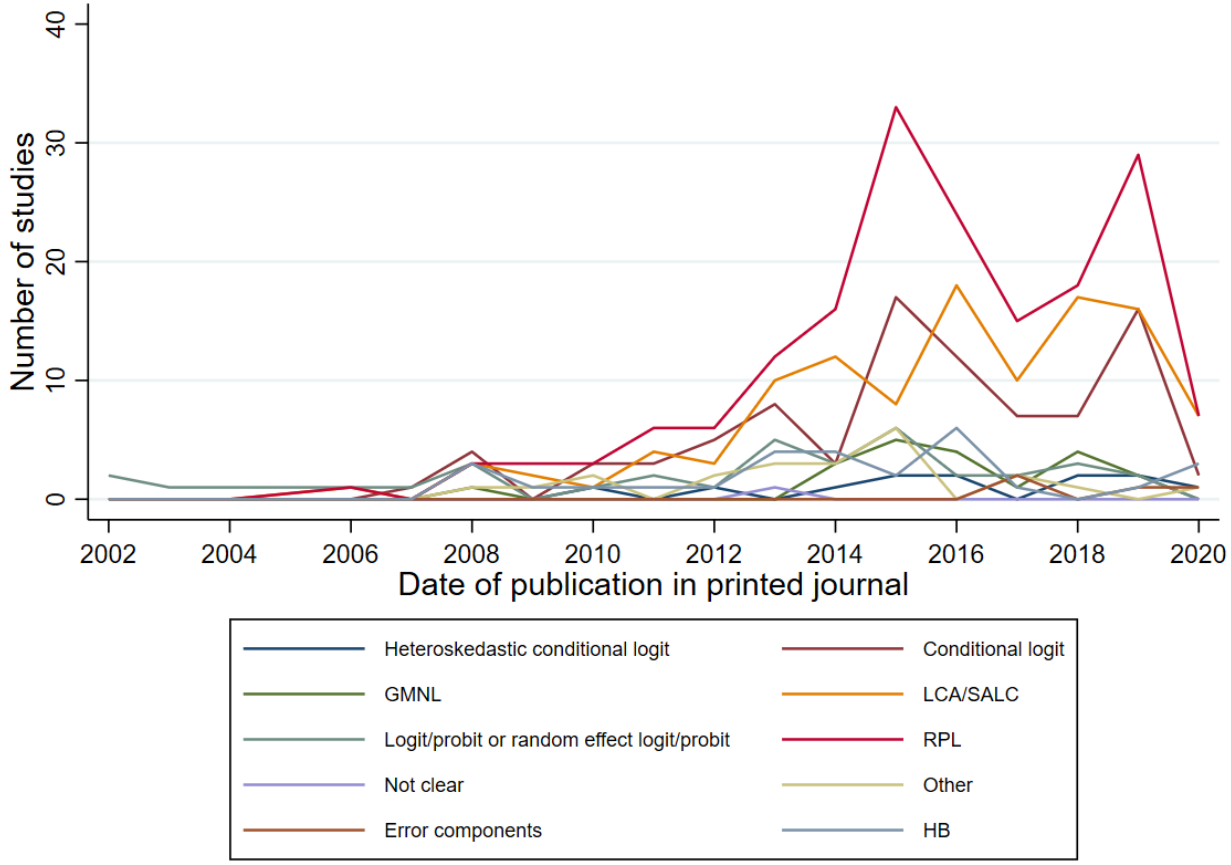
• Peer-reviewed papers, written in English
• Published 01Jan 2000 – 15MAR2020.
• Literature that falls under the Random Utility Model (RUM)
• Discrete choice experiments on health and healthcare, such as health valuation studies, treatment studies, and structure/policy studies (e.g., examine job preferences of health workers - physicians, medical students, and nurses).
• Includes analyses of preference heterogeneity (including explained and unexplained heterogeneity).
  • The unexplained heterogeneity will cover all finite and continuous mixture models with and without covariates effects. Hybrid choice model studies will also be included.
  • Explained heterogeneity will include all interaction studies, i.e., data are pooled.
	• Studies that analyze preference of food (e.g., high sugar), transportation (road safety), and environment (air quality control) that may or may not be related to health, unless addressing health and healthcare audience (health, health economics, or methodological journal).
	• Studies that focus on choice heterogeneity only to evaluate heuristic (e.g., attribute non-attendance), information processing (i.e., differences in utility function), and data mining perspectives.
	• Studies that stratify the data with separate analyses, i.e., data are not pooled (e.g., multiple countries or studies comparing patients and physicians).

