

DE LA DÉMARCHE SCIENTIFIQUE EN PÉRIODE DE CRISE

Arthur Charpentier

Professeur, Université du Québec à Montréal

Dans une conférence donnée le 13 février 2020 ⁽¹⁾, intitulée « Contre la méthode », Didier Raoult affirmait : « moi je n'ai jamais fait d'essais randomisés [...] faire ça sur des maladies infectieuses, ça n'a pas de sens ». Cette vision était reprise dans une tribune plus détaillée, où face à « la méthode » (et aux mathématiques), Didier Raoult défendait (ce qu'il appelait) « la morale [et] l'humanisme » du serment d'Hippocrate. Comme il le rappelle, faire des groupes de contrôle, c'est « dire au malade qu'on va lui donner au hasard, soit le médicament dont on sait qu'il marche, soit le médicament dont on ne sait pas s'il marche » [Raoult, 2020a-2020b]. Alors que cette méthode d'expériences randomisées est aujourd'hui saluée dans toutes les disciplines – comme le rappelle le prix Nobel d'économie attribué en 2019 à Esther Duflo, Michael Kremer et Abhijit Banerjee –, comment un chercheur peut-il prendre une telle position, aujourd'hui ?

Observation, expérimentation, statistique et causalité

La statistique peut être vue comme l'ensemble des outils mathématiques que l'on peut utiliser pour extraire de l'information d'un ensemble de données. Par exemple, on peut se demander si, à la suite d'une opération chirurgicale, le patient doit récupérer chez lui ou à l'hôpital (en vue d'un rétablissement plus rapide ou plus efficace). Le statisticien va naturellement collecter des données, avec autant d'opérations que possible, savoir si les gens ont récupéré chez eux ou à l'hôpital, peut-être

trouver quelques variables supplémentaires – comme la nature de l'opération, l'âge du patient ou la distance entre son domicile et l'hôpital le plus proche (en cas de complication) –, et chercher un indicateur de réussite, comme le fait qu'il ait fallu refaire une opération dans les mois qui ont suivi. Cette démarche peut sembler naturelle, mais permet-elle pour autant de répondre à la question ? Dans ces données collectées, dites d'observation, ne peut-on pas s'attendre à avoir un biais de sélection ? Les personnes qui ont été renvoyées chez elles n'étaient-elles pas en meilleure santé que celles qui sont restées plus longtemps à l'hôpital ? Car la question que se pose le statisticien est fondamentalement « que se serait-il passé si la personne avait choisi l'autre option ? » On parle ici d'inférence

causale, et le patient choisit une option, jamais les deux. On ne peut pas savoir ce qu'aurait donné l'autre option, qui n'a pas été retenue. L'effet causal d'une option, ou d'un traitement, n'est jamais observé car il s'agit de la différence potentielle entre deux alternatives.

La méthode la plus simple consiste à prendre deux patients proches (à défaut d'être identiques), et de voir comment les deux répondent aux deux options, aux deux traitements. John Stuart Mill parlait de « méthode des différences ». La proximité est basée sur les covariables comme l'âge, le genre ou diverses caractéristiques. L'idée des expériences randomisées est de ne plus se contenter d'observer, mais de choisir l'option ou le traitement, de manière aléatoire (et de ne laisser choisir ni le patient ni le médecin). Chaque personne se voit alors affectée dans un groupe, indépendamment de ses caractéristiques. Idéalement, John Stuart Mill voulait comparer des personnes identiques dans le cadre de choix alternatifs : il voulait qu'il n'y ait pas de différences non mesurables avant le traitement entre les personnes recevant le traitement ou le groupe témoin. Si le concept philosophique est louable, John Stuart Mill n'a pas fourni de méthode pour créer cette situation idéale. Au cours du XX^e siècle, la statistique a montré qu'une affectation aléatoire dans les groupes de traitement – on pourrait imaginer davantage de choix que deux alternatives – était très prometteuse.

