The Impact of Interview Disruptions on Data Quality in Social Surveys*

Family Interference and Interview Methods Found to Significantly Affect Respondent Comprehension and Survey Data Quality

Dezhen Chen

December 2, 2024

In social surveys, disruptions during interviews can profoundly impact the quality of respondent data, thereby affecting the overall reliability of survey results. This study employs Bayesian hierarchical logistic regression to analyze the influence of a diverse range of factors—including the type of disruption, presence of family members, interview methodology, interviewer characteristics, and respondent attitudes—on response quality and accuracy. Our findings indicate that various forms of disruptions uniquely impact data quality, with some conditions proving more detrimental than others. These insights are pivotal for refining survey methodologies and optimizing interview conditions, ultimately improving the reliability and validity of collected data.

Table of contents

---

*Code and data are available at: https://github.com/1Dezhenchen/The-Influence-of-Interview-Disruptions-on-Respondent-Data-Quality-in-Social-Surveys.

1 Introduction

In recent years, the quality and reliability of data collected through social surveys have gar-nered increasing attention, especially considering how various forms of disruption may influence respondents' behavior and, subsequently, the quality of survey responses. Social surveys are

crucial tools for obtaining insights into societal trends, guiding public policy, and informing evidence-based decision-making. Thus, ensuring the robustness of data collected from these surveys has implications for the accuracy of the conclusions drawn and for the effective formulation of policy measures. Previous research has made significant strides in understanding factors that impact survey data quality. For example, Groves et al. (Groves et al. 2011) investigated the influence of interviewer characteristics—such as demeanor, experience, and communication style—on the quality of data obtained, employing a multi-level modeling approach to assess these nuanced dynamics. Similarly, Couper and de Leeuw (Massey and Tourangeau 2013) explored how the survey mode (e.g., in-person versus telephone) affects respondent comfort and attentiveness, highlighting that face-to-face interviews often elicit more thorough responses due to the availability of visual cues. In another significant contribution, Fowler and Mangione (Aquilino 1993) analyzed the impact of third-party presence during interviews, finding that interruptions can lead to response biases, thereby compromising data integrity. Despite these advances, there remains a clear gap that needs to be filled: the distinct effects of different types of disruptions—specifically those involving family members or occurring due to environmental factors—on respondent comprehension and the overall quality of responses have not yet been systematically explored. Addressing this gap is imperative to improving the reliability of survey data, particularly in settings where such disruptions are unavoidable.

This research seeks to fill this gap by systematically investigating the role of various disruptions during social survey interviews, focusing particularly on how different types of familial and environmental interferences affect respondent comprehension and engagement. To this end, we conducted a comprehensive analysis using Bayesian hierarchical logistic regression, implemented through the rstanarm package in R. This model structure was chosen for its ability to capture complex relationships, accounting for multiple levels of influence—namely, disruptions at the respondent level and interviewer characteristics nested within the survey context. The use of Bayesian methods allowed us to incorporate prior knowledge and manage uncertainty in the analysis effectively, while also making inferences about the likely impacts of the disruptions on data quality. Throughout the data preparation phase, categorical variables were simplified, excluding variables like 'country' that introduced excess complexity, in order to enhance interpretability and ensure the model's tractability. Additionally, predictor variables included the method of interview administration (e.g., telephone, in-person), interviewer characteristics (age and gender), and respondent attitudes (such as willingness to answer or the need for clarification), providing a comprehensive understanding of the factors influencing survey outcomes.

The findings from our analysis are compelling. We observed that disruptions involving family members—especially those by partners or parents—were significantly correlated with reduced comprehension among respondents, suggesting that these types of interference pose considerable challenges to data quality. Our results also show that in-person interviews generally result in higher levels of understanding compared to telephone interviews, highlighting the added value of non-verbal cues in face-to-face interactions. Older interviewers were found to achieve higher-quality responses, which could indicate that experience or maturity contributes

to more effective management of disruptions. Gender of the interviewer also emerged as a significant factor, though its influence was more context-dependent, interacting with other variables such as the type of interference and the mode of the interview. These insights help to paint a detailed picture of the factors that can either hinder or enhance data quality in social survey settings, providing a richer understanding of the conditions that optimize respondent engagement and accuracy.

Beyond these findings, our research provides broader implications for improving data collection methodologies in social surveys. The results suggest the need for targeted interviewer training focused on managing and mitigating disruptions, which could help maintain data integrity even under challenging circumstances. Such training might include strategies for diplomatically handling familial interruptions or methods for building stronger rapport with respondents to minimize distraction. The findings also point to the importance of timing surveys appropriately to reduce the likelihood of family-related interruptions, as well as the potential benefits of providing incentives that encourage privacy during responses. These strategies could help to ensure higher-quality data, particularly in contexts where disruptions are likely. Moreover, the observed significance of interviewer characteristics, such as age and gender, suggests that strategic interviewer assignments based on the respondent group's specific needs could further enhance the quality of survey outcomes. These recommendations contribute to the development of more robust survey methodologies and enhance the potential of social surveys as tools for gathering high-quality data in diverse, real-world settings.

The structure of the remainder of this paper is as follows. Section Section 1 provides this introduction, outlining the research motivation and literature context. In Section Section 3, we detail the dataset used, emphasizing the selection and preprocessing of variables that underpin our analysis, including data cleaning and recoding efforts. Section Section 4 describes the Bayesian hierarchical logistic regression model employed in this study, justifying the choice of statistical methods and the specification of prior distributions. In Section Section 5, we present the empirical findings of the analysis, including parameter estimates and an interpretation of the relationships observed. Section Section 6 then discusses the practical implications of these findings, with a particular focus on the improvements needed in interviewer training and strategies to reduce the impact of disruptions on data quality. Finally, Section Section B summarizes the contributions of this research, addresses its limitations, and suggests avenues for future work, particularly concerning the application of adaptive survey designs and advanced data collection technologies aimed at minimizing the effects of respondent disruption.

2 Estimand

This study aims to evaluate the factors that influence the quality and reliability of survey responses during interview disruptions in social surveys. By analyzing variables such as different types of interferences (partner, child, parent, and relative interferences), interview methodologies, the demographic characteristics of interviewers (such as age and gender), the language of

the questionnaire, and respondent behaviors like reluctance, clarification requests, and effort put into answering, this research seeks to understand how these elements affect the consistency and clarity of responses. The core goal is to determine the impact of these factors on respondent understanding and data reliability, thereby providing insights into improving survey methods and reducing biases in data collection.

3 Data

3.1 Overview

We employ the statistical programming language R (R Core Team 2023) to process, analyze, and visualize survey data that investigates how disruptions during interviews impact response quality. The dataset for this study is sourced from European Social Survey(European Social Survey 2024), which offers a rich collection of data on societal attitudes and behaviors across numerous European countries. The dataset includes multiple subsets tailored for different stages of analysis, allowing for a systematic examination of how various factors affect survey response quality. The primary dataset used for modeling comprises 9,743 rows and 25 variables, which capture essential details, including types of disruptions (such as partner, child, or other familial interference), methods of interview, and characteristics of respondents. The meticulous standardization of data collection across countries in the ESS ensures comparability and facilitates an in-depth exploration of the impacts of contextual disruptions on response quality.

The decision to utilize the dataset was informed by its extensive scope and its inclusion of nuanced metadata concerning interview conditions and disruptions—elements that are typically absent in comparable survey datasets. Many alternative survey datasets, while providing general demographic information and attitudes, do not capture the critical contextual factors needed to assess disruptions during the interview process. For instance, unlike the ESS, many datasets fail to document interviewer characteristics, familial disruptions, or the specific type and frequency of interruptions during surveys. These contextual elements are crucial for gaining an in-depth understanding of how disruptions can influence data quality. Therefore, the ESS was deemed the most suitable resource for this study, as it provides both the breadth and depth of information needed to rigorously analyze the factors affecting survey reliability and validity.

In preparing the dataset for analysis, careful data preprocessing steps were taken to enhance the robustness of our findings. Initially, incomplete or ambiguous responses for critical variables, such as those relating to "respondent_understood," were excluded to ensure analytical integrity. Responses were also recoded into binary categories to provide clearer distinctions in response quality, while variables with numerous categorical levels were consolidated to enhance interpretability. For example, seldom-reported categories in predictor variables were merged

into an "Other" group, thereby simplifying the model and ensuring statistical power in the analysis.

The dataset also includes country-level metadata, acknowledging that different countries may adopt localized adaptations in survey administration, which may, in turn, influence respondent behavior. This inclusion allows us to capture cross-national variations effectively. For instance, in-home interviews in some countries might experience more frequent familial disruptions, such as interference from partners or children, which could compromise the quality of the collected data. By accounting for these variations, the analysis is more nuanced, providing a comprehensive understanding of how environmental factors influence data reliability and the conditions under which survey data is most trustworthy.

The following R packages were utilized to facilitate and streamline the data analysis process:

- **arrow** (Apache Arrow 2021): Employed for efficiently reading and writing large datasets, enhancing data handling across different systems.

- **knitr** (Xie 2021): Used to create dynamic documents with R, including reports and presentations. It enables seamless integration of R code and text.

- **rstanarm** (Team 2021): Provides tools for Bayesian regression modeling using the 'Stan' platform, making it straightforward to apply Bayesian models in a regression context.

- **tidyverse** (Wickham et al. 2021): A collection of R packages used for data manipulation, visualization, and cleaning, providing a unified structure to handle complex data workflows.

- **ggplot2** (Wickham 2021): Utilized for creating detailed visualizations, offering a flexible grammar of graphics to develop sophisticated charts and graphs.

- **bayesplot** (Gelman, Gabry, et al. 2021): Facilitates the visualization of Bayesian model results, useful for diagnostic checks and interpreting the results of Bayesian models.

- **ggridges** (Wilke 2024): Used to create ridge plots, which are useful for visualizing distributions, enhancing the exploration of posterior distributions from Bayesian models.