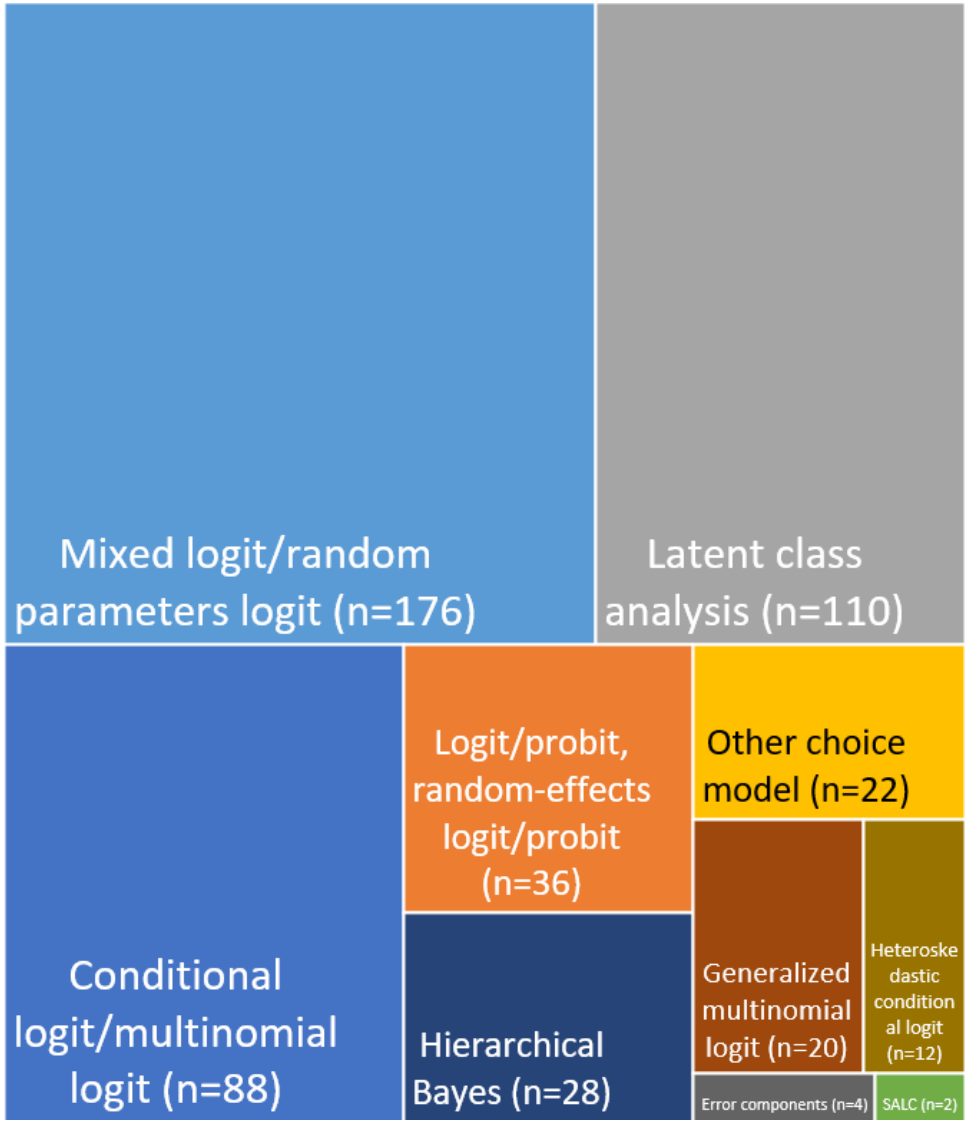
Literature Review - Results

- The search:
  - 1,202 potential titles reduced to 975 (duplicates).
  - 975 abstracts were double screened,
  - 435 full-length texts were screened
  - 379 were included in data extraction
  - A further 44 studies were excluded during data extraction
  - Other 8 papers were identified by working group members
  - In total, 343 studies were included in the final review.
  - Figure 2 shows the PRISMA diagram.



Literature Review - Highlights





Literature Review - Highlights

- Modelling preferences:
  - Model specification (continuous only; categorical only; both continuous and categorical)
    - MNL (n=3; n=25; n=26), Logit/probit/random effects (n=1; n=14; n=7)
    - Nested logits (n=0; n=1; n=3), Error components (n=0; n=3; n=1)
    - RPL (n=7;n=65;n=69), of which hierarchical Bayes (n=1; n=5; n=3)
    - LC analysis (n=3; n=30; n=31)
    - *Overall (n=10; n=107; n=107)*
  - Testing for preference heterogeneity
    - T statistic (n=9, 2.6%), Wald (n=16, 4.7%), F statistic (n=5, 1.5%)
    - Likelihood ratio test (n=34, 9.9%), Lagrange multiplier (n=2; 0.6%), Hausman (n=1, 0.3%)
    - Other test (15, 4.4%), Unclear (n=40; 11.7%)
  - Use of preference heterogeneity in conclusion/policy implications
    - No (n=50; 14.6%)
    - A little (n=158, 46.1%)
    - Yes, extensively (n=135; 39.4%)

Literature Review - Highlights

- Account and explain preference heterogeneity:
  - Use of sample characteristics to explain preference heterogeneity
    - MNL (49 out of 88)
    - RPL (incl HB) and GMNL (95 out of 205)
    - LC analysis including sample characteristics (16 out of 112)
  - Distributions used in RPL models
    - Normal (n=111); lognormal (n=15); triangular (n=2)
    - Others (n=5)
    - 10 studies report correlated distributions (unsure if fully correlated RPL)
  - Number of classes in Latent class analysis (N = 112)
    - Average num of classes = 3.0 (min= 2; max=6)
    - AIC (n=57); 3AIC (n=5); AIC (n=68); CAIC (n=21)

Food for Thought

- Many approaches in practice, no guideline
  - There are multiple concerns emerging that should be addressed
  - How to deal with the challenges in practice should be defined, discussed, addressed
  - Need to find a common way to deal with preference heterogeneity and possibly model preference data from DCE
- Many models do not explain preference heterogeneity (subgroups), but only account for it (RPL or LC without resp. characteristics included)
- Some confusion on models and what they do
  - RPL vs. hierarchical Bayes
  - Scale and preference heterogeneity (i.e. scaled LC, GMNL…)
- Most complete mixed logit model (fully correlated) rarely used
  - Too many parameters (use of categorical coded par.)
  - Too few observations
- How do we (should we) use preference heterogeneity in our conclusion and policy discussions/implications?

Ellen Janssen

2

Practical Considerations

Disclosures

- Ellen Janssen is an employee of Janssen Research and Development, LLC.
- The views and opinions expressed in this presentation are those of the individual presenter and do not necessarily reflect the views of Janssen Research & Development.