On parle d'expérience car l'assignation aléatoire ne se produit que rarement dans un contexte naturel. Les assignations se produisent souvent selon un biais de la personne qui prend la décision. Dans le cas d'un choix entre un traitement agressif et un traitement plus doux, on peut imaginer qu'il serait naturel de traiter moins agressivement un patient dont la maladie est moins grave, ou de traiter plus agressivement un patient qui pourrait survivre mais dont la maladie est plus grave, ou peut-être de traiter moins agressivement un patient qui n'a aucun espoir de survivre. Si les patients gravement malades reçoivent un traitement agressif (et les patients moins gravement malades reçoivent un traitement moins agressif), alors le traitement agressif pourrait paraître nuisible, alors qu'il

est, en réalité, bénéfique, les patients gravement malades étant, a priori, ceux qui ont le plus de chances de mourir. Pour reprendre l'idée de John Stuart Mill, on souhaitait comparer équitablement deux options, mais les comparaisons équitables sont rares, par nature, comme le rappelle Rosenbaum [2017]. Et la randomisation est une des méthodes les plus simples à mettre en œuvre ⁽²⁾, justement parce que le tirage au sort ne favorise aucun des deux traitements. Les concepts d'équité et de justice que nous évoquons ici sont simplement ceux de symétrie – Paternotte [2020] ou Ferry-Danini [2020] reviennent sur l'importance de l'éthique, point que nous effleurons par la suite. Cette symétrie dans la répartition se traduit par le fait qu'un patient a, a priori, autant de chances d'avoir un traitement qu'un autre. Dans les années 1920, on parlera d'« essais d'uniformité ». Les expérimentations aléatoires ont été initialement proposées en agriculture, dans des fermes expérimentales. On divisait la ferme en plusieurs parcelles, et le traitement était choisi de manière aléatoire (il pouvait s'agir d'un engrais et d'un insecticide). Certaines parcelles servaient de contrôle, et en comparant les deux, on pouvait quantifier l'efficacité du traitement, par rapport à la référence qu'est le contrôle – Dehue [2001] revient sur les premières expériences contrôlées.

Expérimentations sociales et évaluation des politiques publiques

Cette approche est aujourd'hui classique pour évaluer l'impact d'une politique publique : on constitue de manière complètement aléatoire deux groupes : un groupe test (qui profitera de la politique) et un groupe témoin (qui n'en bénéficiera pas). La politique peut être une hausse de rémunération de certains fonctionnaires via une prime, des prêts à des femmes dans certaines communautés, etc. À la fin du test, on va comparer les deux groupes afin de voir s'il y a eu – ou pas – un effet bénéfique, éventuellement en affinant

par sous-groupe pour comprendre à qui profite la politique, ou au contraire, à qui elle ne profite pas. On peut alors quantifier l'impact, rarement en expliquer les raisons (si cet impact est jugé significatif). Cela ne permet pas de savoir si une autre politique n'aurait pas eu d'autres effets, ou si un autre levier d'action n'aurait pas été plus efficace. On peut ainsi montrer qu'en rémunérant une visite chez le médecin pour se faire vacciner, le taux de mortalité chute dans certains pays en développement. C'est l'utilisation de ces méthodes qu'Esther Duflo, Michael Kremer et Abhijit Banerjee ont préconisé, pour mesurer l'impact des politiques sur la pauvreté.

Car les essais contrôlés randomisés sont la manière la plus rigoureuse pour déterminer s'il existe une relation de cause à effet entre un traitement et un résultat, et pour quantifier l'impact du traitement. La principale caractéristique est une répartition aléatoire entre les groupes d'intervention, mais elle n'est pas la seule. En médecine, il est souvent demandé que les patients et les chercheurs ne sachent pas quel traitement a été administré tant que l'étude n'est pas terminée (on parle alors d'études en double aveugle). Cette condition n'est souvent pas possible dans le cas des politiques publiques. Les avancées récentes ont permis d'aller plus loin, en relâchant plusieurs hypothèses, comme celle permettant d'adapter les tailles des groupes, dans le cas d'expériences répétées : si on voit qu'un traitement a des effets, mais que l'expérience doit se prolonger dans le temps, il est possible d'en faire profiter davantage de sujets. Pour reprendre la terminologie de Charpentier *et al.* [2020], il est possible de trouver un juste équilibre entre l'exploration et l'exploitation.

Expérimentation et médecine, le cas de la polio

Étrangement, les économistes ont longtemps dit que ces méthodes d'expériences randomisées venaient de la médecine, mais les premières grandes expériences en matière de santé publique datent seulement des années 1950.

