

# La standardisation, outil d'analyse de données transversales en épidémiologie

par B. Junod et V. Wietlisbach

Institut universitaire de médecine sociale et préventive, Lausanne

## 1. Introduction

L'épidémiologie descriptive se fonde le plus souvent sur l'analyse statistique de collections de données de type transversal, c'est-à-dire recueillies à un moment donné. Ce sont par exemple les résultats d'une enquête, d'un examen de dépistage ou bien les données de prévalence d'un registre des tumeurs. Certaines variables comme l'exposition à un facteur de risque ou la nature d'une affection sont alors susceptibles d'être mises en relation avec d'autres variables qui décrivent l'individu concerné, tels l'état-civil, la classe d'âge, la région du domicile ou la profession. L'intérêt de la description consiste principalement à comparer entre différentes catégories de la population certains taux ou certaines valeurs moyennes caractéristiques, et aussi à mettre en évidence des associations statistiques entre variables. Cet article s'attache à montrer quel peut être l'apport de la méthode de la standardisation dans le domaine.

## 2. Comparaison de valeurs standardisées

Pour illustration, on suppose qu'en Suisse une enquête nous donne les consommations moyennes d'alcool par région  $i$  et classe d'âge  $j$ , notées  $\bar{y}_{ij}$ . La différence entre une de ces consommations spécifiques et la consommation suisse moyenne (toutes régions et classes d'âge réunies) peut être attribuée à trois facteurs, explicités par le modèle suivant:

$$\bar{y}_{ij} - \bar{y} = a_i + b_j + c_{ij}$$

où  $a_i$  représente l'effet propre de la région  $i$  et  $b_j$  l'effet propre de la classe d'âge  $j$ , alors que  $c_{ij}$  est l'effet de l'interaction des deux. La comparaison de la consommation d'alcool entre régions peut se faire à l'aide des moyennes simples:

$$\bar{y}_i = \sum_j (n_{ij}/n_{i.}) \bar{y}_{ij}$$

ou des moyennes standardisées par l'âge (avec la méthode directe):

$$\bar{y}_i(st) = \sum_j (n_{.j}/n_{..}) \bar{y}_{ij}$$

où les  $n_{ij}$  représentent les nombres de personnes dans les différentes catégories. Les valeurs standardisées ont l'avantage d'utiliser les pondérations d'une même structure par âge pour toutes les régions, en l'occurrence celle de l'ensemble de la population. En reprenant le modèle d'interaction, la comparaison par soustraction entre deux régions 1 et 2 nous donne:

$$D(st) = \bar{y}_2(st) - \bar{y}_1(st) \\ = (a_2 - a_1) + \sum_j (n_{.j}/n_{..}) (c_{2j} - c_{1j})$$

On remarque que cette différence permet d'éliminer l'effet direct de l'âge par l'annulation des  $b_j$ , mais qu'elle tient compte de la variation de l'interaction entre l'âge et la région. De cette manière, on élimine la différence de consommation d'alcool entre régions attribuable à la simple différence de leur

structure par âge.

## 3. Mise en évidence de l'effet de la variable de standardisation

La soustraction de  $D(st)$  à la différence des moyennes simples, notée  $D = \bar{y}_2 - \bar{y}_1$ , donne:

$$d = D - D(st) \\ = \sum_j (n_{2j}/n_{2.} - n_{1j}/n_{1.}) b_j \\ + \sum_j (n_{2j}/n_{2.} - n_{.j}/n_{..}) c_{2j} \\ - \sum_j (n_{1j}/n_{1.} - n_{.j}/n_{..}) c_{1j}$$

Cette différence des différences ( $d$ ) représente un indice de l'effet de l'âge sur la consommation d'alcool ( $b_j$ ) et de son interaction avec la région ( $c_{ij}$ ). Une condition nécessaire en est que la structure par âge diffère d'une région à l'autre de manière sensible. En plus,  $d$  est indépendant des effets propres des régions ( $a_i$ ); ceux-ci sont souvent liés aux conditions d'observation qui varient de région en région et y entraînent parfois des tendances systématiques à la hausse ou à la baisse.

## Summary

The direct method of standardization is mainly applied for eliminating undesirable effects of confounding variables in the comparison of several rates or means. This paper stresses the usefulness of standardization for detecting the statistical association between two variables. Under the assumption of a simple model implying two main effects and an interaction, it is discussed how the comparison between crude and standardized values measures altogether the main effect of the variable and the interaction. Furthermore, the variation in the population of the differences between crude and standardized values is shown to be an index of the effect of the variable for which it has been standardized.

## Bibliographie

JUNOD B., WIETLISBACH V.: Programme de standardisation des taux; mode d'emploi. Institut universitaire de médecine sociale et préventive, Lausanne (1978)