Considerations

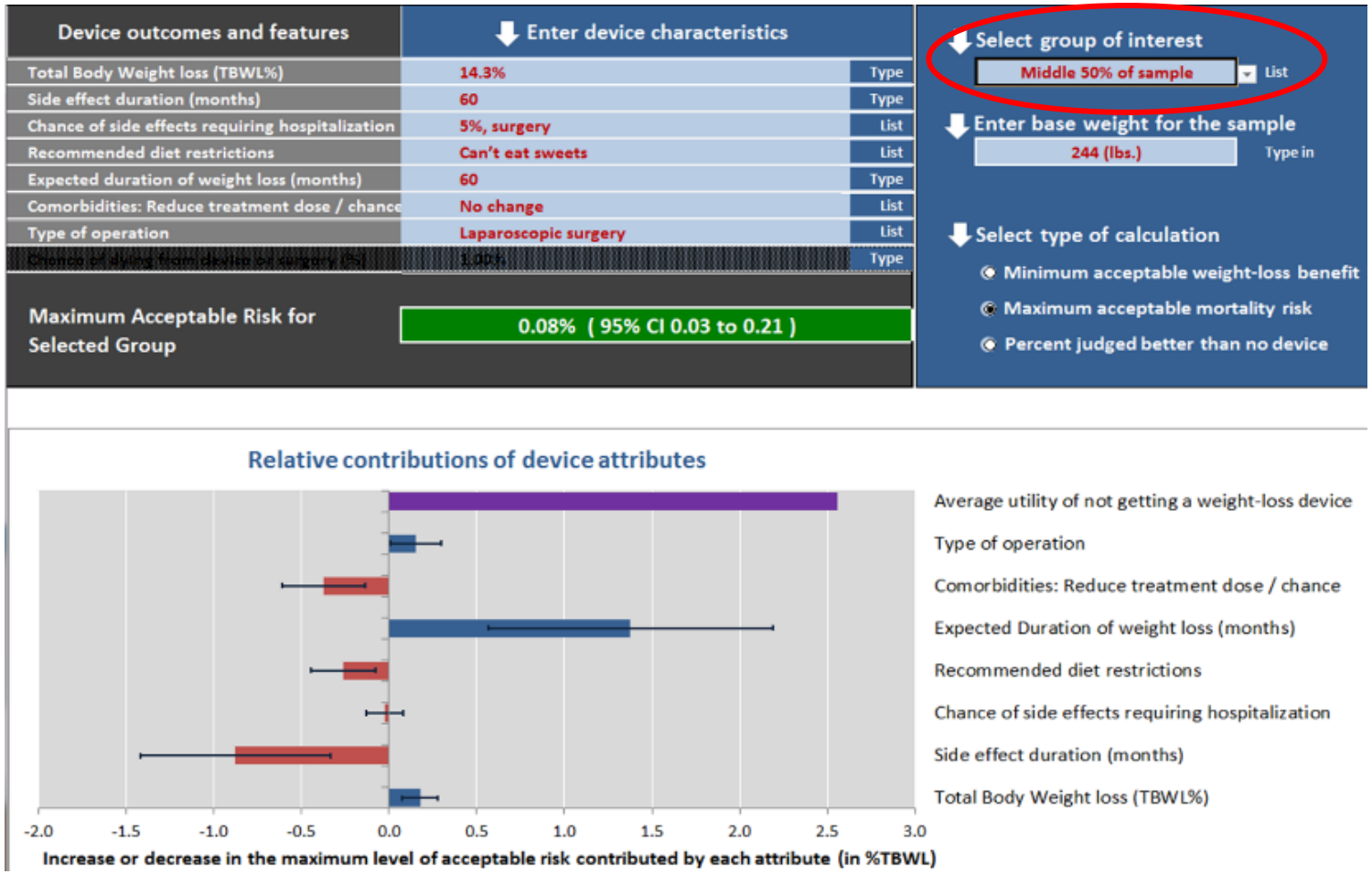
- Ultimate goal of preference studies is to help patients receive appropriate treatments that are in line with their treatment goals:
  - Include the patient perspective in healthcare decision making
  - Improve satisfaction and uptake
  - Improve quality of life
  - Improve health outcomes

- Preference heterogeneity is critical in helping us gain insights that will affect the availability, access, uptake, and satisfaction with treatment
- If preference heterogeneity did not exist, there would be no need for preference studies

Preference heterogeneity and HC decisions

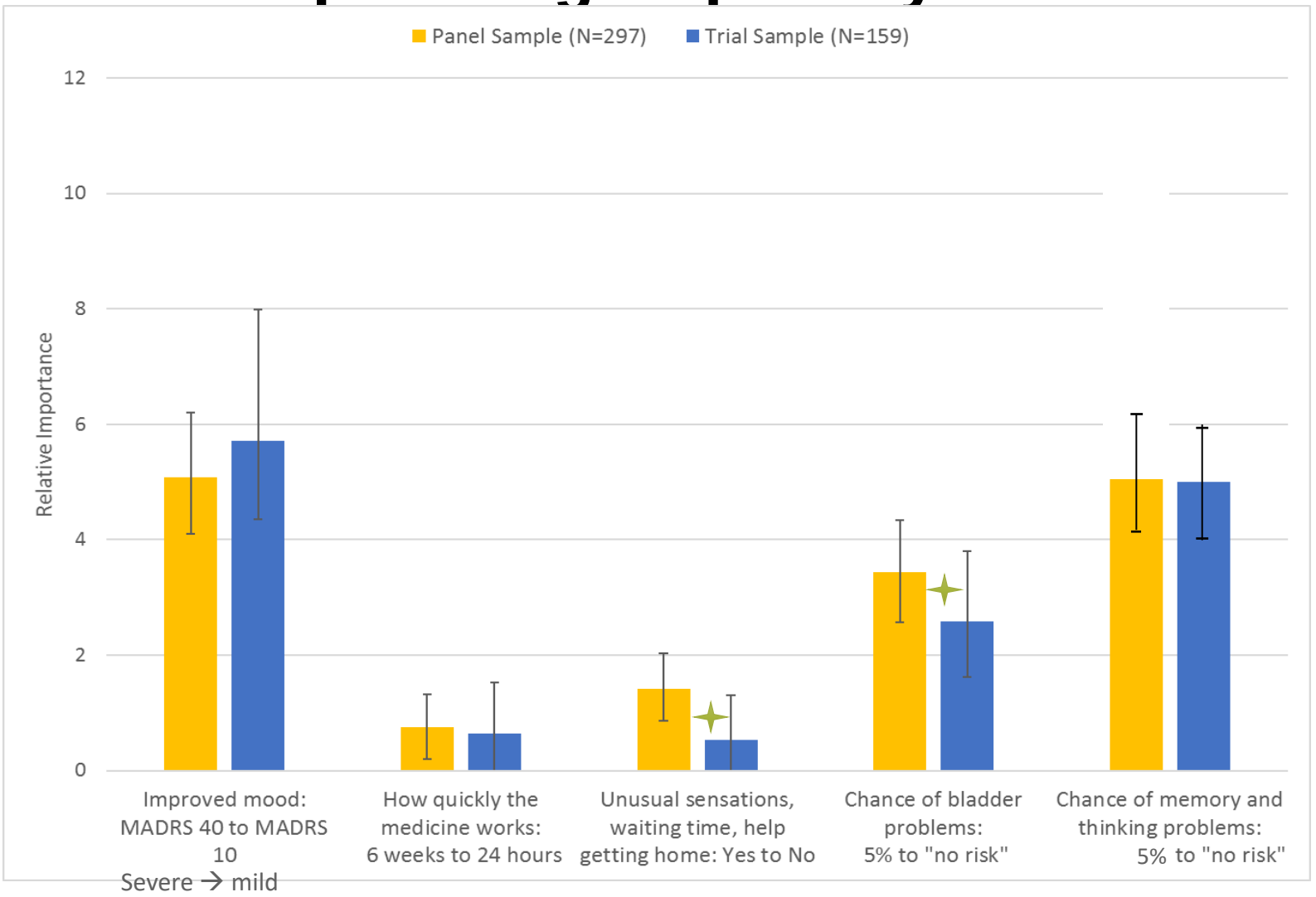
- FDA developed a stated-preference survey to elicit benefit-risk preferences for multiple attributes of hypothetical obesity devices
- Used the results in evaluating the Maestro Rechargeable system
  - *Concluded that a group of patients would accept risks associated with this surgically implanted device for the amounts of weight loss expected to be provided by the device"*