En 1954, plus de 400 000 enfants américains ont participé à une expérience randomisée pour quantifier les effets d'un vaccin, développé par Jonas Salk, pour prévenir la poliomyélite. En moins d'un an, cette expérience a permis de trancher définitivement la question, ce qui était une première pour une question de santé publique. Dans les États qui ont participé à l'étude, la participation n'était pas obligatoire. Un peu plus de 200 000 enfants, sélectionnés au hasard, ont eu accès au vaccin, alors qu'un nombre presque équivalent a eu accès à un placebo, qui était alors une solution d'eau salée. Et un peu plus de 300 000, bien qu'éligibles, n'ont pas participé. On pourrait se poser la question de l'éthique d'une telle procédure, consistant à donner de l'eau salée alors qu'ils auraient pu avoir le vaccin. Mais cette vision est faussée, car basée sur le résultat du test : à l'époque, personne ne savait si le vaccin était efficace, et en particulier, s'il n'aurait pas d'effets secondaires néfastes. Ce principe d'incertitude – Freedman [1987] parlera d'équipose – est souvent considéré comme un prérequis éthique indispensable pour lancer une expérience. Dans l'échantillon ayant obtenu le vaccin (et selon les chiffres donnés par Brownlee [1955], Meier [1990] et Meldrum [1998]), 16 enfants sur 100 000 ont eu une polio paralysante, contre 57 dans le groupe avec le placebo. Statistiquement, une telle différence peut être jugée comme « significative ».

Pour l'anecdote, cette expérience randomisée n'était pas la solution envisagée initialement. La première idée était d'administrer le vaccin à tous les enfants de deuxième année, et d'utiliser les première et troisième années comme groupe témoin. Mais plusieurs chercheurs s'y sont opposés, notant que le passage de première en deuxième année se basait sur des notes, et que les meilleurs enfants (d'un point de vue scolaire) auraient accès au vaccin. Un autre souci était que si la variable de constitution des groupes était rendue publique, ceci aurait entraîné de la triche : les médecins qui vaccinaient connaissaient le niveau scolaire des enfants, et leur opinion sur le fonctionnement des vaccins aurait pu faire induire les résultats, dans un sens ou dans l'autre, parfois simplement en encourageant certains enfants à ne pas

participer à l'expérience. Dans la version randomisée, le choix de participer, ou pas, se faisait avant la constitution des groupes. La participation à l'expérience était liée à certaines variables (en particulier les mères dont le statut socioéconomique était plus faible pensaient davantage que les vaccins étaient dangereux, et ont plus souvent retiré leurs enfants), mais l'affectation dans les groupes était complètement aléatoire, et les deux groupes pouvaient être considérés comme comparables. Cette autosélection n'a causé aucun biais dans l'analyse, contrairement à ce qui se serait passé dans le cas non randomisé (3). Cette première grande expérience randomisée a permis d'établir de manière non ambiguë l'efficacité du vaccin dans la prévention de la polio, et a été un premier pas important dans l'éradication de la maladie. Oui, car la polio est une maladie infectieuse, et cette expérience a été une étape fondamentale dans la recherche scientifique, montrant l'importance de cette méthode quand elle est possible.

Quand la randomisation est impossible

Dans une étude dite d'observation, c'est-à-dire une étude des effets d'un traitement sans attribution aléatoire des traitements, une association entre le traitement reçu et le résultat observé est généralement ambiguë, comme nous l'évoquions initialement. Cette association pourrait refléter un effet causé par le traitement (ce que l'on espère en lançant l'étude), ou un biais non mesuré dans la manière dont les traitements ont été attribués, voire une combinaison des deux. Si Ronald Fisher a posé les bases mathématiques de l'expérimentation aléatoire, William Cochran a formalisé les méthodes d'échantillonnage, expliquant comment analyser les études d'observation.

Car les données d'observation ont un attrait indéniable, comme le notent Moses [1995] ou Benson et Hartz [2000], notamment un coût moindre, une plus grande rapidité pour obtenir des résultats, et

souvent un plus grand nombre de patients. Mais elles peuvent donner des idées fausses. Avant 2002, sur la base d'études observationnelles, les médecins prescrivaient systématiquement une hormonothérapie substitutive aux femmes ménopausées pour prévenir les infarctus du myocarde. Yitschaky *et al.* [2011] rappellent que des expériences randomisées, menées entre 2002 et 2004, ont permis d'établir que certaines de ces femmes présentaient un taux d'infarctus du myocarde plus élevé que les femmes sous placebo, et que l'hormonothérapie substitutive (à base d'œstrogène uniquement) n'entraînait aucune réduction de l'incidence des maladies coronariennes. Comme l'expliquent Sibbald et Roland [1998], seule une expérience randomisée a permis de corriger une pratique, pourtant admise par la profession. On peut aussi penser à l'étude du MRC Vitamin Study Research Group [1991], qui est revenue sur une expérience (non randomisée) visant à établir que l'apport de vitamines pendant la grossesse pourrait prévenir les anomalies du tube neural chez les enfants. Le comité d'éthique, dans les années 1980, n'avait pas voulu priver les patientes de ce traitement potentiellement utile, ce qui a eu des effets secondaires nocifs (non attendus), et a rendu difficile l'analyse des résultats – il a fallu plus de dix ans pour montrer que l'acide folique était la partie efficace du cocktail multivitaminé donné aux femmes enceintes, comme le rappellent Sibbald et Roland [1998].

En pratique, les études d'observation sont utilisées principalement pour identifier les facteurs de risque ou lorsque des essais contrôlés randomisés seraient impossibles, ou contraires à l'éthique. Ainsi, Smith et Pell [2003] posaient – avec humour – la question de l'utilisation d'essais randomisés pour tester l'efficacité des parachutes. Plus sérieusement, Larroque *et al.* [1995] se sont interrogés sur l'impact de la consommation d'alcool pendant la grossesse, en examinant les enfants quelques années plus tard (entre quatre et cinq ans). Ils ont comparé des niveaux de consommation d'alcool modérément faibles et modérément élevés, et ont constaté que les enfants dont les mères consommaient l'équivalent de quatre verres de vin – ou plus – par jour avaient des performances nettement inférieures dans une variété d'évaluations

cognitives. Il a également été noté que les mères qui buvaient plus d'alcool étaient différentes de celles qui en buvaient moins : les plus grosses buveuses étaient moins instruites, plus âgées et fumaient plus souvent des cigarettes. Cette différence implique des biais importants lors de comparaisons. Ils ont proposé des méthodes pour évaluer ces biais, mais une expérience randomisée n'était alors pas envisageable : on ne pouvait pas forcer des mères à boire ou d'autres à arrêter de boire. Comme le notait le Centre pour le contrôle et la prévention des maladies (CDC) ⁽⁴⁾ en 2016 ⁽⁵⁾, même si une inférence causale est impossible sur le sujet : « why take the risk ? » Parfois le principe de précaution s'impose.

De la pratique de la randomisation

Les expériences randomisées sont une méthode qu'il est étonnant de voir rejeter en bloc, comme le fait Raoult [2020a-b]. Elles sont aujourd'hui un outil incontournable en sciences humaines, comme le rappelle Imai [2017]. Mais leur pratique n'est pas sans danger. Comme le montre Stegenga [2018], les expériences randomisées jouent un rôle fondamental non pas « en médecine », mais dans les procédures réglementaires qui encadrent la pratique de la médecine. Pour qu'un médicament (ou disons un traitement) soit approuvé par la Food and Drug Administration (FDA) des États-Unis, il doit généralement y avoir deux essais cliniques randomisés qui suggèrent que le médicament est supérieur à un placebo ; sans aucune contrainte sur le nombre d'essais effectués. Comme les tests négatifs sont souvent non publiés, cette pratique tend à surestimer les bienfaits d'un traitement à cause d'un biais de publication. Plusieurs méta-analyses ont ainsi montré que de nombreuses études randomisées étaient fausses. Mais c'est l'éthique de la pratique, et non pas la méthode, qu'il convient de questionner. Et dans la période de crise sanitaire que l'on connaît actuellement, refuser d'utiliser une telle technique n'est ni sérieux, ni raisonnable.

Notes

1. En ligne sur <https://www.youtube.com/watch?v=7TI3Re57X2Y>

2. Comme le rappelle Headlam [1891], le tirage au sort a longtemps été utilisé dans la Grèce antique pour désigner les représentants. Il est aussi utilisé pour constituer des jurys populaires. Cette méthode permettait à certains d'y voir la possibilité d'une intervention divine.

3. En réalité, certains États ont opté pour cette approche. Plus de 200 000 enfants de deuxième année ont été vaccinés (mais 125 000 enfants environ n'ont pas participé à l'expérience). L'ensemble des enfants de première et de troisième années ont participé à l'expérience, en tant que groupe témoin (non vaccinés). Dans le premier groupe, 17 cas de polio paralysante sur 100 000 ont été observés (comparable aux 16 cas dans le cas randomisé), mais seulement 46 cas dans le groupe supposé témoin. On passe ainsi de $16/57 = 28\%$ à $17/46 = 37\%$ de gain, ce qui est loin d'être comparable.

4. U.S. Centers for Disease Control and Prevention.

5. <https://www.cdc.gov/vitalsigns/pdf/2016-02-vitalsigns.pdf>

Bibliographie

BENSON K. ; HARTZ A.J., "A Comparison of Observational Studies and Randomized, Controlled Trials", *New England Journal of Medicine*, vol. 342, 2000, pp. 1878-1886.

BROWNLEE K.A., "Statistics of the 1954 Polio Vaccine Trials", *Journal of the American Statistical Association*, vol. 50, n° 272, 1955, pp. 1005-1013.

CHARPENTIER A. ; ELIE R. ; REMLINGER C., "Reinforcement Learning in Economics and Finance", 2020. *ArXiv: 2003.10014*

DEHUE T., "Establishing the Experimenting Society: The Historical Origin of Social Experimentation according to

The Randomized Controlled Design”, *American Journal of Psychology*, vol. 114, n° 2, 2001, pp. 283-302.

FERRY-DANINI J., « Petite introduction à l'éthique des essais cliniques », medium.com, 1^{er} avril 2020. <https://medium.com/@ferry.danini/petite-introduction-%C3%A0-l%C3%A9thique-des-essais-cliniques-d1b6d9f0bbb2>

FREEDMAN B., “Equipose and the Ethics of Clinical Research”, *New England Journal of Medicine*, vol. 317, 1987, pp. 141-145.

HEADLAM J.W., *Election by Lot at Athens*, Cornell University Press, 1891.

IMAI K., *Quantitative Social Science: An Introduction*, Princeton University Press, 2017.

MRC Vitamin Study Research Group, “Prevention of Neural Tube Defects”, Results of the Medical Research Council Vitamin Study, *Lancet*, vol. 338, n° 131, 1991, p. 7.

LARROQUE B. ; KAMINSKI M. ; DEHAENE P. ; SUBTIL D. ; DELFOSSE M.-J. ; QUERLEU D., “Moderate Prenatal Alcohol Exposure and Psychomotor Development at Preschool Age”, *American Journal of Public Health*, vol. 85, n° 12, 1995, pp. 1654-1661.

MEIER M., “Polio Trial: An Early Efficient Clinical Trial”, *Statistics in Medicine*, vol. 9, 1990, pp. 13-16.

MELDRUM M., “A Calculated Risk: The Salk Polio Vaccine Field Trials of 1954”, *British Medical Journal*, vol. 317, 1998, pp. 1233-1236.

MOSES L., “Measuring Effects without Randomized Trials? Options, Problems, Challenges”, *Medical Care*, vol. 33, 1995, pp. AS8-AS14.

PATERNOTTE C., « Contre la méthode ? », medium.com,

5 avril 2020. <https://medium.com/@cedpatern/contre-la-m%C3%A9thode-cf4b173c0bc0>

RAOULT D., « Le médecin peut et doit réfléchir comme un médecin, et non pas comme un méthodologiste », *Le Monde*, 2020a, 25 mars 2020. https://www.lemonde.fr/idees/article/2020/03/25/didier-raoult-le-medecin-peut-et-doit-reflechir-comme-un-medecin-et-non-pas-comme-un-methodologiste_6034436_3232.html

RAOULT D., « L'éthique du traitement contre l'éthique de la recherche », le Pr Didier Raoult critique les “dérives” de la méthodologie », *Le quotidien du médecin*, 2020b, 2 avril 2020. <https://www.lequotidiendumedecin.fr/specialites/infectiologie/lethique-du-traitement-contre-lethique-de-la-recherche-le-pr-didier-raoult-critique-les-derives-de>

ROSENBAUM P., *Observation and Experiment: An Introduction to Causal Inference*, Harvard University Press, 2017.

SIBBALD B. ; ROLAND M., “Understanding Controlled Trials: Why are Randomised Controlled Trials Important?”, *British Medical Journal*, vol. 316, 1998, p. 201.

SMITH G. ; PELL J.P., “Parachute Use to Prevent Death and Major Trauma Related to Gravitational Challenge: Systematic Review of Randomised Controlled Trials”, *British Medical Journal*, vol. 327, 2003, p. 1459.

STEGENGA J., *Medical Nihilism*, Oxford University Press, 2018.

YITSCHAKY O. ; YITSCHAKY M. ; ZADIK Y., “Case Report on Trial: Do You, Doctor, Swear to Tell the Truth, the Whole Truth and Nothing but the Truth?”, *Journal of Medical Case Reports*, vol. 5, n° 179, 2011.