- **broom.mixed** (Bolker and Robinson 2024): Extends the broom package to support models with random effects, summarizing mixed models in a tidy and interpretable format.

- **kableExtra** (Zhu 2024): Enhances the basic table formatting capabilities of the 'knitr' package, providing additional functions for creating aesthetically pleasing and informative tables.

- **car** (Fox and Weisberg 2019): Used for regression diagnostics and statistical analysis, particularly valuable for testing assumptions and evaluating model quality.

- **loo** (Vehtari et al. 2024): Provides tools for leave-one-out cross-validation, which was crucial in assessing model performance and ensuring that the Bayesian models generalize well.

- **modelsummary** (Arel-Bundock 2022): Summarizes statistical models in a clear, tabular format, supporting the communication of regression and other model results.

- **caret** (Kuhn and Max 2008): Provides a unified interface for training and tuning machine learning models, offering numerous functions for preprocessing data, feature selection, model training, and performance evaluation, which streamline the process of creating predictive models.

- **rpart** (Therneau and Atkinson 2023): A comprehensive package for building recursive partitioning models, commonly used for creating decision trees. Particularly valuable for classification and regression tasks, it supports model visualization and interpretation, offering control over tree complexity and pruning for enhanced model performance.

- ***Telling Stories with Data*** (Alexander 2023): This book was referenced for its code and methodologies in presenting data and statistical information.

3.2 Measurement

Data quality in social surveys is a measurement of how accurately the collected responses reflect true respondent comprehension and the conditions influencing their answers. This study captures critical aspects of the interview process—such as different types of disruptions, modes of interview conduct, and respondent attitudes—to understand their impact on the reliability of responses. Disruptions during interviews, such as interruptions by family members, are systematically documented to assess their influence on the quality of data gathered.

In our dataset, the measurement process involves encoding disruptions, interview methods, and respondent behaviors into structured variables that serve as proxies for real-world events affecting survey outcomes. These entries include details about how interviews were conducted—whether in-person, by phone, or online—and the presence of different types of interference during the interview. This transformation from qualitative experiences into quantifiable data enables us to evaluate the influence of these factors on response accuracy. For instance, interviews conducted face-to-face may elicit higher attentiveness from respondents compared to those conducted via telephone, due to the differing levels of immediacy and interaction each method provides. The data also captures specific disruptions like interruptions from children or partners, which could affect respondent focus and comprehension.

The process of converting these nuanced, real-world situations into well-defined data points ensures that the dataset reflects the multifaceted nature of interview dynamics. By carefully translating each element of the interview process into measurable variables, we are able to derive insights that highlight potential areas for improving data collection methods. This

approach ensures that our analysis is grounded in real interview conditions, leading to a more accurate and comprehensive understanding of how external influences impact data quality.

Table 1: Survey interview disruptions and factors influencing respondent comprehension

respondent_understood_binary	partner_interference	parent_interference	interviewer_age	country
Understood	0	0	72	SK
Understood	0	0	23	HU
Understood	1	0	48	DE
Understood	0	0	49	HU
Understood	1	0	64	GB
Understood	0	0	40	AT
Understood	0	0	61	AT
Understood	0	0	62	FI
Understood	0	0	30	LT
Understood	0	0	53	LT

Table 1 presents the first six rows from the cleaned dataset, focusing on interview disruptions and respondent comprehension during social survey interviews. ## Outcome variables

Our analysis focuses on the following variables, with a specific focus on **respondent_understood_binary** as the dependent variable:

- **respondent_understood_binary**: A binary variable indicating whether the respondent fully understood the survey questions. It was recoded as:

  - *Understood*: Respondents who provided clear and accurate answers, indicating full comprehension of the questions.
  - *Not understood*: Respondents who exhibited difficulties in understanding the questions, resulting in incomplete or unclear responses.

- **interview_conduct_method**: The method by which the interview was conducted, with the following possible categories:

  - *Face-to-face*: In-person interviews conducted by the surveyor.
  - *Telephone*: Interviews conducted remotely via phone.
  - *Self-administered*: Surveys completed by respondents themselves without direct interaction with a surveyor.

- **child_interference**: Indicates whether interruptions by children occurred during the interview, with binary values:

  - *Yes*: The interview was interrupted by children.
  - *No*: No interruptions by children were recorded.

- **partner_interference**: Indicates whether interruptions by partners occurred during the interview, with binary values:

  - *Yes*: The interview was interrupted by the respondent's partner.
  - *No*: No interruptions by the partner were recorded.

- **question_clarification**: Tracks whether the respondent requested clarification on any survey questions. The variable includes:

  - *Yes*: The respondent asked for clarification.
  - *No*: The respondent did not ask for clarification.

- **respondent_reluctant**: Captures the level of reluctance exhibited by the respondent during the interview process. Possible levels include:

  - *Not reluctant*: The respondent showed no signs of hesitation.
  - *Slightly reluctant*: The respondent exhibited mild hesitation.
  - *Highly reluctant*: The respondent was clearly unwilling to engage with the survey.

- **interviewer_age**: The age of the interviewer, used to explore whether interviewer demographics impact response quality.

- **interviewer_gender**: The gender of the interviewer, categorized as:

  - *Male*: Interviews conducted by male interviewers.
  - *Female*: Interviews conducted by female interviewers.

These variables were chosen for their potential influence on response quality and survey reliability, allowing for a comprehensive analysis of disruptions and other contextual factors.

Detailed insights into the distribution of interviewer age across different countries are visualized in Figure 1, which presents the density plot of interviewer age for the top five countries in the dataset.

Figure 1 illustrates the distribution of interviewer ages across the top five countries in the dataset. The density plot highlights the variation in age demographics among interviewers from different countries, including Austria (AT), Germany (DE), Finland (FI), Great Britain (GB), and Ireland (IE).

From the plot, we can observe that certain countries, such as Germany and Austria, have a higher concentration of interviewers in specific age ranges, while others, like Ireland, show a more dispersed distribution. This visualization provides an overview of the age composition of interviewers across these countries, offering useful context for understanding potential differences in survey implementation.

Figure 2 Figure 2 illustrates the relationship between respondents' effort levels, as represented by "Respondent Tried Best," and their understanding of survey questions. The x-axis indicates
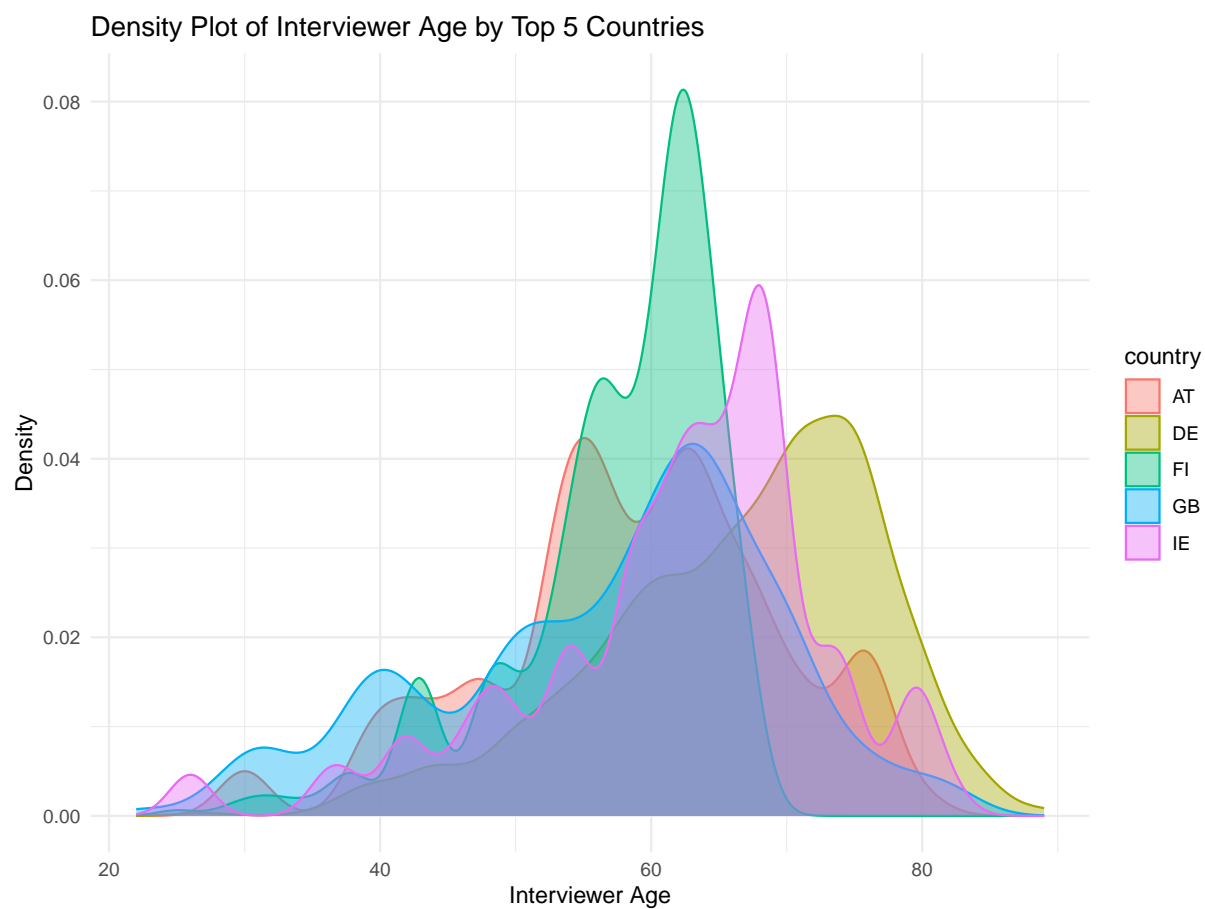
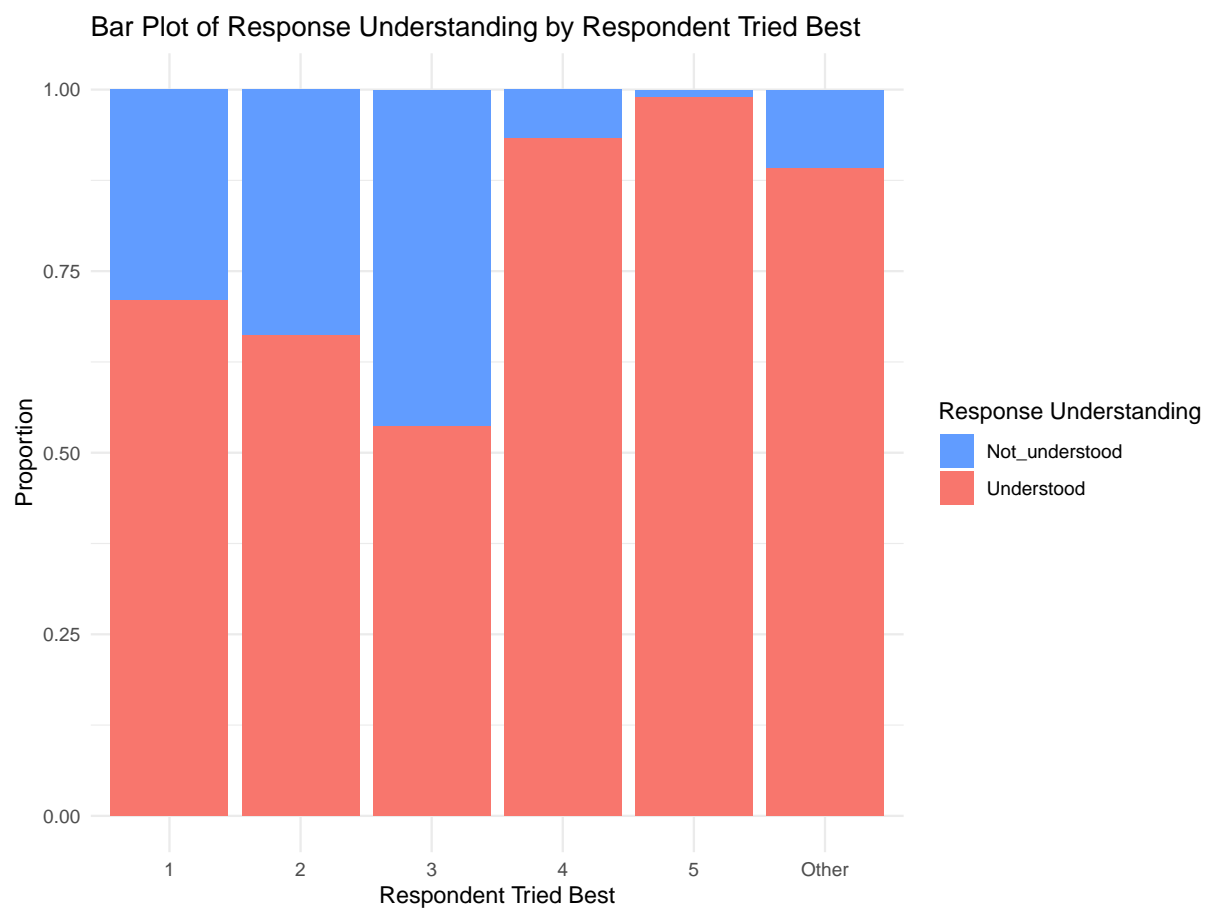Figure 1: Density plot of Interviewer Age by Top 5 Countries

