

Introduction

Lors d'enquêtes longitudinales, les premiers participants ne sont pas toujours disponibles par la suite et **l'échantillon s'amenuise**. On parle alors d'**attrition**. De plus, il existe des problèmes de **non-réponse à certaines questions**. Nous étudions les méthodes de traitement de ces données manquantes et cherchons des solutions performantes dans le cas longitudinal.

Motivations

Notre travail est motivé par l'enquête "**devenir parent**" menée par le centre lémanique d'étude des parcours et modes de vie (PAVIE). Durant cette enquête, des couples devaient répondre à des questions à trois reprises (vagues 1, 2 et 3). Sur 236 couples ayant participé en vague 1, seulement 168 ont répondu à l'appel en vague 2 et environ 174 lors des derniers entretiens.

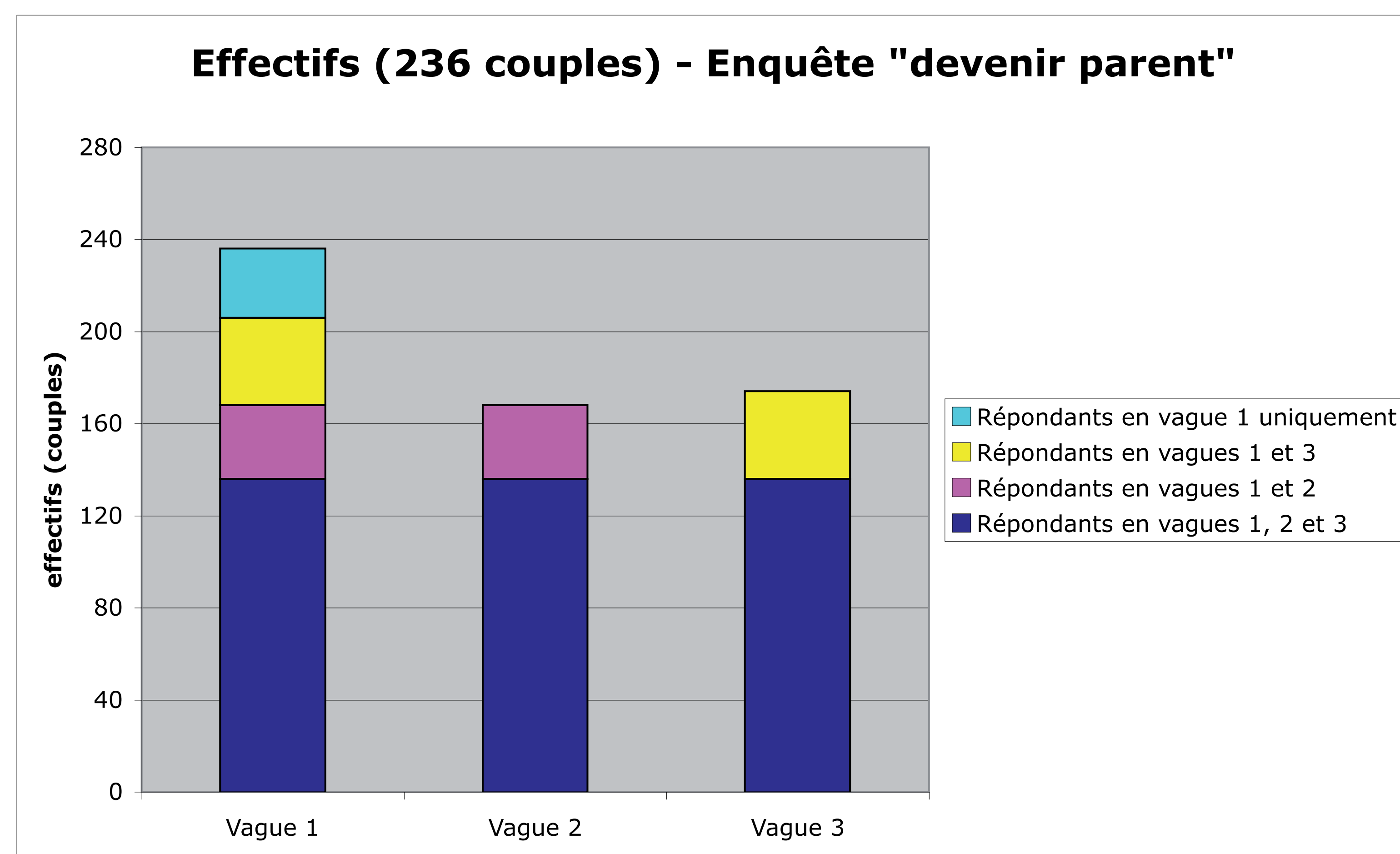


FIGURE 1: Nombre de couples participant à l'enquête "devenir parents" en fonction de la vague.

Enjeux

L'attrition est responsable de **biais internes et externes** : cela introduit une distorsion des relations entre les variables. De plus, si l'échantillon est réduit systématiquement selon certaines