But which patients are the most risk tolerant?



[1]Ho et al. Surgical Endoscopy. 2015. [2]http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm430223.htm

Different patient groups may have different preferences



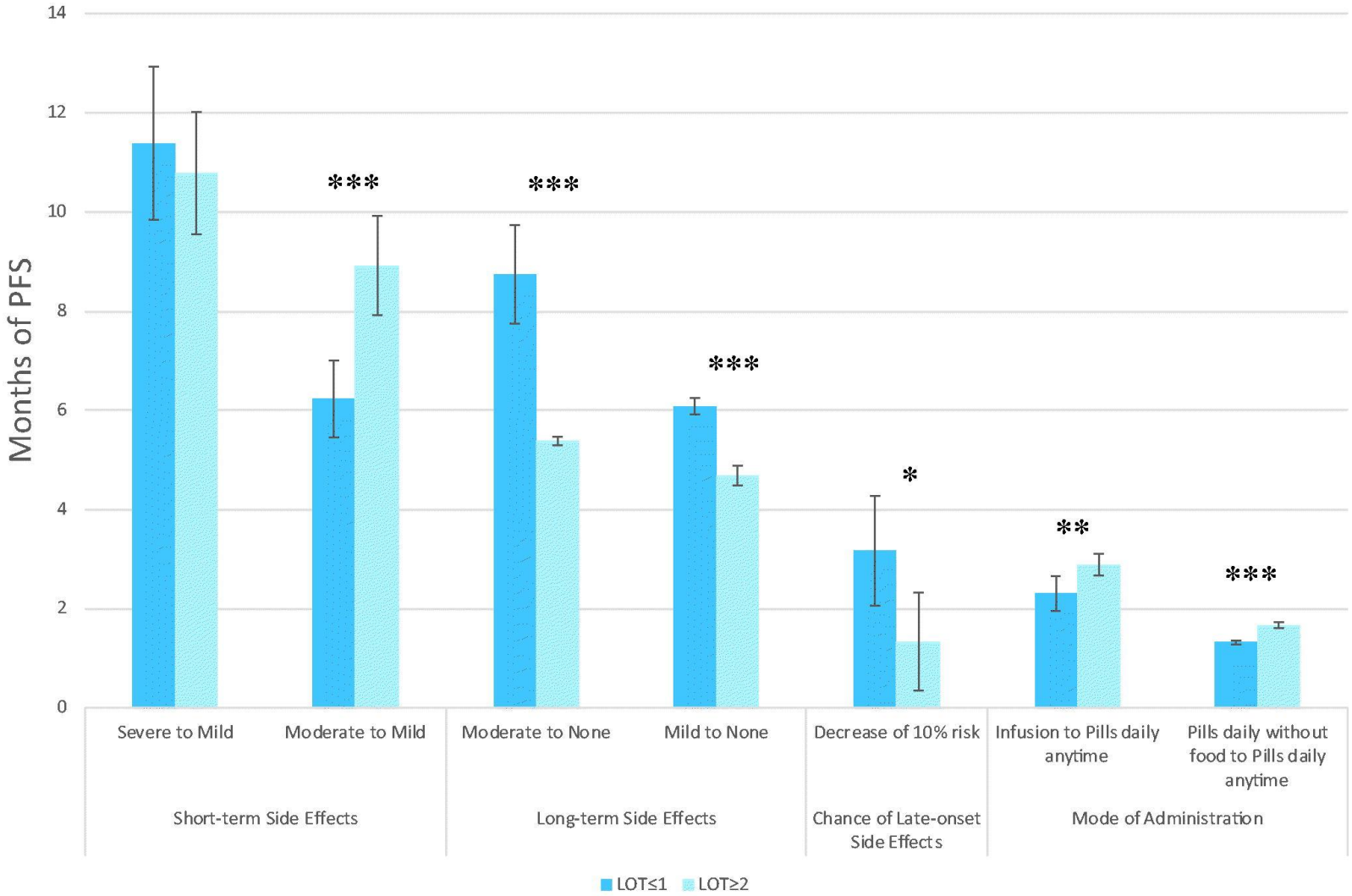
Why do we see these differences?

