ont un caractère hypothético-déductif. Qu'est-ce que cela signifie ?

Pour déterminer si le prénom Irène cause la mention « Très bien », un raisonnement hypothético-déductif débuterait par la formulation d'une hypothèse sur les causes de la mention (lesquelles pourraient inclure le capital culturel des parents, la réputation de l'établissement fréquenté, l'âge du candidat, etc.) et sur les relations causales qu'elles entretiennent avec le prénom. Généralement, on nomme « modèle causal » une telle hypothèse, que l'on représente par un graphe : les nœuds de ce graphe correspondent aux phénomènes pris en compte (le prénom, la mention, les causes de la mention…) et les flèches d'un nœud à l'autre représentent des relations de cause à effet *(voir la figure 4)*.

En utilisant les théories probabilistes de la causalité, on tire des conclusions probabilistes d'un tel modèle. Si ces conclusions contredisent les observations, le modèle est rejeté ; si elles sont valides, il peut être considéré comme confirmé. Ici, le problème de la circularité et celui de l'absence de principe de clôture sont résolus dès lors que l'on définit un modèle causal. D'une part, en effet, tout modèle causal porte sur un certain nombre de phénomènes bien déterminés et, de ce fait, il véhicule une réponse à la question de la clôture laissée ouverte par les théories probabilistes. D'autre part, pour chacun des phénomènes qu'il représente, le modèle formule une hypothèse relativement à l'ensemble de ses causes. La circularité des théories probabilistes n'intervient donc plus.

Reste que l'expérimentation et l'hypothético-déduction ont chacune des limites. Un essai randomisé n'est que rarement envisageable – par exemple, il ne l'est pas dans le cas du prénom et de la mention. De même, le succès des approches hypothético-déductives dépend largement des hypothèses qui auront pu être formulées.

Toutefois, des algorithmes développés depuis les années 1990 par les équipes de Judea Pearl, à l'Université de Californie à Los Angeles, et de Clark Glymour, Peter Spirtes et Richard Scheines, à l'Université Carnegie Mellon de Pittsburgh, visent à contourner les obstacles que nous avons mis en évidence d'une façon qui soit à la fois observationnelle et inductive.

Des algorithmes utiles

Ces algorithmes prennent pour entrée les corrélations entre phénomènes d'un ensemble préalablement fixé. Le problème constitué par l'absence de principe de clôture est donc résolu exactement de la même façon que dans le cadre hypothético-déductif : en isolant d'emblée un ensemble fini de phénomènes. Mais ici, aucune hypothèse causale n'est formulée. À partir des seules corrélations – les informations en termes de probabilités –, on construit tous les graphes susceptibles d'être le modèle causal correct pour l'ensemble de phénomènes considéré. L'obstacle de la circularité des théories probabilistes de la causalité est donc contourné.

Ces algorithmes font appel à une analyse un peu différente (et un peu moins satisfaisante) de celle qu'utilisent les théories probabilistes circulaires, avec la conséquence importante que des informations purement probabilistes suffisent à déterminer si deux phénomènes de l'ensemble considéré ont un lien de causalité direct, sans l'intermédiaire d'un autre phénomène.

Cependant, l'analyse qui fonde ces algorithmes admet des contre-exemples, qui ont fait l'objet d'un vif débat dans les années 2000. Notamment, elle suppose

LES ESSAIS CLINIQUES RANDOMISÉS

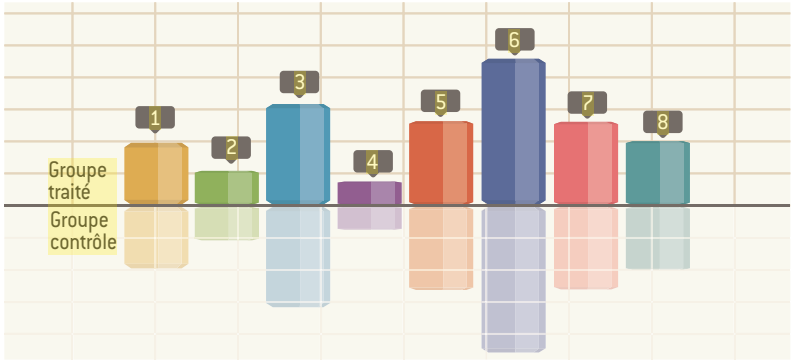
Le premier essai clinique randomisé ayant donné lieu à une publication scientifique a été mené en Grande-Bretagne dans les années 1940*. L'étude portait sur l'effet d'un antibiotique, la streptomycine, sur la tuberculose pulmonaire.

Cent neuf patients, âgés de 15 à 25 ans, y ont participé. Ils ont été répartis en deux groupes, l'un recevant le traitement, l'autre non (le « groupe contrôle »). La répartition a été effectuée selon une procédure aléatoire garantissant que deux patients du même sexe et relevant du même hôpital aient la même probabilité d'appartenir au groupe recevant la streptomycine.

Cette façon de procéder vise à répartir de manière similaire dans les deux groupes les autres facteurs que la streptomycine, le sexe et l'hôpital de prise en charge susceptibles d'affecter causalement l'évolution de l'état de santé des patients.

Imaginons que, dans le cas considéré, il existe trois facteurs de ce type : l'âge, le régime alimentaire et le niveau culturel. Si chacun de ces facteurs est susceptible de deux modalités, huit situations peuvent se présenter. La randomisation (le choix aléatoire des patients à traiter parmi les participants à l'essai clinique) vise à faire en sorte que ces huit situations soient représentées de manière comparable dans le groupe traité et dans le groupe contrôle *(voir la figure)*. Cette procédure est désormais la règle dans les essais cliniques rigoureux, qui doivent par ailleurs porter sur un nombre suffisant de patients pour que la randomisation soit efficace.

*G. Marshall *et al.*, *British Medical Journal*, vol. 2(4582), pp. 769-782, 1948.



EN SUPPOSANT QUE L'ÂGE, le régime alimentaire et le niveau culturel puissent influer sur l'évolution de l'état de santé, et que chacun de ces facteurs ait deux valeurs pertinentes seulement, l'ensemble des patients constitue huit sous-populations. Grâce à la randomisation, les individus de chaque sous-population sont à peu près également répartis entre le groupe traité et le groupe contrôle.