

# Live Session 5: Explainability

Session will begin at 16:00

Dr Zak Varty

# Welcome!

## This Session:

1. Group Discussion: Explainability Methods (15 mins)
2. Feedback from Groups (15 mins)
3. Group Discussion: LIME paper (15 mins)
4. Feedback from groups (15 mins)

# Explainability Methods

# ICE, PDP, LIME, SHAP

In breakout rooms discuss:

- What does this method show?
- When might this be a good method to use?
- What are the drawbacks/limitations of this method?

# Group 1 - ICE

- What does this method show?
- When might this be a good method to use?
- What are the drawbacks/limitations of this method?

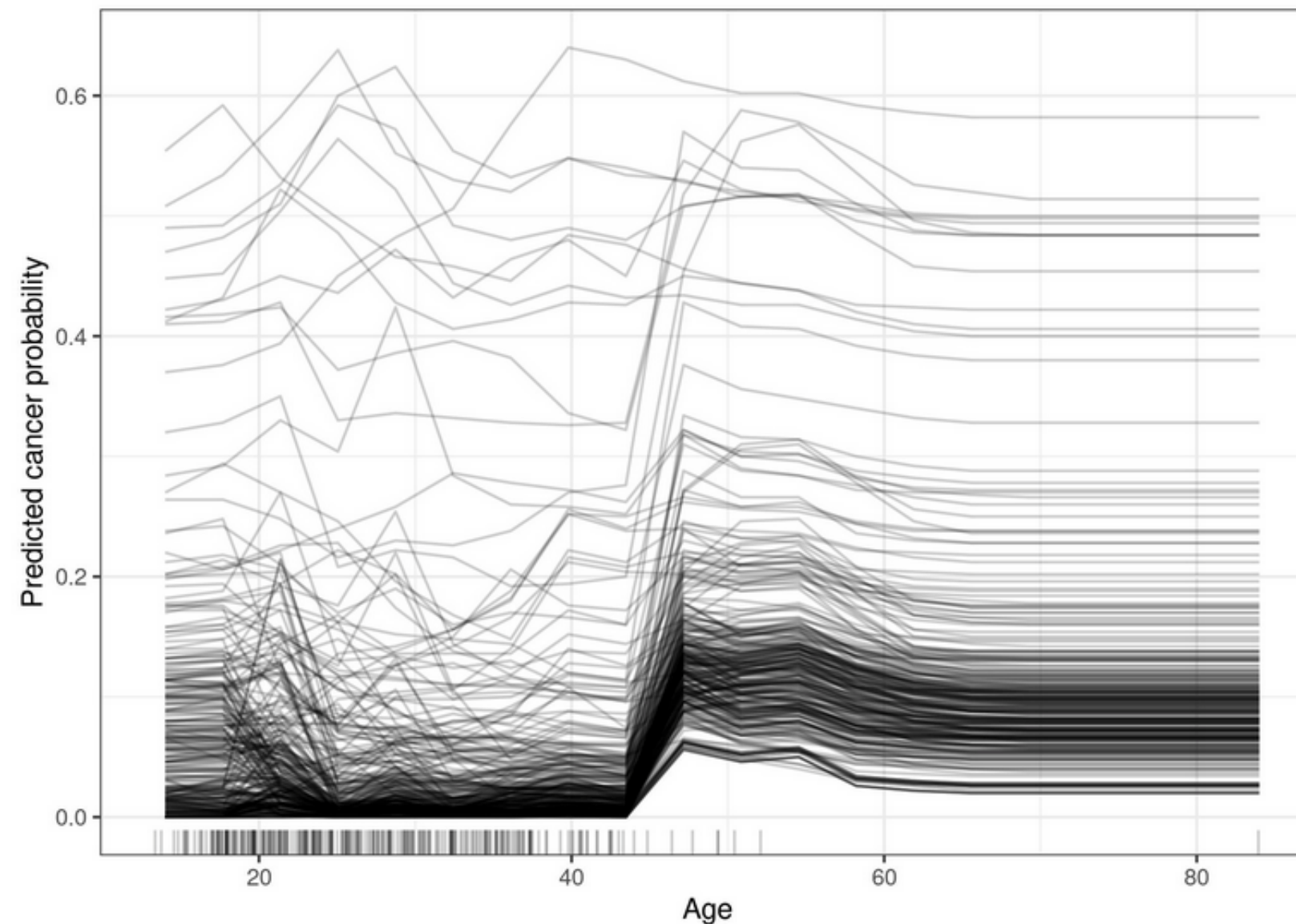


FIGURE 9.1: ICE plot of cervical cancer probability by age. Each line represents one woman. For most women there is an increase in predicted cancer probability with increasing age. For some women with a predicted cancer probability above 0.4, the prediction does not change much at higher age.