- Are patients that enroll in trials simply different from patients in online panels?
- Does treatment experience affect preferences?

✦ Statistically significant difference across samples

Fairchild et al. Patient preferences for ketamine-based antidepressant. Neurology, Psychiatry and Brain Research. 2020.

How does treatment experience affect preferences?

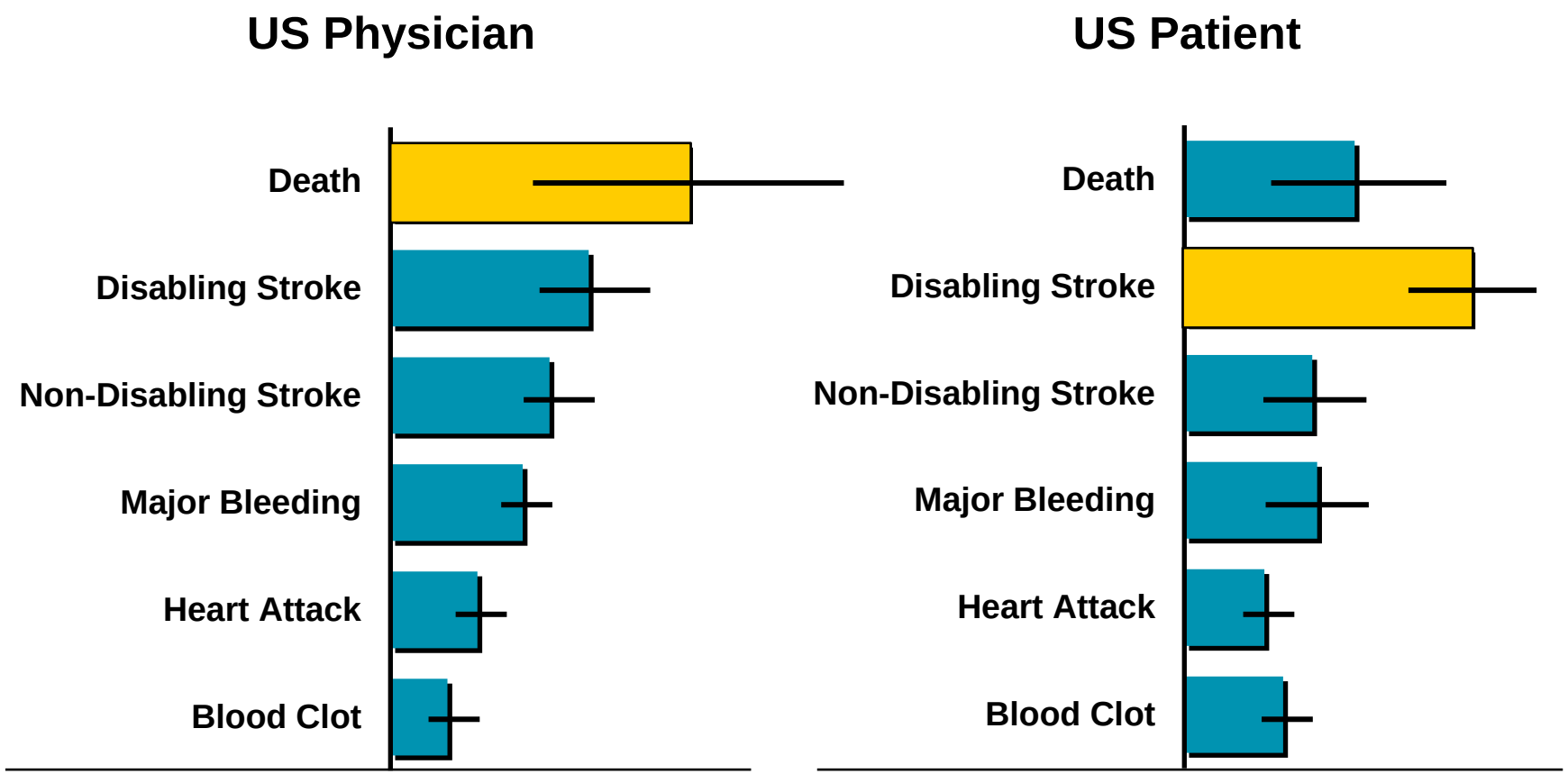


If treatment experience affects preferences, how may that affect decisions on availability and access to treatments?

Janse, Janssen, et al. Line of therapy and patient preferences regarding lung cancer treatment: a discrete-choice experiment. CRMO. 2021

Preference heterogeneity across stakeholder groups

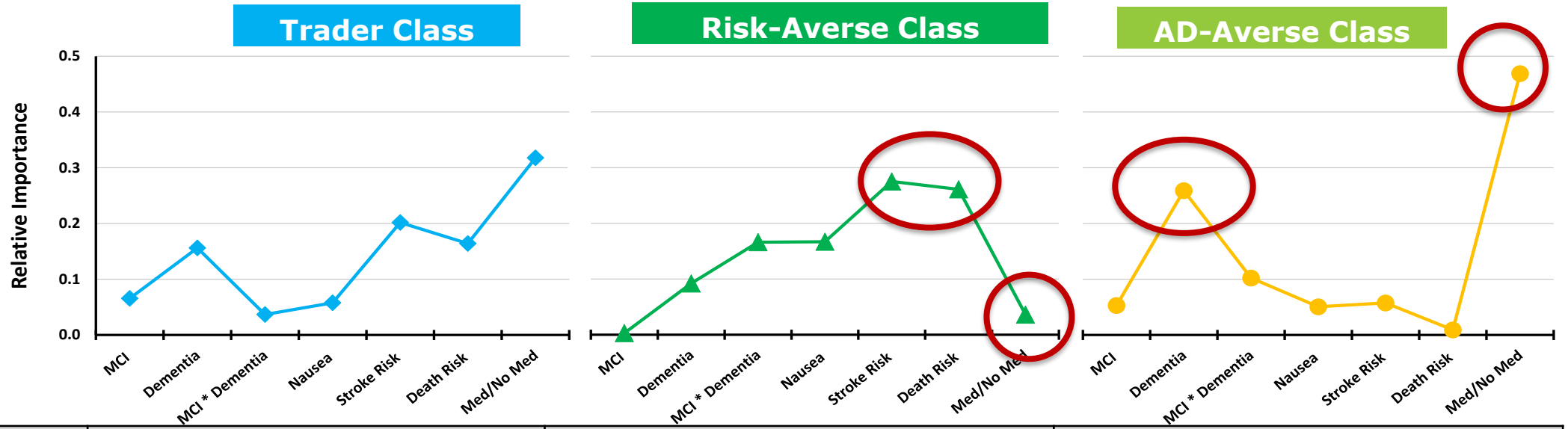
Preferences for Anticoagulants in Atrial Fibrillation