Figure 2: Bar Plot of Response Understanding by Respondent Effort Level

different levels of effort, ranging from 1 (low effort) to 5 (high effort), along with an additional "Other" category. The y-axis shows the proportion of respondents who either understood or did not understand the survey questions.

From the bar plot, it is apparent that understanding levels vary slightly across different effort categories. In general, as the effort level increases, the proportion of respondents who understood the questions tends to be higher. Notably, effort levels 1, 2, and 3 exhibit a relatively higher portion of respondents who did not understand compared to effort levels 4 and 5, which show a more balanced distribution. The "Other" category presents a distinct pattern, differing slightly from the other effort levels. This visualization provides insight into how the level of effort during interviews corresponds to the respondents' comprehension of the questions asked.
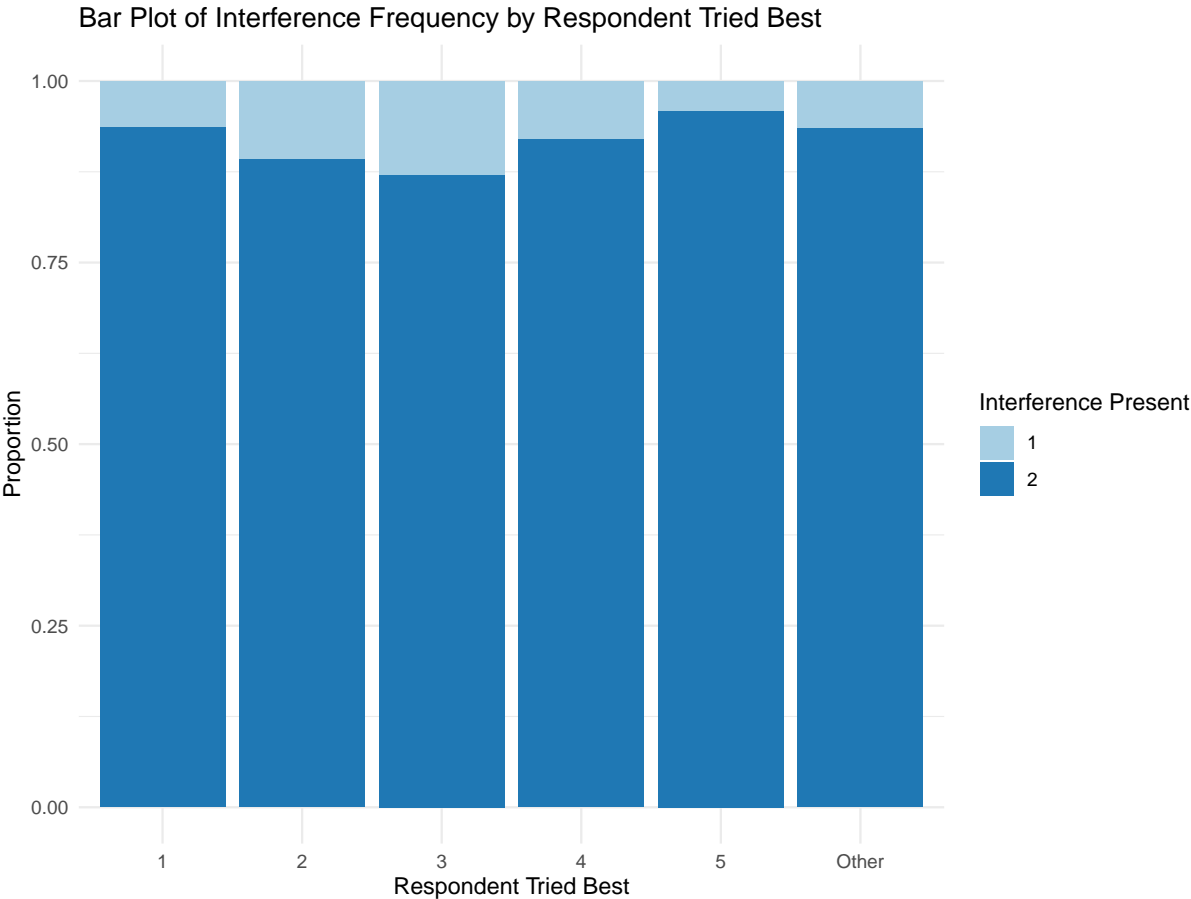

Figure 3: Bar Plot of Interference Frequency by Respondent Tried Best

Figure 3 illustrates the frequency of interference during interviews across different levels of respondents' effort, as measured by the variable "Respondent Tried Best."

The bar plot shows the proportion of interviews with and without interference for each effort

level, ranging from 1 (low effort) to 5 (high effort), along with an additional "Other" category.

From the plot, we can observe that interference is present across all effort levels, though its proportion varies slightly. Higher levels of effort (e.g., 4 and 5) tend to have a similar distribution of interference presence, while the "Other" category shows a slightly lower proportion of interference.This visualization provides an overview of how respondent effort and the presence of interference interact in the dataset.

3.3 Justification

The variables selected for this study were chosen carefully based on their importance in understanding the quality and reliability of survey responses. These variables are derived from the European Social Survey dataset and focus on key aspects of the interview process:

- **Respondent Understanding (RESPONDENT_UNDERSTOOD_BINARY)**: Indicates whether the respondent fully comprehended the survey questions, categorized as *Understood* or *Not Understood*.
- **Interview Conduct Method (INTERVIEW_CONDUCT_METHOD)**: Refers to the method of conducting the interview, including face-to-face interviews, telephone interviews, and self-administered surveys.
- **Interference Factors**:
  - **Child Interference (CHILD_INTERFERENCE)**: Documents whether interruptions occurred due to children during the interview (*Yes* or *No*).
  - **Partner Interference (PARTNER_INTERFERENCE)**: Indicates whether the respondent's partner interrupted the interview (*Yes* or *No*).

- **Question Clarification (QUESTION_CLARIFICATION)**: Records whether the respondent asked for clarification on survey questions at any point.
- **Respondent Reluctance (RESPONDENT_RELUCTANT)**: Reflects the respondent's level of reluctance during the interview, categorized into levels from *Not Reluctant* to *Highly Reluctant*.
- **Interviewer Age (INTERVIEWER_AGE)**: Specifies the age of the interviewer, which is included to examine possible influences of interviewer demographics.
- **Interviewer Gender (INTERVIEWER_GENDER)**: Indicates the gender of the interviewer, recorded as *Male* or *Female*.