# Group 2 - PDP

- What does this method show?
- When might this be a good method to use?
- What are the drawbacks/limitations of this method?

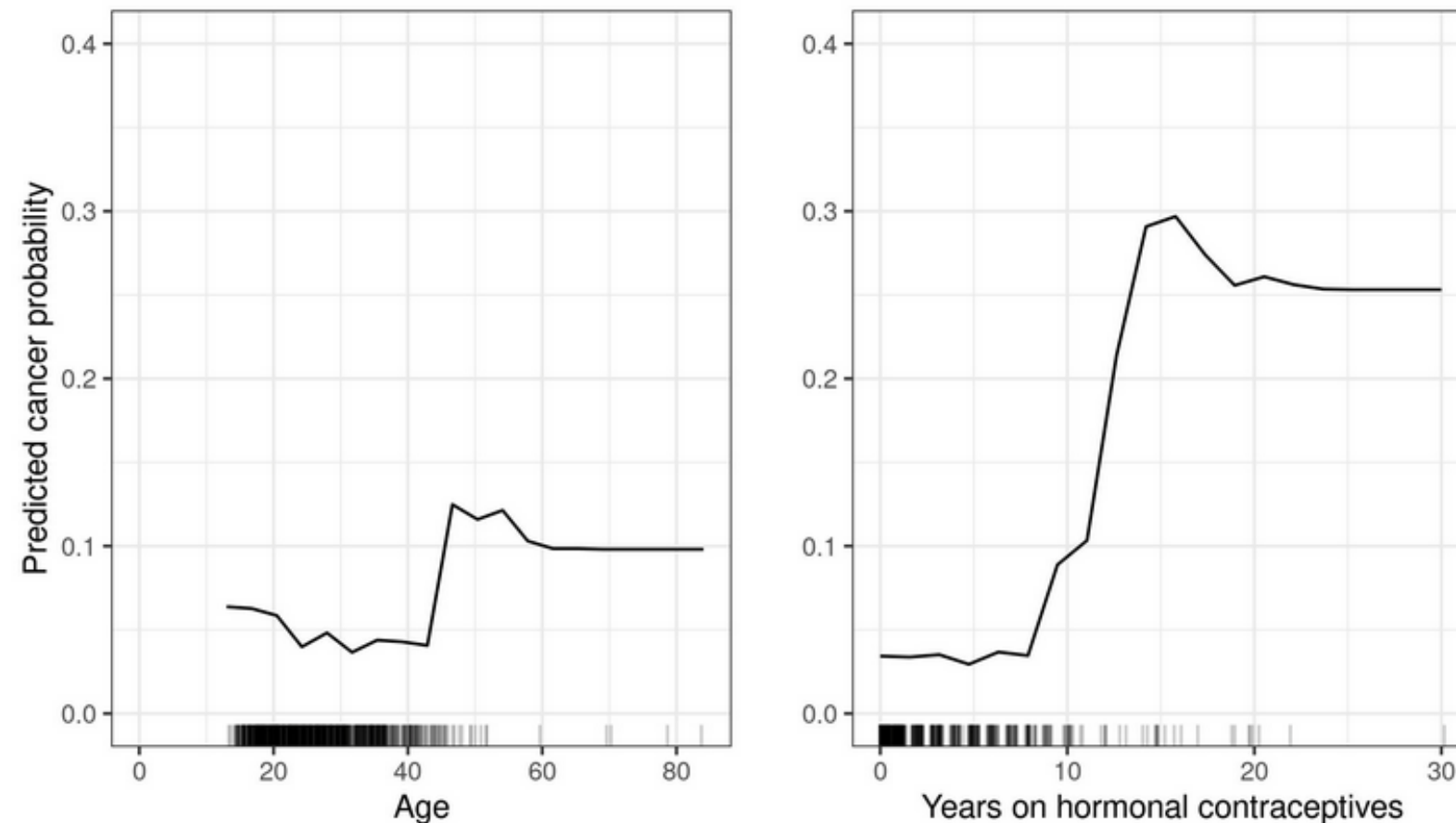
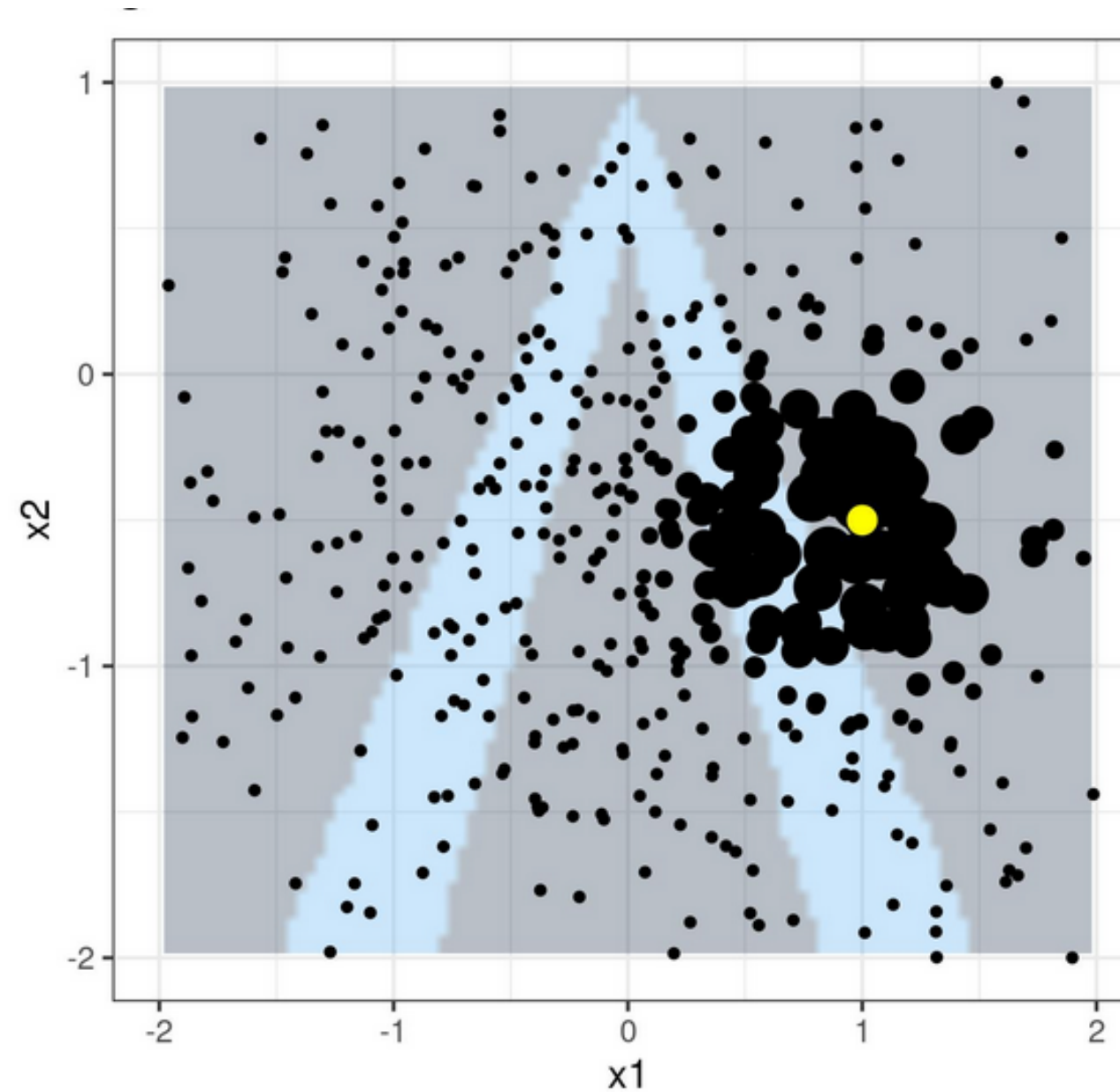


FIGURE 8.3: PDPs of cancer probability based on age and years with hormonal contraceptives. For age, the PDP shows that the probability is low until 40 and increases after. The more years on hormonal contraceptives the higher the predicted cancer risk, especially after 10 years. For both features not many data points with large values were available, so the PD estimates are less reliable in those regions.