Can awareness into preference heterogeneity across stakeholders increase patient-centricity?

Levitan, Yuan, González, et al., ISPOR 18th Ann Int Mtg, 2013

What if preference heterogeneity is not determined by one observable characteristic?



Sample proportion	40%	33%	27%
Primary concerns	• Prefer medication	• Prefer no medication	• Strongly prefer medication
	• Trade off among all attributes	• More concerned about risks	• More concerned about efficacy
Statistically significant participant-level covariates	• Younger		• Older
	• Health problems reported	• No health problems reported	
	• Less likely to have AD caregiving experience	• Least likely to have AD caregiving experience	• Most likely to have AD caregiving experience
		• Assigned to 16-year version	• Assigned to 12-year version

Johnson et al. Something Is Better Than Nothing…. Value Health. 2019 Sep;22(9):1063-1069. doi: 10.1016/j.jval.2019.03.022.

Practical questions for identifying preference heterogeneity

- Which preference elicitation method to use?
- What sample size is needed to get meaningful insights on heterogeneity?
- How do we interpret model results?
  - Is the heterogeneity identified relevant? Is it actionable?
  - What if we don’t identify heterogeneity?
- How do we balance between model complexity and interpretability of results?

Stephane Hess

3

Best-Practice

Sources of heterogeneity

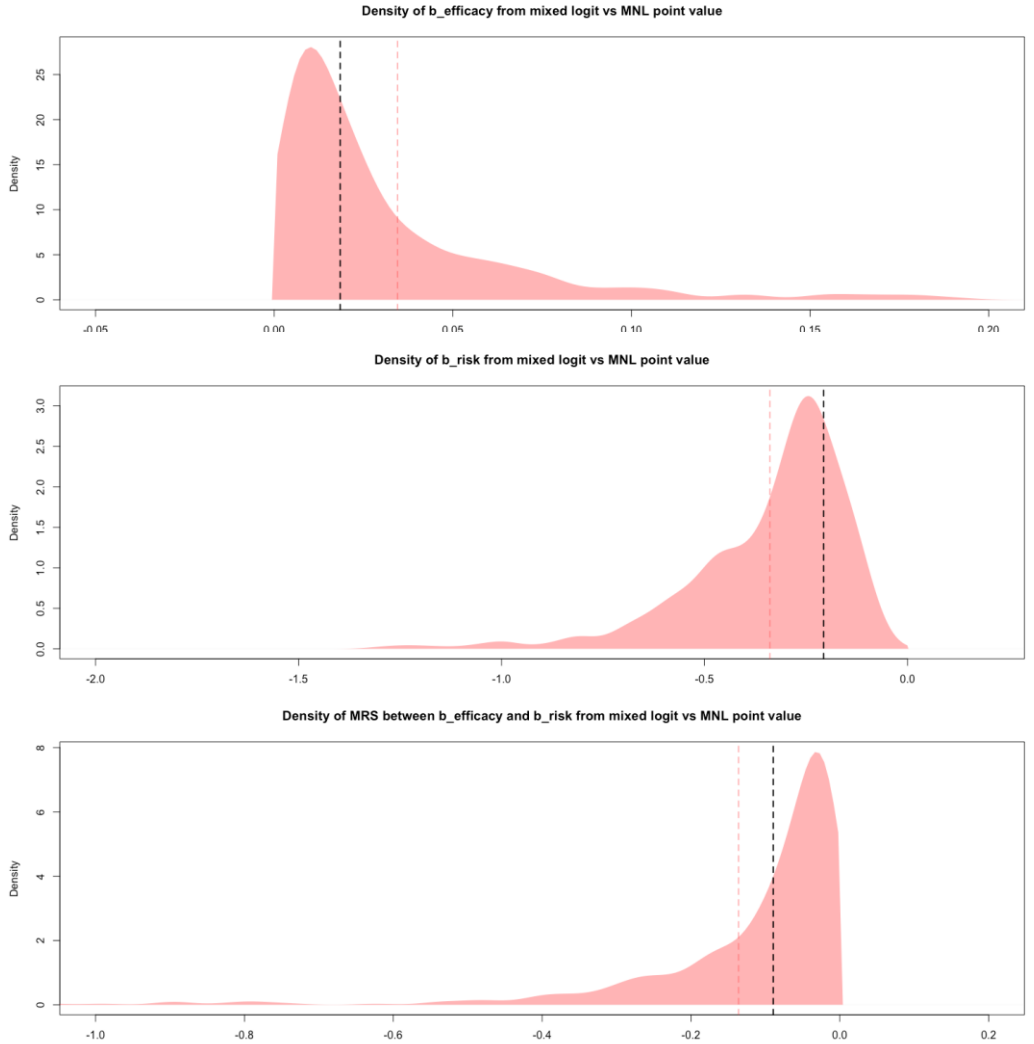
- Earlier speakers have made a clear case in support of the notion that heterogeneity exists
- This can relate to heterogeneity in:
  - sensitivities to individual attributes
    - some patients may care more about risk, others more about efficacy
  - baseline preferences for different options
    - some patients may have an overall preference for surgery, others for non-surgical treatment
  - noise
    - An analyst cannot completely explain behaviour, and the behaviour (from the analyst's perspective) may be more random for some patients than for others
- As we will see, in practice, these different sources are inter-related, and disentangling them is in often not possible