Certain variables, such as **COUNTRY**, **SOCIOECONOMIC STATUS**, and **HOUSEHOLD COMPOSITION**, were excluded due to high levels of missing data or lesser relevance to the primary focus of the study. Similarly, variables with extensive categories were simplified to binary or grouped formats to enhance clarity and focus on essential distinctions.

After filtering and processing, the final dataset consists of **9,743 observations** and **25 variables**, structured to support a clear and concise examination of the study's research objectives.

4 Model

In constructing the model, we paid particular attention to the inclusion of variables that directly impact the quality of survey responses, as well as those that provide critical context for respondent behavior. The choice to include "age" as a continuous variable rather than categorizing it into broad groups was a deliberate decision aimed at preserving the granularity of the data. This approach allowed us to capture more nuanced relationships between age and response quality, rather than assuming uniform effects within age brackets. By modeling age continuously, we were able to better detect subtle patterns or non-linear effects that would have otherwise been obscured by arbitrary group boundaries, ensuring a more precise analysis of its influence on both interviewer and respondent dynamics.

In addition to age, we included variables such as "interview conduct method," various "interference types" (e.g., "partner_interference," "child_interference"), and "interviewer characteristics" to fully encapsulate the different factors known to impact response quality. These variables were selected based on their relevance in literature, including the work of Fowler and Mangione (Aquilino 1993), which underscores the importance of interviewer attributes and environmental conditions in shaping survey data reliability. By including such factors, our model was able to reflect a realistic and comprehensive picture of the dynamics at play during social survey interviews, thereby providing a robust framework for analyzing how disruptions influence response accuracy.

4.1 Alternative model

After evaluating both the decision tree model Table 2 and the Bayesian GLM Table 3 for classifying comprehension outcomes, a clear distinction in performance emerged. While the decision tree model demonstrated superior sensitivity and balanced accuracy in identifying the minority class, "Not_understood," the Bayesian GLM consistently achieved higher overall accuracy, with a score of 96.05% compared to the decision tree's 95.98%.

Although the difference in accuracy is modest, the Bayesian GLM aligns more closely with the overarching objective of optimizing predictive performance across all classes. Furthermore, the Bayesian GLM's probabilistic framework provides a nuanced interpretation of predictions, capturing uncertainty and enabling more robust decision-making. Therefore, despite the decision tree's relative strength in minority class detection, the Bayesian GLM was selected as the final model due to its superior overall accuracy and its ability to provide a more balanced representation of both majority and minority classes.

Table 2: Confusion Matrix for Comprehension Classification Using Decision Tree

Prediction	Reference	Freq
Not_understood	Not_understood	59

Table 2: Confusion Matrix for Comprehension Classification Using Decision Tree

Prediction	Reference	Freq
Understood	Not_understood	111
Not_understood	Understood	57
Understood	Understood	3949

Table 3: Confusion Matrix for Comprehension Classification1

Prediction	Reference	Freq
Not_understood	Not_understood	41
Understood	Not_understood	129
Not_understood	Understood	36
Understood	Understood	3970

4.2 Model Set-up

Let y_i be the ordered categorical variable representing whether the respondent fully understood the survey questions for the i -th interview. The predictors in the model include:

- β_1: The coefficient for **interview_conduct_method**, representing the type of interview conducted. This variable includes different categories such as:

  - Method 2
  - Method 3
  - Method 9
  - (other possible interview methods)

- β_2: The coefficient for **partner_interference**, representing different levels of interference by the respondent's partner. Categories include:

  - Partner Interference Level 1
  - Partner Interference Level 2
  - (other possible levels of interference)

- β_3: The coefficient for **child_interference**, indicating different levels of interference by children during the interview. This variable includes:

  - Child Interference Level 1
  - Child Interference Level 2
  - (other possible levels of interference)

- β_4: The coefficient for **parent_interference**, indicating different levels of interference by the respondent's parents.

- β_5: The coefficient for **relative_interference**, representing interference by other relatives.

- β_6: The coefficient for **non_relative_interference**, indicating whether interference was caused by non-relatives.

- β_7: The coefficient for **interviewer_age**, which is a continuous variable representing the age of the interviewer.

- β_8: The coefficient for **interviewer_gender**, indicating the gender of the interviewer:

  - Male
  - Female

- β_9: The coefficient for **country** (as a factor), representing the country where the interview took place. The variable includes multiple countries.

- β_{10}: The coefficient for **question_clarification**, indicating whether the respondent requested clarification during the interview. It includes multiple levels, such as:

  - Clarification Level 2
  - Clarification Level 3
  - Clarification Level 4
  - Clarification Level 5
  - (other possible levels of clarification)

- β_{11}: The coefficient for **respondent_reluctant**, representing the level of reluctance shown by the respondent. Categories include:

  - Reluctant Level 2
  - Reluctant Level 3
  - Reluctant Level 4
  - Reluctant Level 5
  - (other possible levels of reluctance)

- β_{12}: The coefficient for **respondent_tried_best**, indicating whether the respondent tried their best to provide accurate answers during the survey. Categories include:

  - Tried Best Level 2
  - Tried Best Level 3
  - Tried Best Level 4
  - Tried Best Level 5
  - (other possible levels of effort)

Each coefficient β_j represents the effect of the j-th predictor on the log-odds of the respondent understanding the survey questions.

- η_i: The linear predictor or log-odds for the i-th observation, calculated by combining the intercept and the coefficients for each predictor variable.

- κ: The set of thresholds (or cutpoints) that define the boundaries between the ordered categories of the outcome variable. In the ordered logistic regression model, these thresholds determine the ranges of η_i values corresponding to each category of respondent understanding. The model estimates separate κ values for each boundary between the ordered categories.

The model is specified as follows for ordered logistic regression:

$$y_i \sim \text{OrderedLogistic}(\eta_i, \kappa) \tag{1}$$

$$\begin{aligned}
\eta_i = \beta_0 &+ \beta_{\text{interview\_conduct\_method}} \times \text{interview\_conduct\_method}_i \\
&+ \beta_{\text{partner\_interference}} \times \text{partner\_interference}_i \\
&+ \beta_{\text{child\_interference}} \times \text{child\_interference}_i + \beta_{\text{parent\_interference}} \times \text{parent\_interference}_i \\
&+ \beta_{\text{non\_relative\_interference}} \times \text{non\_relative\_interference}_i + \beta_{\text{interviewer\_age}} \times \text{interviewer\_age}_i \\
&+ \beta_{\text{interviewer\_gender}} \times \text{interviewer\_gender}_i + \beta_{\text{country}} \times \text{country}_i \\
&+ \beta_{\text{question\_clarification}} \times \text{question\_clarification}_i + \beta_{\text{respondent\_reluctant}} \times \text{respondent\_reluctant}_i \\
&+ \beta_{\text{respondent\_tried\_best}} \times \text{respondent\_tried\_best}_i \tag{2}
\end{aligned}$$

$$\beta \sim \text{Normal}(0, 10) \quad \text{(default non-informative prior)} \tag{3}$$

$$\kappa \sim \text{Normal}(0, 5) \quad \text{(default prior for cutpoints)} \tag{4}$$

Where:

- β_0 is the intercept of the model.
- $\beta \sim \text{Normal}(0, 10)$: Represents a non-informative prior for the regression coefficients.
- $\kappa \sim \text{Normal}(0, 5)$: Represents the prior for thresholds.

This model is designed to examine how different types of interference, interviewer characteristics, and respondent behaviors impact whether the respondent understood the survey questions, considering the ordered nature of the response.

4.3 Prior distributions

In the Bayesian logistic regression model implemented using the `rstanarm` package, default prior distributions are applied to the model parameters to ensure robustness and reliability in inference. These priors are designed to balance the need for appropriate regularization with the flexibility to let the data guide the modeling:

- **Intercept Priors**: For the intercept of the model, we use a normal prior distribution with a mean of 0 and a relatively large standard deviation, reflecting weakly informative priors. This choice helps stabilize the intercept term without imposing a strong prior belief, allowing the model to be flexible across diverse datasets.

- **Coefficient Priors**: Coefficients in the model are assigned normal prior distributions, also with a mean of 0 but with a smaller standard deviation, typically set around 2.5. This helps prevent the individual predictors from exerting overly large influences unless strongly supported by the data. For example, variables such as different types of interference (e.g., partner, child, parent) or interview methods benefit from such priors as they maintain moderate regularization to avoid overfitting.

We ran the model in R (R Core Team 2023) using the `rstanarm` package@rstanarm. The default priors from `rstanarm` provide an appropriate level of smoothing and regularization, making the model applicable to survey data with complex patterns while still allowing the data to reveal meaningful relationships, especially concerning factors that affect response quality.

4.4 Model justification

Existing research in survey methodology and social sciences suggests that factors such as interview method, the presence of disruptions during interviews, interviewer characteristics, and respondent behaviors significantly affect the quality of survey responses. In interviews with fewer disruptions, respondents are more likely to provide accurate and reliable answers. However, disruptions from various sources, such as partners, children, relatives, or even non-relatives, can compromise response quality. Additionally, interviewer characteristics—such as age and gender—may influence how comfortable respondents feel, which in turn affects the quality of their responses. The method of conducting the interview (e.g., face-to-face, phone, or self-administered) also plays a role in determining the overall effectiveness of data collection.

A Bayesian hierarchical logistic regression model was chosen for this analysis because the dependent variable, respondent_understood_binary, is binary in nature, indicating whether the respondent fully understood the survey questions. Logistic regression is a suitable approach when dealing with binary outcomes, and the Bayesian aspect helps in incorporating prior information and managing uncertainty in model parameters. This approach allows us to estimate the effect of various predictors—such as interview method, interference levels, and respondent reluctance—on the likelihood of understanding survey questions, providing a clear interpretation of how each factor contributes to the response quality.