# Group 3 - LIME

- What does this method show?
- When might this be a good method to use?
- What are the drawbacks/limitations of this method?



# Group 4 - SHAP

- What does this method show?
- When might this be a good method to use?
- What are the drawbacks/limitations of this method?

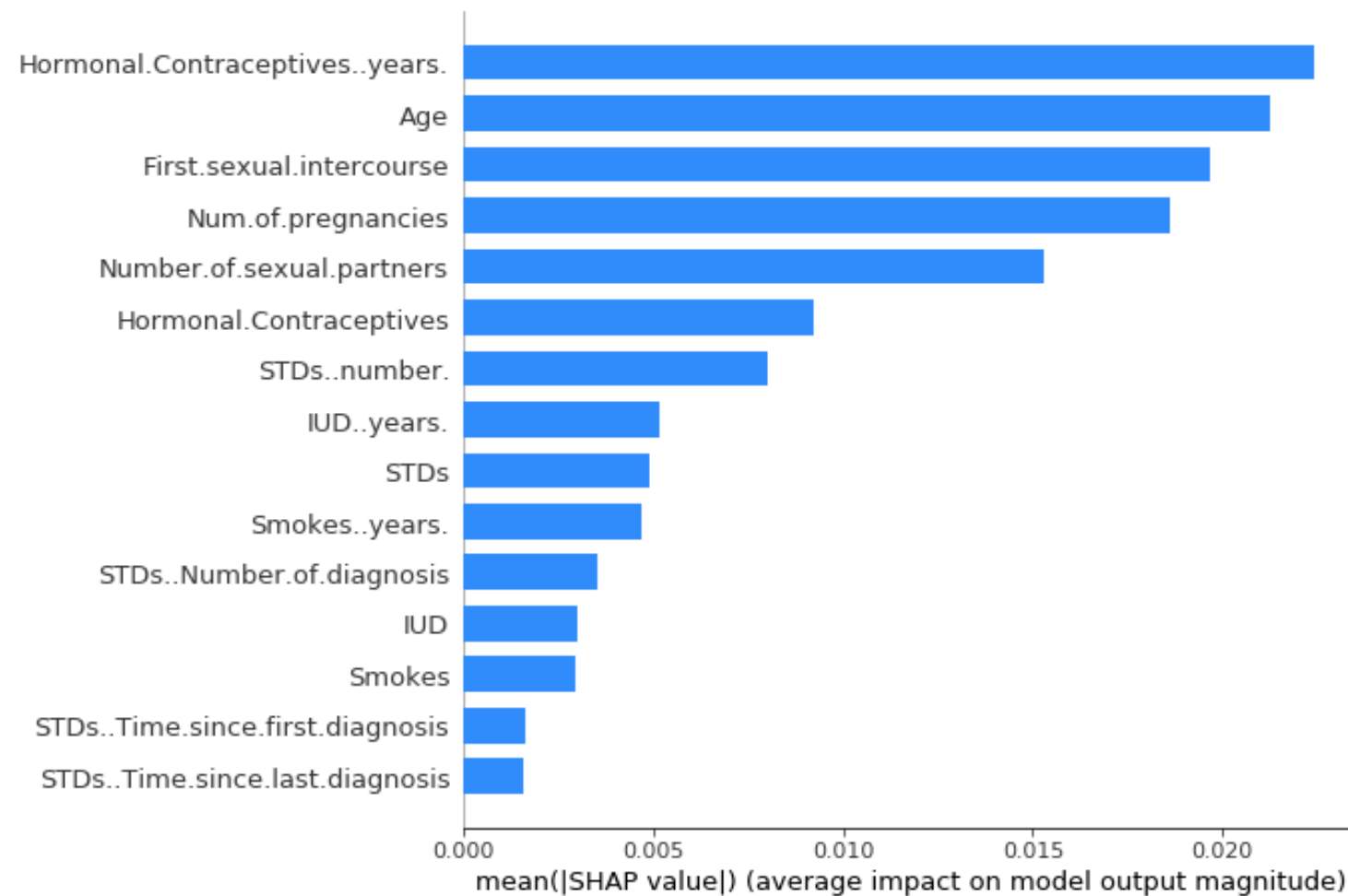