Why does heterogeneity matter

- Understanding that different patients have different sensitivities can be very valuable
  - Provides insights into why some people make specific choices
    - e.g. underlying health conditions reducing the likelihood of certain treatment options
  - Can ensure a more robust appraisal of cost vs benefits
    - e.g. if we know how relative importance of waiting time and risk varies across patients
  - Can help in the context of shaping a more efficient provision of patient services
    - e.g. make a variety of services available, where the mix is such that it can help cover the widest range of preferences
  - Can help shape strategies to nudge behaviour
    - e.g. identify drivers of vaccine hesitancy

But what if I'm only interested in an overall effect?

- Offering more than a single treatment plan may not be possible
- May thus conclude that only aggregate preferences matter
- Aggregate preferences from a model allowing for heterogeneity may be different (and less biased) than those from a model assuming homogeneity



There's another reason

- The reason that we are using random utility theory is that we recognise that utility theory is only an approximation of the real behavioural process
- We thus incorporate an error term in our utility functions

$$U_{i,n,t} = V_{i,n,t} + \varepsilon_{i,n,t}$$

- Deterministic part of utility is a function of attributes of the alternatives, characteristics of the individual, and estimated parameters

$$V_{i,n,t} = f(x_{i,n,t}, z_n, \beta)$$

- Now imagine that the sensitivities to attributes are distributed in the population

$$V_{i,n,t} = f(x_{i,n,t}, z_n, \beta_n)$$

- But we do not incorporate that, i.e., use a deterministic β

$$U_{i,n,t} = f(x_{i,n,t}, z_n, \beta) + \varepsilon^*_{i,n,t}$$

$$\varepsilon^*_{i,n,t} = f(x_{i,n,t}, z_n, (\beta - \beta_n)) + \varepsilon_{i,n,t}$$

- Error term becomes correlated with the deterministic part of utility, breaking one of the core assumptions of random utility models, leading to bias

Despite developing advanced models myself, I have three simple rules

Observation 1: complexity itself should not be the aim	Observation 2: always start with a simple model	Observation 3: explaining heterogeneity through covariates is better than throwing random distributions at a model
• Do not use an advanced model because it is complex, but use it if the complexity provides additional benefits • Too many studies use Mixed Logit (or more recently Hybrid choice) because they see this as a *requirement* to be published	• I have estimated thousands of models, on hundreds of datasets • I have NEVER strayed from my rule that the first model I estimate on a dataset is the simplest possible MNL model • Even before that, should analyse data in a non-parametric way	• Which result is more useful to a policy maker? ○ relative importance of efficacy compared to risk varies as a function of age and gender ○ relative importance follows a negative lognormal distribution • Choice modellers all too often (especially in health) focus on random het. first and foremost

- However, I have also never come across a dataset where random heterogeneity does not make a difference
  - Key aim should be to explain as much of the heterogeneity as possible through covariates, leaving only what you must to random heterogeneity

So how should we capture the heterogeneity?

- Step 1: study the data
- Step 2: deterministic heterogeneity
- Step 3: random heterogeneity

Step 1: study the data

- Before fast computers came along, this was part of the normal workflow
- Tabulate the data
  - Look at who chooses what in what settings
  - Do patients with specific characteristics choose certain treatments more often that others?
  - Do they avoid options with specific characteristics, etc?
- This might sound tedious, but it's easy to operationalize in an efficient manner
- Much preferable to the trial-and-error approach of blindly throwing dozens of covariates at a model

Step 2: deterministic heterogeneity

- Use the insights from step 1, alongside general intuition, to introduce deterministic heterogeneity
- DO NOT use a brute force approach
  - This involves interacting every patient covariate with every attribute
- Likely to lead to convergence issues as well as the majority of effects not being meaningful or attaining reasonable levels of significance
- Instead gradually build up complexity
- Make decisions on retaining interactions in the model jointly on statistical and behavioural/clinical grounds

Step 3: random heterogeneity

- This is the step that takes the least analyst time but most computational time
- Also the step likely to give the biggest mathematical gains
- But interpretation is lacking, and there is much more scope for mistakes
- Which parameters should be random?
  - if only some parameters are specified to be random, the model is likely to be inferior, and heterogeneity is likely to be misattributed
- What distributions should be used?
  - Too many people use Normal distributions
  - Is the use of a wrong distribution better than than not allowing for random heterogeneity?
- Should multivariate distributions be used?
  - i.e. is the heterogeneity likely to be correlated across attributes
- How should the model be estimated?
  - Classical vs Bayesian vs EM algorithm

What type of random heterogeneity?

- The big decision is between continuous random heterogeneity and discrete random heterogeneity (i.e. Latent Class)
- A latent class model is also a Mixed Logit model, but with discrete distributions
- Both types of models also allow an analyst to include deterministic heterogeneity (which they should!), but in a very different way
  - Mixed logit would still use interactions at the level of individual attributes
  - LC uses covariates in class allocation model
- Especially in a health context, where many attributes are categorical, I often see inherent advantages for LC

Common misunderstanding 1: Hierarchical Bayes (HB)

- HB is not a different model
- It is simply Bayesian estimation of a Mixed Logit model
- People talk about individual-level estimates, but there is no such thing
  - They are simply individual level posterior means, still derived from a sample level distribution
  - Can also get this from a model estimated using classical techniques

Common misunderstanding 2: scale heterogeneity

- Heterogeneity can relate to relative as well as absolute sensitivities
  - choices are more or less random for some people (from analyst's perspective)
- Clearly a possibility, but IMPOSSIBLE to disentangle from other heterogeneity
- This inability to disentangle the effects also applies to deterministic heterogeneity, so even allowing for scale differences between groups may just be picking up other heterogeneity
- For example, men might be less risk sensitive and risk might be most important attribute
  - Then scale heterogeneity might pick that up
- It's not a bad thing to allow for scale heterogeneity. But must never think you can disentangle it, or that you should try to
- Safest approach is to allow for heterogeneity in all individual parameters, including correlation between them

How important is it to know the source of the heterogeneity?