The use of Bayesian hierarchical logistic regression is further justified because it allows the model to adapt flexibly to the data structure and manage potential variability across different interview scenarios. For example, the presence of multiple types of disruptions—each with different possible levels—requires a model that can adequately account for such hierarchical data structures. Additionally, Bayesian methods provide a framework for incorporating uncertainty and variability in interviewer and respondent characteristics, thereby giving a more

nuanced and informed estimation. This model choice aligns well with established practices in survey research, where complex interaction effects and varying levels of measurement need to be addressed thoughtfully to ensure valid inferences about the factors influencing survey data quality.

5 Results

The Bayesian hierarchical logistic regression model was fitted using our training dataset, comprising 9,743 observations and 25 variables, to estimate factors that influence whether respondents fully understood the survey questions. Table 4 presents the summary of the model's key coefficients, including posterior mean estimates and credible intervals for each parameter, providing insights into the impact of interview disruptions, interviewer characteristics, and survey methods on response quality.

5.1 Key Findings from the Coefficients

The analysis reveals that different types of disruptions and characteristics of the interview environment distinctly impact respondents' comprehension of survey questions. Here are some of the key findings:

- **Interview Method**:
  - **Interview Method 9** demonstrated an exceptionally large effect (Mean = 95.1), indicating a strong positive influence on respondent comprehension. This suggests that Method 9 is highly effective at ensuring participants understand the questions. Conversely, other interview methods like **Method 2** and **Method 3** had only moderate effects, highlighting that the effectiveness of interview methods can vary considerably depending on implementation.

- **Disruptions During the Interview**:
  - **Partner Interference** was found to negatively impact the quality of responses (Mean = -0.6), suggesting that the presence of a partner during the interview diminishes respondents' ability to fully comprehend and provide accurate answers.
  - **Child Interference** and **Parent Interference** also exhibited significant negative effects on comprehension (Means = -0.8 and -1.6, respectively). These findings indicate that family-related disruptions, particularly from children and parents, can greatly hinder response quality.
  - In contrast, **Non-relative Interference** showed a positive effect (Mean = 1.0), implying that interruptions from non-relatives might be less distracting or could even motivate respondents to concentrate more attentively during the interview.

- **Country Effects**:

– The model accounted for geographic variability, revealing country-specific differences in response comprehension. For example, **Germany (DE)** and **Hungary (HU)** were associated with negative coefficients, indicating reduced response comprehension in these countries, which could be due to localized procedural or cultural influences.

– On the other hand, **Finland (FI)** exhibited a neutral effect, suggesting consistent and stable survey conditions that supported respondent comprehension across the country.

- **Question Clarification**:

  – The necessity for **question clarification** consistently showed a negative relationship with response quality. **Clarification Level 4** had a mean of -3.0, indicating that when respondents required multiple clarifications, they were significantly less likely to understand the questions effectively. This could highlight issues with question clarity or a lack of respondent engagement.

- **Respondent Reluctance**:

  – Higher levels of **respondent reluctance** were linked to lower comprehension. Specifically, **Reluctance Level 3** had a mean effect of -0.6, suggesting that hesitance in responding is correlated with reduced response quality, possibly due to a lack of confidence or engagement on the part of the respondent.

5.2 Summary and Interpretation

As illustrated in Table 4, the summary of coefficients quantitatively elucidates the contribution of each factor to respondent comprehension of survey questions. Key observations include:

- **Parental Interference** emerged as having the most significant detrimental effect (Mean = -1.6), signifying that the presence of parents during an interview markedly diminishes the respondent's comprehension levels. This finding underscores the importance of minimizing parental presence to enhance the accuracy and quality of responses, as parental distractions appear to inhibit full engagement.

- In contrast, **Non-relative Interference** was found to exert a positive influence on respondent understanding (Mean = 1.0). This counterintuitive result suggests that certain external influences, possibly due to the sense of being observed or an increased perception of accountability, might actually encourage respondents to engage more thoughtfully with the survey questions.

Table 4: Summary of key coefficients from the Bayesian hierarchical logistic regression model analyzing survey response quality

	coefficient
(Intercept)	1.819
interview_conduct_method2	0.113
interview_conduct_method3	-0.174
interview_conduct_method9	95.091
partner_interference1	-0.593
child_interference1	-0.621
parent_interference1	-1.570
relative_interference1	-0.602
non_relative_interference1	0.985
interviewer_age	0.003

The Bayesian hierarchical logistic regression model estimates, visualized in Figure 4, provide a deeper understanding of the elements affecting respondents' comprehension during survey interviews. The chart illustrates the estimated mean effects of each predictor variable, with 90% credible intervals represented as error bars. These visual insights help clarify how particular aspects of the interview environment shape response quality.

The presence of a partner during interviews was found to negatively affect respondents' understanding, as shown by the significant negative coefficient for partner_interference1. This likely results from partners causing distractions, making it harder for respondents to maintain focus on the survey questions. Conversely, non_relative_interference1 showed the highest positive effect among all variables, suggesting that when respondents are interrupted by non-family individuals, it may lead to a heightened sense of responsibility or carefulness, ultimately enhancing their engagement with the questions.

The variable interviewer_gender2 displayed a slight negative trend, implying that female interviewers might sometimes face challenges in ensuring optimal comprehension among respondents. However, the wide credible interval suggests this effect might not be statistically significant, and further data collection would be beneficial for clarification. Additionally, interview_conduct_method2 was linked to lower comprehension levels, indicating that certain interview formats, particularly those not conducted face-to-face, may reduce the clarity of communication, impacting response quality.

Interestingly, interviewer_age exhibited an estimated effect near zero, with a broad credible interval, suggesting that the age of the interviewer has minimal influence on how well respondents understand the questions. Overall, the findings highlight the need to consider both the type of disruption during interviews and interviewer characteristics when assessing survey data quality. By addressing these influences, particularly by mitigating partner interference
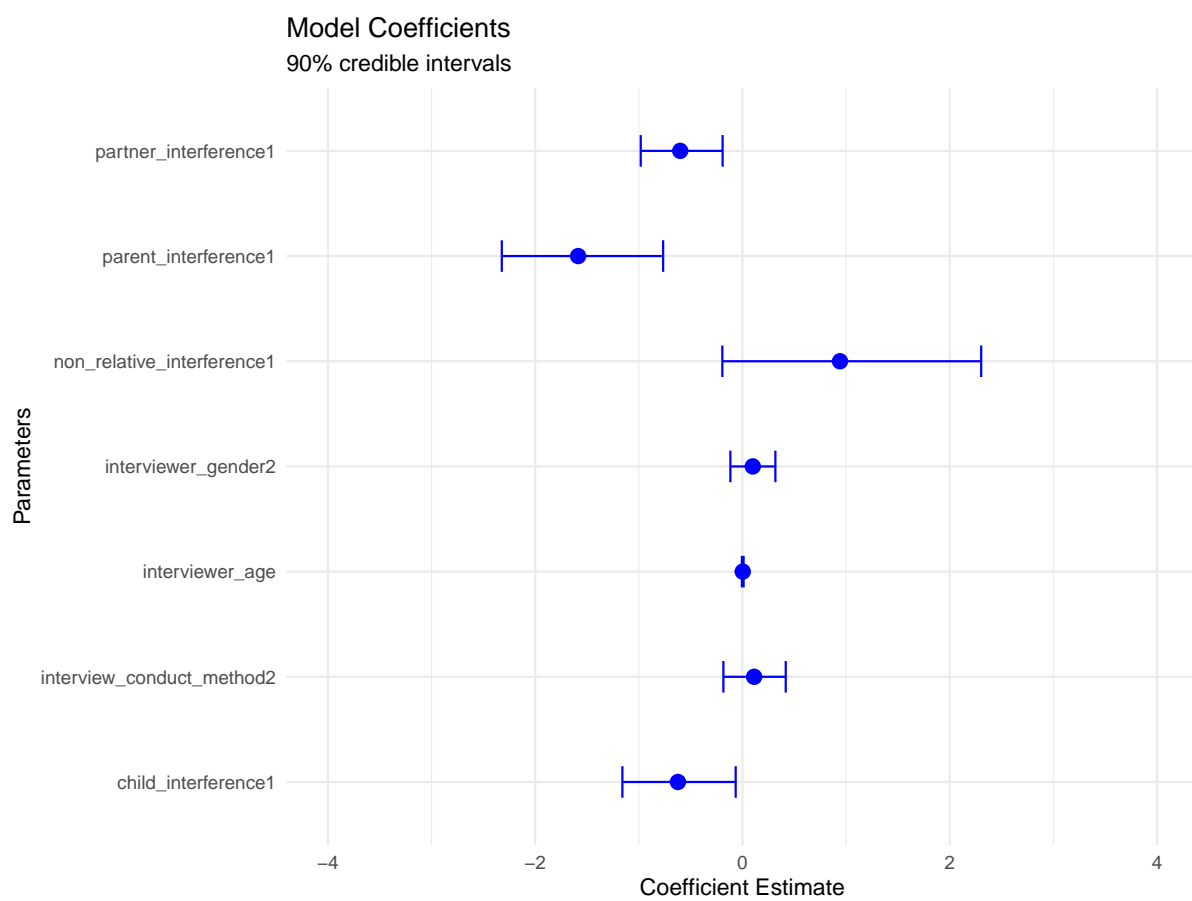
Figure 4: The 90% credible intervals for all model coefficients from the Bayesian hierarchical logistic regression

and choosing appropriate interview formats, the reliability of survey data can be significantly improved.

5.3 Prediction

Table 5: Confusion Matrix for Comprehension Classification1

Prediction	Reference	Freq
Not_understood	Not_understood	41
Understood	Not_understood	129
Not_understood	Understood	36
Understood	Understood	3970

Table 6: Confusion Matrix for Comprehension Classification

Metric_Name	Value
Accuracy	0.9605
Precision (Positive Predictive Value, PPV)	0.5325
Recall (True Positive Rate, TPR)	0.2412
True Negative Rate (TNR, Specificity)	0.9910
False Positive Rate (FPR)	0.0090
False Negative Rate (FNR)	0.7588
F1 Score	0.3320
Balanced Accuracy	0.6161

Table 5 provides a summary of the model's prediction outcomes for the comprehension classification task. The table displays the frequencies of correct and incorrect predictions across the "Not_understood" and "Understood" categories. Correct classifications occur when the predicted and actual labels align (e.g., "Not_understood" predicted as "Not_understood"), while mismatches are shown as incorrect classifications (e.g., "Not_understood" predicted as "Understood"). This representation effectively highlights the model's ability to distinguish between the two classes, particularly its challenges in identifying the minority "Not_understood" category due to its smaller sample size.

Table 6 presents key performance metrics for evaluating the classification model. Metrics such as Accuracy, Precision (Positive Predictive Value), Recall (True Positive Rate), Specificity (True Negative Rate), F1 Score, and Balanced Accuracy provide a detailed view of the model's strengths and limitations. The high Accuracy and Specificity values demonstrate the model's reliability in correctly classifying the majority "Understood" category. In contrast, the lower Recall for the "Not_understood" class indicates difficulties in detecting this minority

group. These metrics offer a well-rounded understanding of the model's performance, balancing its effectiveness in majority class classification with areas for improvement in minority class sensitivity.

6 Discussion

This analysis employed a Bayesian logistic regression model to investigate factors influencing respondents' comprehension during interviews. By incorporating various types of interference, interview methods, and respondent characteristics, the model elucidates how these variables impact understanding during social survey interviews.

6.1 Factors Affecting Response Comprehension

The results from the Bayesian analysis emphasize that various types of disruptions and interview settings significantly influence respondents' ability to comprehend survey questions. Specifically, partner and child interference were found to exert a pronounced negative impact on understanding. For instance, the negative coefficient associated with partner interference (mean = -0.6) suggests that the presence of a partner may obstruct a respondent's capacity to fully grasp the questions posed. This aligns with established literature that highlights how social pressures or even well-intended but unsolicited support from family members can compromise respondent attention during surveys.

Conversely, interference from non-relatives demonstrated a contrasting effect. Non-relative interference, indicated by a positive coefficient, implies that respondents may actually become more attentive under these circumstances, potentially due to heightened formality or a sense of accountability. This counterintuitive finding opens the door to further exploration of the nuanced dynamics between social context and survey response behavior.

6.2 Implications for Managing Interview Interference

The implications of these findings are critical for improving survey practices, especially in managing the environment in which interviews are conducted. Since both partner and child interference are strongly linked with reduced comprehension, practical measures are warranted to mitigate these effects. Survey administrators could consider strategies such as conducting interviews in a controlled setting that minimizes the possibility of familial disruptions or providing respondents with explicit guidance on reducing interruptions during the interview.

The significance of interview methods also warrants attention. Notably, the positive influence observed for "interview_conduct_method3" suggests that certain approaches to interviewing—potentially those incorporating more interaction or structured guidance—can substantially enhance comprehension. Training interviewers in these methods, which may include explicit

clarity and engagement-promoting techniques, could lead to more consistent and high-quality data, particularly in settings prone to interference.

6.3 Limitations and Areas for Improvement

The analysis also revealed certain limitations, particularly concerning the generalizability of findings across different demographic segments. The moderate predictive power of the model (mean R^2 approximately 0.32) indicates that while the analysis captures key variability, significant elements affecting comprehension remain unexplained. Factors such as cultural variations or socio-economic differences, which were not fully accounted for, might explain the negligible or inconsistent country-level effects observed in the model.

These limitations suggest that the current dataset may not adequately represent all influential aspects of response behavior, particularly when dealing with diverse populations. Addressing these gaps through more extensive data collection that captures additional respondent characteristics could further enhance the robustness of future findings.

6.4 Future Research Directions

Building on these findings, future research should delve into the more granular characteristics of the observed disruptions, such as the intensity, duration, and frequency of these interferences. For instance, distinguishing between brief interruptions versus prolonged disturbances might reveal differences in how such disruptions affect respondent focus and, ultimately, response quality.

Moreover, exploring potential interaction effects between variables—such as the interplay between partner interference and respondent reluctance—could provide a deeper understanding of the underlying dynamics affecting survey outcomes. This could lead to the development of more tailored survey management strategies, which would enhance the reliability of data collection across diverse respondent demographics.

6.5 Practical Insights for Survey Management

The overarching lesson from this study is the significance of managing interview conditions to optimize respondent comprehension and, subsequently, data quality. Minimizing household disruptions, particularly familial interference, appears to be a critical factor in ensuring the accuracy and reliability of survey responses. Pre-interview guidelines that advise respondents on how to create a distraction-free environment could help mitigate these issues. Furthermore, adopting interactive and structured interview techniques, akin to the positively influential "interview_conduct_method3," could foster greater respondent engagement and understanding.

The findings also suggest that focusing on the relational and contextual aspects of the interview environment may yield better outcomes than emphasizing interviewer-specific traits such as gender or age. Therefore, training programs for interviewers should prioritize environmental control and interaction quality to maximize the clarity and effectiveness of survey questioning.

Overall, by systematically addressing interference and enhancing interviewer practices, survey administrators can significantly elevate the quality of the data they collect. These insights contribute to refining survey methodologies and ensuring that the conclusions drawn from social research are both credible and impactful.

A Appendix

A.1 Data Cleaning Notes

We began by importing the raw dataset using the read.csv() function from base R, with a focus on keeping the process flexible for further manual manipulation. Initially, we removed columns that contained more than 80% missing values to retain only the most relevant information for analysis.

To handle missing data, numeric variables were filled with their median values, while categorical variables were imputed with their mode. This approach aimed to maintain data integrity while minimizing assumptions about the distribution of missing values.

Subsequently, additional data cleaning measures were taken. We removed duplicate entries to ensure uniqueness and eliminated outliers in numeric columns using the Z-score method, retaining only those rows with Z-scores within ± 3. This step was important for reducing the influence of extreme data points on subsequent analyses.

To enhance clarity and usability, several column names were renamed to be more descriptive. For example, resundq was renamed to respondent_understood. Additionally, the respondent_understood variable was recoded into a binary form (Understood vs. Not_understood) to simplify future analyses.

To further refine the dataset, we consolidated categories for variables with more than five distinct values, grouping the least frequent values into an "Other" category. The country variable was excluded from this process to preserve its geographic specificity, which was deemed valuable for the analysis.

After cleaning, the dataset was split into training and testing sets in a way that ensured the country variable was well-represented in both. This stratified split preserved regional diversity, which was essential for maintaining representativeness in model training and evaluation.

Lastly, the cleaned dataset was saved in both Parquet and CSV formats for use in future stages of the analysis. This ensured compatibility across different software tools and provided efficient access for future work.

A.2 European Social Survey Methodology

The European Social Survey (ESS) is a cross-national survey conducted biennially across European countries, aimed at measuring the attitudes, beliefs, and behaviors of European citizens. To ensure data accuracy and quality, the ESS employs rigorous survey methodologies and standardized procedures, ensuring comparability across participating countries and providing a reliable data source for social science research and policymaking (European Social Survey, 2024).

The ESS targets individuals aged 15 and older living in private households across participating European countries, using a carefully designed sampling framework to ensure national representativeness. Countries employ multi-stage probability sampling to select respondents, ensuring that each individual has an equal chance of being selected, thereby minimizing selection bias and enhancing the generalizability of survey results.

To collect the data, the ESS uses trained interviewers to conduct face-to-face interviews. These interviews are conducted in the respondents' native language, ensuring that participants can fully understand the questions being asked. The questionnaires are translated following rigorous standardized guidelines to ensure equivalence across languages and cultures (European Social Survey, 2024). Additionally, questionnaires are pre-tested before the official survey is launched to identify and resolve potential misunderstandings, ensuring consistent understanding among respondents.

During data collection, the ESS implements strict quality control procedures, including systematic interviewer training, ensuring they conduct interviews following standardized techniques to reduce bias and maintain data consistency. To ensure sample representativeness, stratified sampling is used in some countries based on demographic characteristics like age, gender, and region, ensuring sufficient representation of different population groups.

Like many survey research initiatives, the ESS faces challenges related to declining response rates. To mitigate non-response, the ESS requires interviewers to make multiple contact attempts, particularly in cases where initial attempts are unsuccessful, including follow-up visits. To further reduce non-response bias, post-stratification weights are applied to the collected data to adjust for differences between the sample and the overall population.

The survey design of the ESS aims to provide a comprehensive overview of social attitudes while maintaining brevity to minimize respondent fatigue. Most interviews are conducted within a reasonable timeframe to ensure high response quality. The ESS places significant emphasis on accurately capturing every viewpoint, ensuring that the voices of marginalized groups are also adequately represented, which helps policymakers and researchers better understand the dynamics of social attitudes in Europe.

By maintaining strict methodologies regarding sampling frameworks, translation procedures, interviewer training, and addressing non-response bias, the European Social Survey has become a trusted source of social data, providing essential support for research into cross-national comparisons and understanding social attitudes.

A.3 Ideal Methodology

For studying factors that affect respondent understanding during social surveys, an ideal methodology would involve using data from the European Social Survey (ESS). The goal is to make sure we represent people from different countries across Europe fairly, considering

things like age, gender, education, and location, so that our findings are accurate and reflect various social contexts.

Sampling Method: We'll use stratified random sampling, which means we'll group people by country first, and then by characteristics like age, gender, and education within each country. This helps make sure that each group is represented well in the sample, reducing bias and making our findings more generalizable to the whole population.

Sample Size: To make sure our analysis is meaningful, we plan to include at least 1,500 participants per country. For larger and more diverse countries, we may increase this number to properly represent smaller groups.