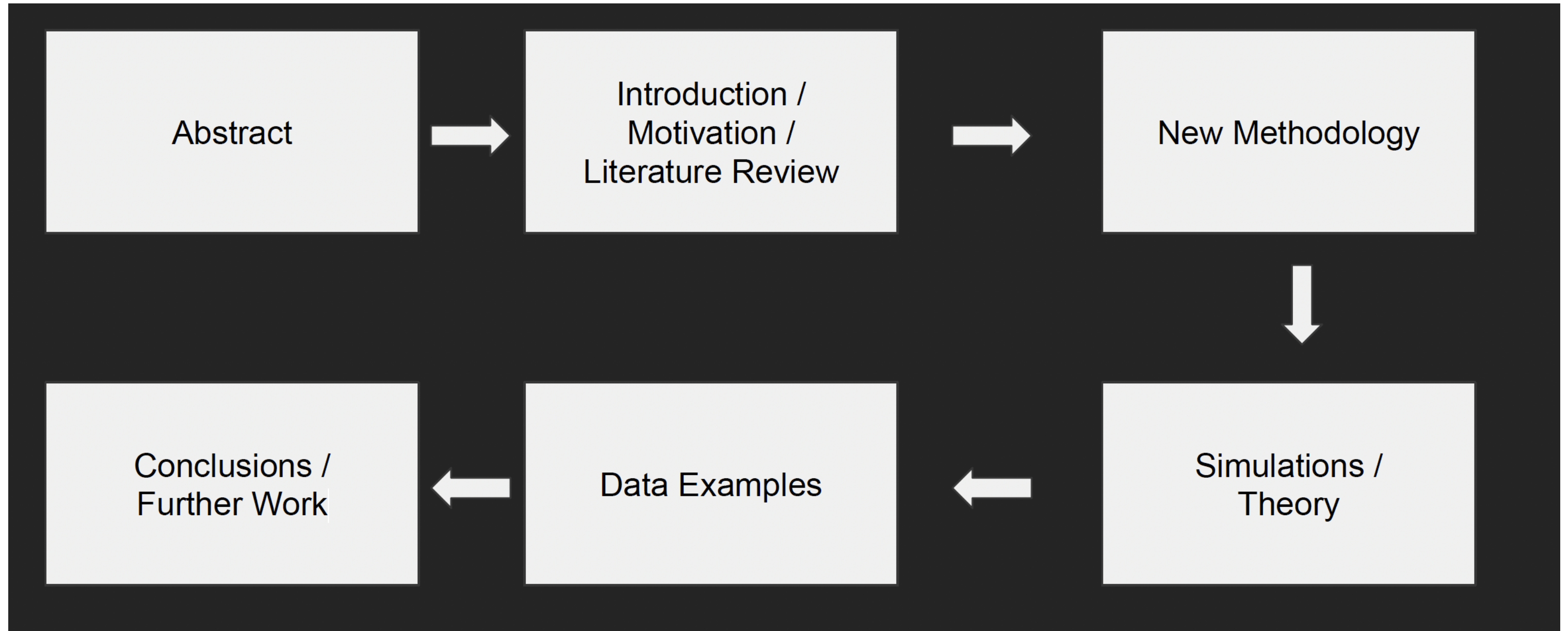


FIGURE 9.25: SHAP feature importance measured as the mean absolute Shapley values. The number of years with hormonal contraceptives was the most important feature, changing the predicted absolute cancer probability on average by 2.4 percentage points (0.024 on x-axis).