caractéristiques des individus, les conclusions théoriques ne seront plus valides. **L'analyse des données est compromise et peut mener à des conclusions parfois aberrantes.**

Méthode

Certaines méthodes sont plus adéquates que d'autres pour traiter le problème des données manquantes : par exemple, **ne pas systématiquement réduire l'échantillon** à ceux qui ont répondu à toutes les questions, et **éviter de remplacer les données manquantes par la moyenne ou la médiane**.

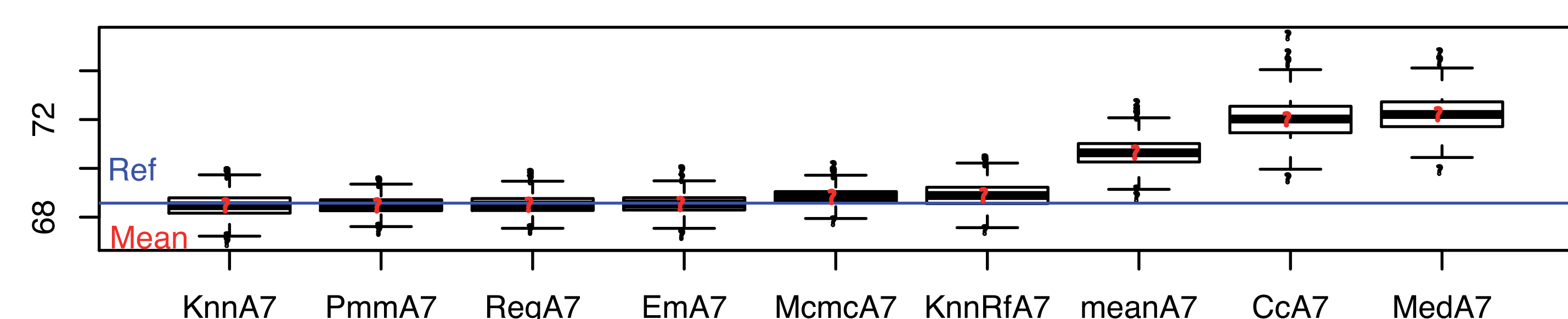


FIGURE 2: Moyenne de la variable calculée sur 1000 échantillons à l'aide de différentes méthodes. PmmA7 : imputation multiple + predictive mean matching ; EmA7 : imputation multiple + EM ; MeanA7 : imputation par la moyenne ; CcA7 : aucune imputation ; MedA7 : imputation par la médiane.

Dans le cas longitudinal, nous comparons certaines méthodes déjà étudiées ainsi que des méthodes plus adaptées au longitudinal, comme le remplacement par l'observation précédente (**LOCF**) ou par la **régression**. Nous attendons des résultats mitigés pour ces deux méthodes, puisque l'enquête porte sur les changements.