Data Collection Approach: We'll use a mixed-method approach for data collection, including face-to-face interviews and self-completed questionnaires to reach different types of people. Interviews will be conducted by trained local interviewers to prevent language barriers or cultural differences from affecting the data quality. We'll also use self-administered questionnaires for follow-ups to improve the response rate among harder-to-reach groups, like younger people or those who move around frequently.

Survey Process and Quality Checks: Interviewers will use standardized scripts to make sure everyone hears the questions the same way. Surveys will be recorded digitally to make data handling easier and to minimize errors. To avoid any bias, interviewers are trained to read questions in a neutral manner.

Oversampling Certain Groups: To better understand specific impacts on comprehension, we will oversample some groups, like the elderly or those with lower education levels. This gives us more precise data about these groups, but we will use weight adjustments afterward to ensure the final dataset reflects the actual population structure.

Strategies to Improve Response Rates: We'll use several strategies to get as many people as possible to participate:

- **Multi-language Support**: Questionnaires will be available in different languages so that non-native speakers can participate.

- **Follow-up Contact**: We'll follow up with those who initially refuse or don't respond, trying at least twice to increase participation.

- **Incentives**: We'll offer small rewards like gift cards or raffle entries to encourage people to take part.

Data Analysis and Verification: Before analyzing, we'll clean up the dataset, deal with missing values, and make sure demographic information is accurate. We'll use statistical adjustments to correct for any missing data, and weighting will be used to ensure that our data accurately represents the population.

A.3.1 Budget Breakdown for Survey Implementation

To ensure we collect and analyze quality data, we will allocate the budget as follows:

1. **Survey Design**: $8,000

   - This budget includes designing the questionnaire, translating it into multiple languages, and making sure the questions are clear and unbiased. It also covers interviewer training costs.

2. **Sampling and Recruitment**: $15,000

   - This will go towards sampling costs, including **stratified sampling** to make sure all groups are represented proportionally. It also includes acquiring relevant demographic data for the 23 countries participating in ESS.

3. **Recruiting Respondents**: $30,000

   - This covers contacting participants, managing interviews, and compensating them for follow-ups across the countries.

4. **Incentives**: $20,000

   - Incentives are crucial for getting people to take part, especially those harder to reach. We plan to provide small rewards like gift cards or raffle opportunities to encourage participation.

5. **Data Quality Assurance and Analysis**: $12,000

   - This includes data cleaning, quality checks, and analysis costs. We'll hire statistical consultants to ensure proper data handling, make sure there are no inconsistencies, and apply adjustments to correct biases.

By using this methodology, we aim to get reliable and fair insights into how well people understand survey questions across different contexts, especially in the diverse societies of Europe.

A.4 Idealized Survey

An example of the survey can be found at this link: https://docs.google.com/forms/d/1BWPU4vFYpzYnkg1nzvXu9b_P12UOXFKcOh4muV7wVQ0/prefill

2024 Comprehension and Interview Experience Survey
Conducted by: DEZHEN CHEN
Contact: dezhen.chen@mail.utoronto.ca

Welcome, and thank you for taking part in the **2024 Comprehension and Interview Experience Survey**. Your participation will help us understand how respondents perceive and engage with survey questions, and will improve the quality of interview interactions.

As a thank you for your time, all survey respondents will be automatically entered into a sweepstakes, with a **chance to win $200**. Ten winners will be selected randomly.

Your responses will remain strictly confidential and anonymous, and any data collected will be securely stored and used solely for academic and research purposes. No identifying information will be connected to your survey responses.

A.5 Section 1: Demographics

Country of Residence:
Please specify your country of residence:
- [Your answer]

A.6 Section 2: Interviewer Details

Interviewer Age:
What is the age of the interviewer?
- [Please specify in years]

Interviewer Gender:
What is the gender of the interviewer?
1. Male
2. Female

A.7 Section 3: Interview Process Information

Interview Conduct Method:
What was the method used to conduct this interview?
1. Face-to-face
2. Telephone
3. Online video interview
4. Other

Questionnaire Language:
What language was used during the interview?
- [Please specify the language]

Question Clarification:
Did you need clarification on the questions during the interview?
1. No clarification needed
2. Minor clarification needed
3. Moderate clarification needed
4. Extensive clarification needed
5. Completely unclear, repeated explanations needed
6. Other

Respondent Understanding Level:
How well do you think the respondent understood the questions?
1. Fully understood
2. Mostly understood
3. Needed some clarification
4. Partially understood
5. Did not understand at all

Respondent Tried Their Best:
Do you think the respondent tried their best to answer the questions?
1. Very unwilling to answer
2. Somewhat reluctant
3. Neutral
4. Willing to answer
5. Very willing to answer
6. Other

A.8 Section 4: Interference During Interview

Interference Presence:
Was there any interference during the interview process?
1. Yes
2. No

Partner Interference:
Was there any interference by a partner during the interview?
1. Yes
2. No

Child Interference:
Was there any interference by children during the interview?
1. Yes
2. No

Parent Interference:
Was there any interference by parents during the interview?
1. Yes
2. No

Relative Interference:
Was there any interference by other relatives during the interview?
1. Yes
2. No

Non-relative Interference:
Was there any interference by non-relatives during the interview?
1. Yes
2. No

Unknown Interference:
Was there any unknown source of interference during the interview?
1. Yes
2. No

Interference Not Applicable:
Was interference not applicable during this interview?
1. Yes
2. No

A.9 Section 5: Additional Feedback

Additional Comments:
Do you have any suggestions or feedback regarding the interview process?
- [Please provide your comments]

Interference Details:
If there was interference, could you please describe it in more detail?
- [Please provide your comments]

Suggestions for Improvement:
Do you have any suggestions to improve the clarity or format of the questions in this survey?
- [Please provide your suggestions]

A.10 Important Notes

- All responses are confidential and will be used for research purposes only.
- This survey will take approximately 10 minutes to complete.
- Your participation is entirely voluntary, and you may withdraw at any time.
- All questions marked with an asterisk (*) are required.
- There are no right or wrong answers; we appreciate your honest opinions.

Thank You Message:
Thank you once again for your participation. Your feedback is invaluable in helping us improve the quality of survey interactions and comprehension. We appreciate your time and effort in completing this survey.

B Model details

B.1 Model summary

Table 7 presents the coefficients from our model analyzing factors that influence response comprehension during interviews. The intercept value, 1.823, represents the baseline level of response comprehension when all other predictors are at their reference or baseline levels. Key variables include "interview_conduct_method9," which has a highly positive coefficient of 81.309, indicating that this particular interview method greatly enhances respondent comprehension. In contrast, factors such as "parent_interference" and "child_interference" exhibit

Table 7: Coefficients from a Bayesian logistic regression model examining factors affecting respondents' understanding during interviews, including interview methods, interference types, and country-level indicators.

	Bayesian Logistic Regression
(Intercept)	1.823
interview_conduct_method2	0.113
interview_conduct_method3	−0.179
interview_conduct_method9	81.309
partner_interference1	−0.601
child_interference1	−0.623
parent_interference1	−1.586
relative_interference1	−0.609
non_relative_interference1	0.941
interviewer_age	0.003
interviewer_gender2	0.100
as.factor(country)CH	0.695
as.factor(country)DE	−1.302
as.factor(country)FI	−0.998
as.factor(country)GB	−0.941
as.factor(country)HR	−1.338
as.factor(country)HU	−1.262
as.factor(country)IE	−1.274
as.factor(country)LT	−0.633
as.factor(country)NL	−0.795
as.factor(country)NO	−1.286
as.factor(country)SI	−2.181
as.factor(country)SK	−0.697
question_clarification2	−0.284
question_clarification3	−1.205
question_clarification4	−2.965
question_clarification5	−2.969
question_clarificationOther	−2.038
respondent_reluctant2	−0.138
respondent_reluctant3	−0.564
respondent_reluctant4	−0.418
respondent_reluctant5	−0.120
respondent_reluctantOther	0.708
respondent_tried_best2	0.308
respondent_tried_best3	0.528
respondent_tried_best4	2.943
respondent_tried_best5	4.204

negative coefficients (-1.586 and -0.623 respectively), suggesting that these types of interference are associated with lower levels of understanding during interviews.

Additionally, non-relative interference, which shows a positive coefficient of 0.941, suggests that having individuals outside the immediate family present may be less disruptive, or even beneficial, for the comprehension of survey questions. Interviewer-related variables, such as "interviewer_gender," appear to have minimal effects on comprehension, as evidenced by small coefficients like 0.100, suggesting that interviewer attributes play a relatively minor role compared to environmental and relational factors during social surveys ## VIF

Table 8: Variance Inflation Factors (VIF) for Each Predictor in the Model.

	Variable	GVIF	Df	GVIF..1..2.Df..
interview_conduct_method	interview_conduct_method	1.755	3	1.098
partner_interference	partner_interference	1.061	1	1.030
child_interference	child_interference	1.049	1	1.024
parent_interference	parent_interference	1.026	1	1.013
relative_interference	relative_interference	1.082	1	1.040
non_relative_interference	non_relative_interference	1.088	1	1.043
interviewer_age	interviewer_age	1.342	1	1.159
interviewer_gender	interviewer_gender	1.236	1	1.112
as.factor(country)	as.factor(country)	4.256	12	1.062
question_clarification	question_clarification	2.103	5	1.077
respondent_reluctant	respondent_reluctant	2.725	5	1.105
respondent_tried_best	respondent_tried_best	3.186	5	1.123

Table 8 presents the Generalized Variance Inflation Factors (GVIF) for each predictor in the Bayesian logistic regression model. GVIF helps assess multicollinearity among predictors, indicating how much the variance of a regression coefficient is inflated due to correlations between variables.

Most predictors, such as respondent_tried_best (GVIF = 3.186, Df = 5) and respondent_reluctant (GVIF = 2.725, Df = 5), show moderate GVIF values, suggesting manageable multicollinearity. The adjusted GVIF values, which are generally close to 1, indicate that the inflation effect is controlled across variables.

In summary, the GVIF analysis indicates that multicollinearity is present but not severe enough to threaten model reliability. Further analysis may focus on predictors with slightly higher GVIF to ensure stability.
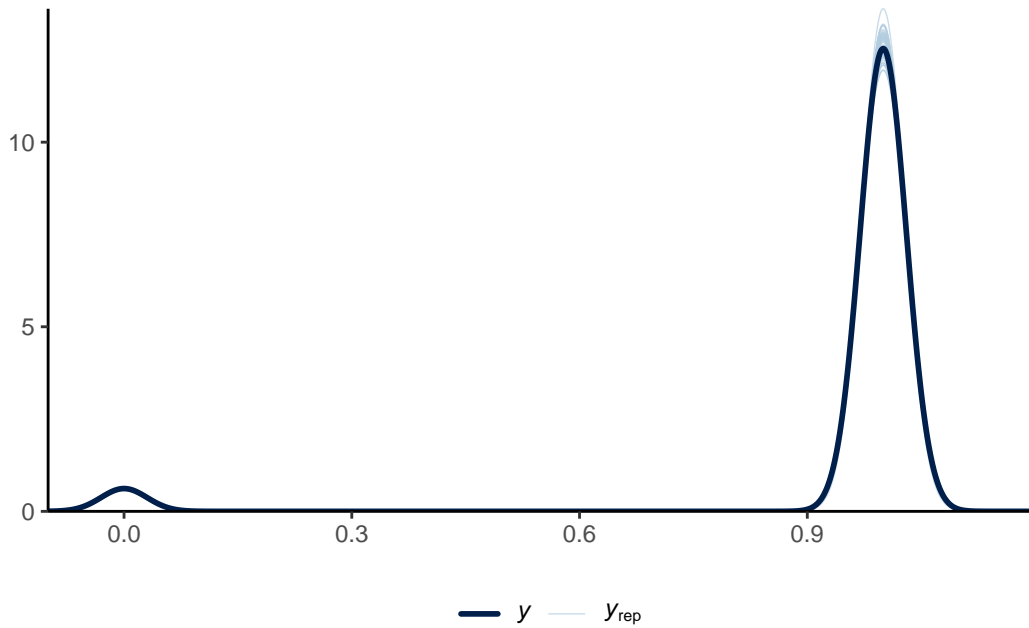
Figure 5: Posterior Predictive Check: The comparison between observed response values (dark line) and model-generated replications (light lines) demonstrates the model's ability to replicate the observed response patterns, indicating a good fit and capturing the main features of the response distribution.

B.2 Posterior predictive check

In Figure 5, we implement a posterior predictive check, which displays the overlap between the observed data (denoted by y and the replicated data generated from the model (denoted by (y_rep). The replicated lines closely follow the shape and central trend of the observed data, particularly around the main peak at 0.9, which suggests that the model captures the overall distribution effectively. This alignment indicates that the model successfully represents the underlying data-generating process, especially in regions with high data density. While the model's performance appears strong, any visible discrepancies—particularly in areas of lower density—may highlight aspects of the data structure that are not fully captured. Addressing such discrepancies could involve refining the model by introducing additional parameters or reconsidering underlying assumptions. Overall, this comparison provides evidence that the model fits well, though there remain potential areas for enhancement to better encompass the full range of data complexities.

B.3 Diagnostics

B.3.1 Validation

(a) Residual Histogram
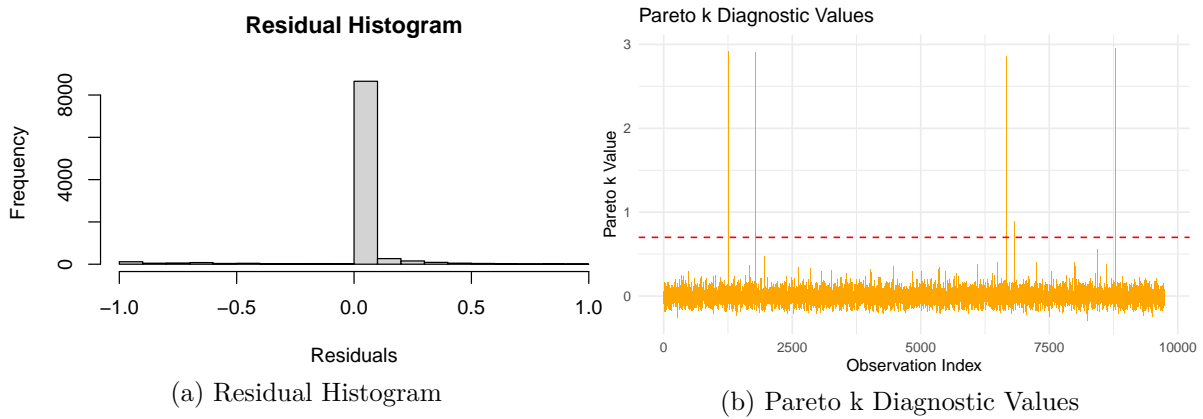
(b) Pareto k Diagnostic Values

Figure 6: Comparing Diagnostic Visualizations for Model Checking

In Figure Figure 6a, the residual histogram displays the distribution of residuals from the model. The concentrated frequency around zero suggests that the model effectively captures much of the variability in the data, with fewer extreme deviations, thereby indicating a reasonable level of fit.

Figure Figure 6b presents the Pareto k diagnostic values, which help assess the reliability of the estimated leave-one-out cross-validation (LOO) predictions. The majority of Pareto k values fall below the threshold of 0.7, indicating that the model is well-calibrated with only minimal issues in terms of influential observations.

38

(a) Leave-One-Out Cross Validation (ELPD_LOO)
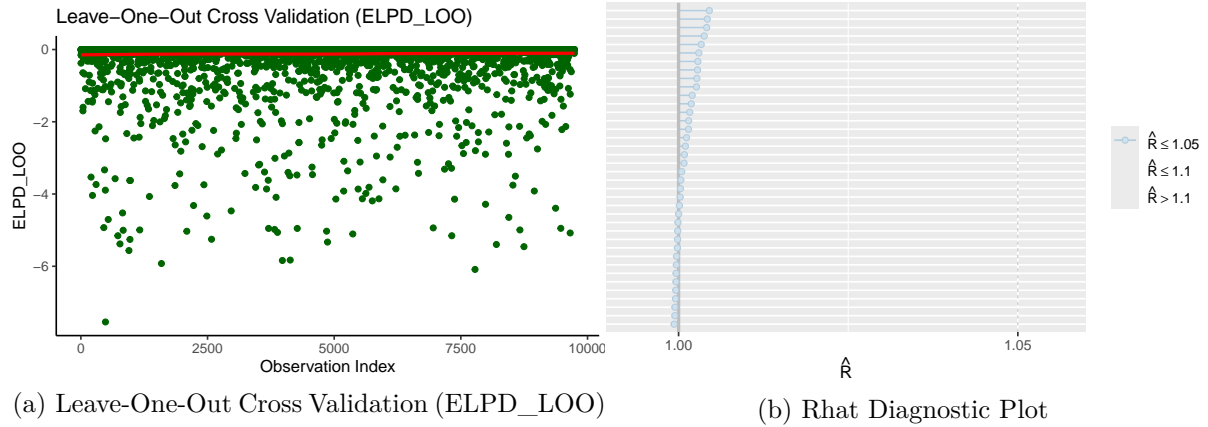
(b) Rhat Diagnostic Plot

Figure 7: Comparing Diagnostic Visualizations for Model Checking

Figure Figure 7a is a scatter plot for leave-one-out cross-validation (ELPD_LOO) estimates. The consistency of the points around the red loess smooth line suggests that the model's predictions are largely consistent across different observations, thus reinforcing the model's robustness.

Figure Figure 7b is an Rhat plot, showing that all values of the Rhat diagnostic are close to 1. This indicates good mixing of the Markov Chain Monte Carlo (MCMC) chains and suggests that the model has likely converged appropriately, making the posterior estimates trustworthy.

References

Alexander, Rohan. 2023. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellingstorieswithdata.com.

Apache Arrow. 2021. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Aquilino, William S. 1993. "Effects of Spouse Presence During the Interview on Survey Responses Concerning Marriage." *Public Opinion Quarterly* 57 (3): 358–76.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models.* https://CRAN.R-project.org/package=broom.mixed.

European Social Survey. 2024. "ESS Data File: 0c3f5fda-4fe0-42f2-B4c8-4ef093c0508f." https://ess.sikt.no/en/datafile/0c3f5fda-4fe0-42f2-b4c8-4ef093c0508f/7?tab=0&elems=de8c08b6-e406-4c56-92ee-4272818cae96_1.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression.* Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gelman, Gabriel, Jonah Gabry, et al. 2021. *Bayesplot: Plotting for Bayesian Models.* https://mc-stan.org/bayesplot.

Groves, Robert M, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey Methodology.* John Wiley & Sons.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Massey, Douglas S, and Roger Tourangeau. 2013. "Where Do We Go from Here? Nonresponse and Social Measurement." *The ANNALS of the American Academy of Political and Social Science* 645 (1): 222–36.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Team, Stan Development. 2021. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

Therneau, Terry, and Beth Atkinson. 2023. *Rpart: Recursive Partitioning and Regression Trees.* https://CRAN.R-project.org/package=rpart.

Vehtari, Aki, Jonah Gabry, Måns Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. 2024. "Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models." https://mc-stan.org/loo/.

Wickham, Hadley. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* https://ggplot2.tidyverse.org.

Wickham, Hadley et al. 2021. *Tidyverse: Easily Install and Load the 'Tidyverse'.* https://tidyverse.tidyverse.org.

Wilke, Claus O. 2024. *Ggridges: Ridgeline Plots in 'Ggplot2'.* https://CRAN.R-project.org/package=ggridges.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.*

https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.