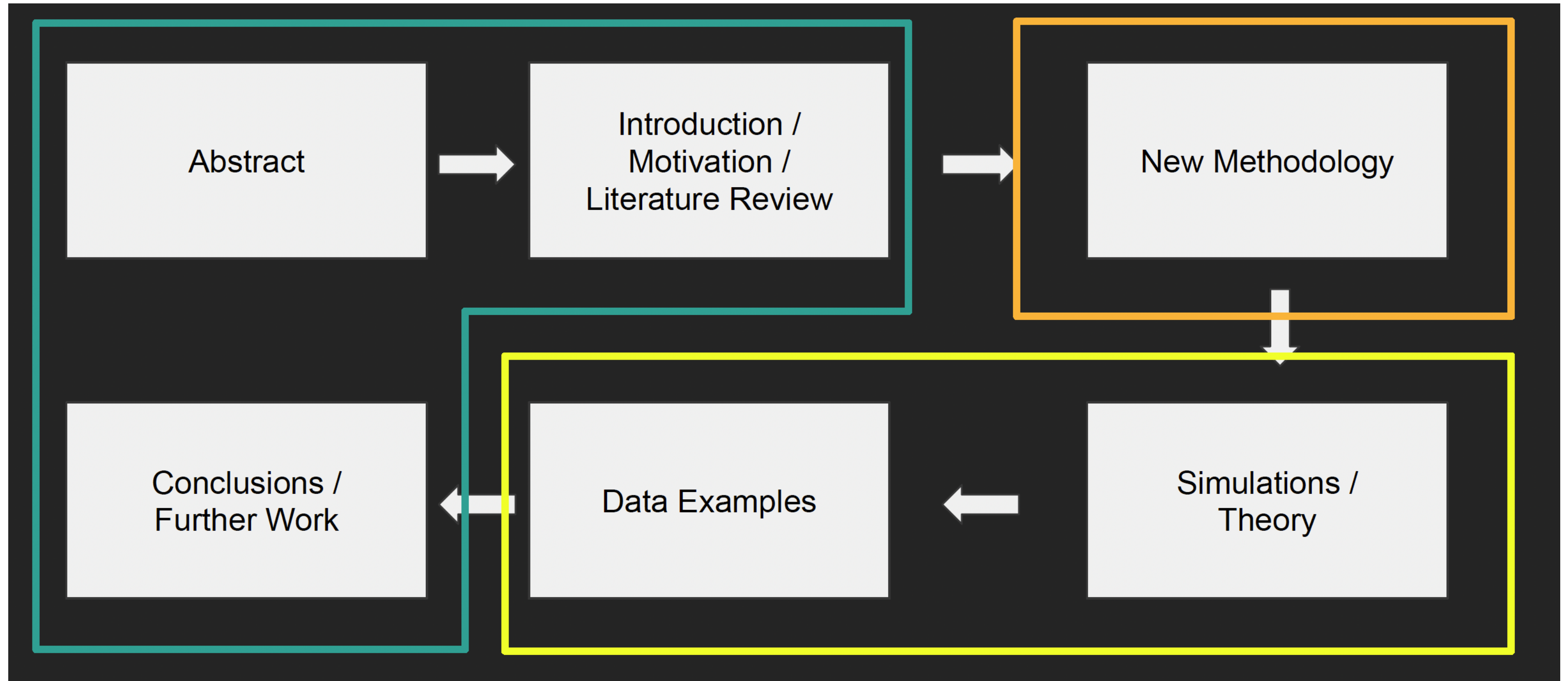


# Methods Papers

# Typical Strucutre of a Methods Paper



# Typical Strucutre of a Methods Paper



# Group 1 - End matter

- How do the authors use the end-matter to “sell” their new work?
- What other work has been done in this area?
- What limitations does the method have?
- How could this work be taken further?

# Group 2 - Methodology

- What new methods are introduced in this paper?
- Is it clear which aspects of this are novel?
- Were there any parts that you found particularly clear or confusing?

# Group 3 - Theory / Evidence

- What simulations or theoretical results do the authors present to justify their new method?
- What data examples do they use to demonstrate their new method?
- Do you think that these are effective, or can you think of better examples?