- Clearly very important in theory
- So easy to misattribute!
- Any specification is an approximation! The model itself is an approximation
- But if only allowing for some effects, cannot attribute source!
- I am a great believer in gradually building up complexity to avoid this as much as possible
- Simple illustration using a binary vaccine choice setting (simulated data)
  - One vaccine has greater efficacy
  - The other has lower side effects

Easy to misattribute heterogeneity to wrong attribute

Number of individuals	500		500		500		500	
Number of modelled outcomes	5000		5000		5000		5000	
Estimated parameters	2		3		3		4	
LL(final)	-2167.405		-2083.096		-2061.137		-2060.733	
Adj.Rho-square (0)	0.374		0.3981		0.4044		0.4042	
AIC	4338.81		4172.19		4128.27		4129.47	
BIC	4351.84		4191.74		4147.82		4155.54	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficiency	0.0199	11.78	0.0279	16.64	0.0202	14.1	0.0212	11.78
risk	-0.3651	-25.18	-0.3647	-26.37	-0.2958	-23.36	-0.3025	-20.12
efficacy interaction for men			-0.021	-12.79			-0.0029	-0.99
risk interaction for men					-0.2033	-13.33	-0.1816	-6.57

The same happens in Mixed Logit (so best to try all parameters random!)

Number of individuals	500		500		500		500	
Number of modelled outcomes	5000		5000		5000		5000	
Estimated parameters	3		3		4		5	
LL(final)	-2160.457		-2150.033		-2150.017		-2149.356	
Adj.Rho-square (0)	0.3758		0.3788		0.3785		0.3784	
AIC	4326.91		4306.07		4308.03		4308.71	
BIC	4346.47		4325.62		4334.1		4341.3	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficacy (mean for lognormal)	-0.9883	-23.84	-0.9401	-22.92	-0.94	-22.92	-0.9301	-21.87
efficacy (sd for lognormal)	0	NA	0.262	8.54	0.2614	8.55	0.3277	5.27
risk (mean for lognormal)	-3.9673	-40.54	-3.8255	-48.46	-3.8257	-48.52	-3.7959	-46.23
risk (sd for lognormal)	-0.3579	-6.19			-0.0316	-0.83	-0.0928	-0.97
cholesky term							0.158	1.29

Can even find scale heterogeneity that isn't there, or random heterogeneity that isn't (MNL wins here!)

Number of individuals	500		500	
Number of modelled outcomes	5000		5000	
Estimated parameters	4		3	
LL(final)	-2060.733		-2077.792	
Adj.Rho-square (0)	0.4042		0.3996	
AIC	4129.47		4161.58	
BIC	4155.54		4181.14	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficacy	0.0212	11.78	0.0126	10.41
risk	-0.3025	-20.12	-0.255	-19.99
efficacy interaction for men	-0.0029	-0.99		
risk interaction for men	-0.1816	-6.57		
scale interaction for men			2.087	17.45

Number of individuals	500		500	
Number of modelled outcomes	5000		5000	
Estimated parameters	5		7	
LL(final)	-2149.356		-2060.553	
Adj.Rho-square (0)	0.3784		0.4034	
AIC	4308.71		4135.11	
BIC	4341.3		4180.73	
	estimate	Rob.t-ratio(0)	estimate	Rob.t-ratio(0)
efficiency (mean for lognormal)	-0.9301	-21.87	-1.1901	-23.81
efficiency (sd for lognormal)	0.3277	5.27	0.1249	1.11
risk (mean for lognormal)	-3.7959	-46.23	-3.8435	-45.49
risk (sd for lognormal)	-0.0928	-0.97	-0.0011	-0.2
cholesky term	0.158	1.29	0.107	1.04
efficiency interaction for men			-0.003	-1.03
risk interaction for men			-0.1809	-6.53

Data requirements

- Capturing heterogeneity places additional demands on data
- Need richer data, with trade-offs that allow us to observe different outcomes for different people
- Heterogeneity in sensitivities alone may not lead to different outcomes if stimuli are not strong enough
  - One person may care more about risk than another person, but not enough to make a different choice given the scenarios presented in our data
  - And that would prevent us from recognising the impact of the heterogeneity in forecasting
- Focus on categorical variables in many health studies makes capturing heterogeneity especially hard
  - Proliferation of parameters
  - And possibly lack of statistical support for continuous distributions

How do we use results from models allowing for heterogeneity?

- Having heterogeneity in our models is valuable for both valuation work (e.g. cost/benefit analysis) and forecasting
- We've gone to a lot of trouble, so let's use the results in the right way
- Should really not just use the mean outcomes
- Key advantage of deterministic heterogeneity is for sample enumeration
  - Our samples are never fully representative, plus populations change, and we may want to predict choices in a different population, etc
  - Can adjust the sample, and interactions with covariates will then lead to different outcomes
  - This is not possible if only using random heterogeneity!

Summary

- Heterogeneity is a crucial component of behaviour
- Allowing for it in choice models has many benefits
  - Even reduced bias in cases where just a single value is needed in output
- But important to recognise that:
  - Source of heterogeneity is often difficult to explain and there is substantial risk of misattribution
  - Linking as much of the heterogeneity as possible to observable characteristics is always best
  - Gradually building up model complexity is key
- Greater insights are also possible in model application (e.g. forecasting), but again, care is required

Questions?

www.ispor.org

ISPOR

Improving healthcare decisions

Thank you!

For questions on content: