

Les approches quasi expérimentales d'une nouvelle génération d'études pour évaluer les effets des politiques de lutte contre la pollution atmosphérique

Quasi-experimental methods as a new generation of studies to evaluate the health effects of interventions to reduce air pollution

Tarik Benmarhnia¹

Résumé

L'évaluation des effets d'une politique publique permet de conclure sur le rôle effectif d'une politique publique dans la réduction de polluants atmosphériques et sur les bénéfices sanitaires y étant possiblement associés. Cet article vise à introduire la notion d'inférence causale et le modèle contrefactuel dans le contexte de l'évaluation des politiques de lutte contre la pollution atmosphérique. Puis nous présentons différentes approches quasi expérimentales qui sont jusqu'à présent très peu utilisées en ce qui concerne les politiques de lutte contre la pollution atmosphérique. Nous détaillons ainsi les approches de variables instrumentales et de régression avec discontinuité avec des exemples, afin d'encourager l'utilisation de ce type d'approche pour qu'une nouvelle génération d'études pour évaluer les effets des politiques de lutte contre la pollution atmosphérique voie le jour.

Mots-clés

inférence causale, pollution de l'air et santé, évaluation de politiques publiques, méthodes quasi expérimentales

Abstract

Impact evaluation aims at providing evidence about the potential effectiveness of a public policy that can be applied in the context of air pollution reduction and related health benefits. This article aims at introducing the notion of causal inference and the counterfactual framework in the context of public health policy for outdoor air quality. We introduce the notion of quasi experiment which is rarely used in this context. We thus highlight the use of instrumental variables and regression discontinuity methods in order to encourage further applications of these methods for a new generation of impact evaluation studies in the context of public health policy for outdoor air quality.

Keywords

causal inference, outdoor air pollution and health, public policy evaluation, quasi-experimental methods

(1) University of California, San Diego (USA)

Aujourd'hui, il est incontestable que les niveaux ambiants de pollution atmosphérique sont bien moins élevés qu'il y a plusieurs décennies, bien que ce constat ne soit pas valable pour certains pays en voie de développement, comme l'Inde par exemple (Ravindra, Sidhu *et al.* 2015). Ces réductions sont très probablement associées aux nombreuses réglementations qui ont été mises en place, comme le *Clean Air Act* aux USA (Fann, Fulcher *et al.*, 2009 ; Dominici, Greenstone *et al.*, 2014) ou celles mises en place en Europe (Crippa, Janssens-Maenhout *et al.*, 2015). Étant donné les relations entre les différents polluants atmosphériques et la santé des populations, une réduction des niveaux de pollution pourra être associée à de nombreux bénéfices sanitaires.

Toutefois, sans une évaluation précise des effets, il est difficile de conclure sur le rôle effectif d'une politique publique dans la réduction de polluants atmosphériques donnés et encore moins sur les bénéfices sanitaires y étant possiblement associés. L'évaluation des effets vise à répondre précisément à ce type de question et se distingue d'autres types d'évaluations, telles que l'évaluation d'implantation (qui vise à comprendre les mécanismes de mise en place d'une intervention), par exemple (Champagne, Contandriopoulos *et al.*, 2009, Gertler, Martinez *et al.*, 2016). Il est fondamental de conduire des évaluations des effets en lien avec une politique publique précise qui vise à réduire les niveaux de pollution atmosphérique pour différentes raisons (Dominici, Greenstone *et al.*, 2014). Cela permet notamment de s'assurer que les ressources publiques sont utilisées de manière efficace et également d'expérimenter de nouvelles interventions et de décider quant au maintien, à la modification ou à l'arrêt d'interventions existantes.

Les études ayant porté sur l'évaluation des effets sanitaires de ce type de politique publique s'appuient pour la plupart sur des approches méthodologiques « classiques » en épidémiologie, dont un certain nombre est décrit dans l'article de Segala et Duchesne. Or ces méthodes se heurtent souvent à la difficulté de contrôle des facteurs de confusion, qu'ils soient mesurés ou non. Des méthodes, empruntées à l'économétrie, sont aujourd'hui en développement en santé publique et s'avèrent prometteuses pour pallier cette difficulté. Cet article propose un éclairage sur les concepts auxquels elles font appel et fournit des conseils et des exemples de mise en œuvre pour évaluer les effets des politiques de lutte contre la pollution atmosphérique.

Inférence causale et modèle contrefactuel

Il est ici question d'attribuer un effet causal à une intervention précise sur un indicateur de santé donné dans un contexte particulier. La notion d'inférence causale est centrale dans de nombreuses disciplines aujourd'hui car elle permet de questionner le caractère causal d'une association entre une intervention et des changements vis-à-vis d'un indicateur de santé particulier, par exemple. L'un des modèles les plus souvent utilisés pour identifier des effets de nature causale est basé sur la notion de contrefactuel. Il s'agit de comparer deux groupes (le groupe factuel et le groupe contrefactuel) qui seront « identiques » à une exception près : le groupe contrefactuel n'aura pas reçu l'intervention d'intérêt. On dit notamment que ces deux groupes doivent être échangeables, dans la mesure où l'intervention, qu'elle survienne dans l'un ou l'autre groupe, aurait les mêmes effets observés.

Ainsi, avec le modèle contrefactuel, il s'agit de se questionner sur l'évolution d'un indicateur de santé donné, pour la même population, dans une situation hypothétique dans laquelle l'intervention d'intérêt n'aurait pas eu lieu. Ce questionnement nécessite donc de changer la manière dont sont formulées les questions d'évaluation (Zigler, Dominici, 2014). En effet, la majorité des études conduites jusqu'à présent (Henschel, Atkinson *et al.*, 2012) visent majoritairement à comparer des indicateurs de santé avant et après la mise en place d'une intervention. Ce type de raisonnement est particulièrement problématique lorsqu'il s'agit d'attribuer des changements d'un indicateur de santé à une intervention particulière afin d'en isoler l'effet, car plusieurs facteurs (mesurés ou non) peuvent contribuer aux changements sans qu'ils soient pour autant intrinsèques à l'intervention d'intérêt. Ainsi, il est fondamental que des évaluations basées sur le modèle contrefactuel soient dorénavant conduites pour qu'une nouvelle génération d'études visant à évaluer les effets des politiques de lutte contre la pollution atmosphérique puisse voir le jour.

En utilisant le modèle contrefactuel, il sera ainsi possible de mesurer l'effet causal d'une intervention *via* le contraste entre le groupe factuel et le groupe contrefactuel. Seulement, le problème fondamental de la causalité (Holland, 1986 ; Kaufman, 2007) est que, par définition, ce groupe contrefactuel n'est pas directement observable. La randomisation a justement été proposée comme solution pour pallier ce problème. En effet, elle permet de résoudre un enjeu capital : la prise en compte des facteurs de confusion, qu'ils soient mesurés ou non (Bonell, Hargreaves *et al.*, 2011). Il est important de distinguer les facteurs de confusion selon qu'ils sont mesurés ou non mesurés. Dans une situation improbable dans laquelle il n'existerait pas de facteurs

de confusion non mesurés, il sera toujours possible de pallier les biais de confusion mesurés par ajustement ou appariement. Cependant, une telle situation dans laquelle il n'y a pas de facteurs de confusion non mesurés, en plus d'être improbable, demeure invérifiable empiriquement.

La randomisation comme variable instrumentale

Les Essais Contrôlés Randomisés (ECR) ont été introduits et utilisés en médecine et sciences sociales, où un groupe est assigné aléatoirement à l'intervention (ou au traitement) et un groupe ne l'est pas. Cette répartition aléatoire permet ainsi de s'assurer qu'il n'y a pas de différence entre les deux groupes avant l'intervention, autrement dit qu'il n'y a pas de biais de confusion, mesurés ou non. Pour illustrer comment la randomisation permet de considérer la présence de facteurs de confusions mesurés ou non, il est nécessaire d'introduire la notion de variable instrumentale. Une variable instrumentale peut être définie simplement comme une variable (qu'on désignera IV) qui sera fortement corrélée à une exposition donnée (disons un niveau de polluant atmosphérique X) sans affecter l'indicateur de santé Y d'intérêt, autrement dit au travers de la variable X. Autrement dit, la variable IV affectera la distribution de Y exclusivement au travers de son influence sur la variable d'exposition X. Ainsi, en considérant cette variable IV, il sera possible de considérer l'ensemble des facteurs de confusion mesurés ou non qui existent entre les variables X et Y, à condition de respecter certaines hypothèses (Jackson, Swanson, 2015) et de mener les analyses statistiques appropriées (Angrist, Imbens 1995 ; Angrist, Krueger, 2001 ; Angrist, Pischke, 2008).

Ceci étant dit, il est possible de représenter la randomisation comme une variable instrumentale idéale, dans la mesure où le fait qu'un groupe soit assigné aléatoirement à un groupe ou l'autre prédira parfaitement la

probabilité de recevoir le traitement (ou l'intervention) en s'assurant que les deux groupes sont échangeables et qu'il n'y a pas de biais liés aux facteurs de confusion mesurés ou non. Le graphe orienté acyclique (*Directed Acyclic Graph* ou DAG en anglais) (Shrier, Platt, 2008) ci-dessous représente cette situation. Dans la figure ci-dessous, IV représente une variable instrumentale, X l'exposition (e.g. monoxyde de Carbone, CO) et Y un indicateur de santé donné, tandis que C représente un ensemble de facteurs de confusion mesurés (e.g. jour de la semaine) et U un ensemble de facteurs de confusion non mesurés. Pour simplifier, on considère ici une adéquation parfaite (i.e. toutes les unités dans le groupe traitement suivent le traitement) et aucune hétérogénéité dans l'effet de Y sur X. Si l'on considère que IV est un processus de randomisation (IV prédit parfaitement la probabilité de recevoir X ou non) alors, on voit bien que l'utilisation de IV permet de contourner les potentiels biais de confusion liés à C et U permettant d'estimer un effet causal de X sur Y.

Ainsi, la randomisation permet, pour les raisons énoncées précédemment, d'estimer un effet causal d'une exposition (ou une intervention) sur un indicateur de santé. C'est la raison pour laquelle on parle souvent de *gold standard*.

La notion d'expérimentation naturelle et les méthodes quasi expérimentales

Cependant, ce type d'expérimentations, où l'on choisit délibérément d'intervenir sur un groupe et non sur un autre (c.à.d. contrôle) n'est pas faisable dans de nombreuses situations, que ce soit pour des raisons de coût ou des raisons évidentes d'éthique (Hawe, Shiell *et al.*, 2004, Moore, Moore, 2011 ; Petticrew, Chalabi *et al.*, 2011). Ainsi, il est difficile de concevoir des situations réelles dans lesquelles il serait envisageable de conduire une randomisation de ce type en ce qui

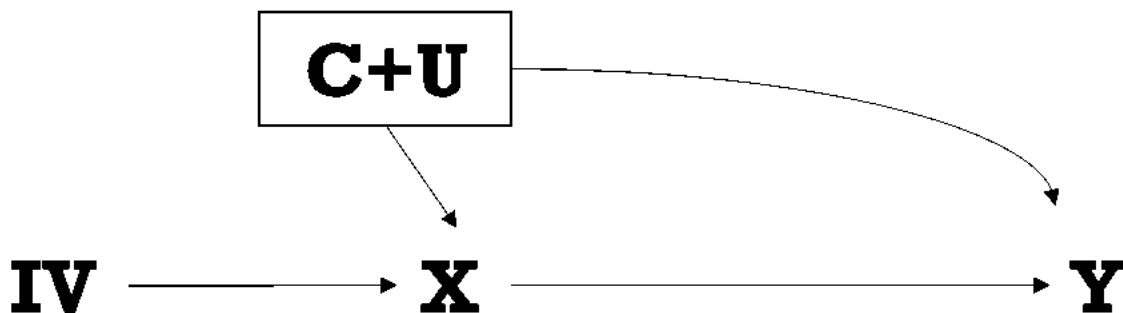


Figure 1. DAG représentant l'utilisation d'une variable instrumentale pour estimer l'effet de X sur Y.
Directed Acyclic Graph representing the use of an instrumental variable to estimate the effect of X on Y.

concerne l'évaluation d'une intervention qui viserait à réduire les niveaux de pollution atmosphérique sur une population d'une ville, par exemple.

Heureusement, il existe des situations dans lesquelles une répartition aléatoire non délibérée a pu avoir lieu, permettant de distinguer deux groupes (*i.e.* factuels et contrefactuels) et de conduire une évaluation des effets d'une intervention. Cela fait référence à la notion « d'expérimentation naturelle ». Les « expérimentations naturelles » constituent un excellent devis d'étude alternatif aux ECR, car elles permettent, en se basant sur des situations observées où il n'y a pas eu de répartition délibérée et aléatoire des groupes d'intervention et de contrôle, de se baser également sur la notion de contrefactuel pour estimer des effets de nature causale (Bor, 2016). Pour cela, les Méthodes Quasi Expérimentales (MQE) ont été développées pour concevoir des groupes contrefactuels en se basant uniquement sur des données observées. L'expression « quasi expérimental » fait ainsi référence à une « *expérimentation avec une intervention donnée, des résultats de santé mesurés, et deux unités expérimentales (groupe recevant l'intervention et groupe contrôle) mais sans recours à la randomisation pour créer les conditions d'inférence causale* » (Cook, Campbell *et al.*, 1979 ; Shadish, Cook, 2009). Ainsi, les MQE permettent d'imiter les résultats que produirait un ECR considérant les facteurs non mesurés, tout en ne servant que de situations existantes.

Variables instrumentales et pollution atmosphérique

Il existe traditionnellement plusieurs types de MQE. La principale approche consiste à utiliser une variable instrumentale issue de la configuration de l'expérimentation naturelle. Des variables géographiques (frontières, altitude, météorologie...) pourront notamment être utilisées comme variables instrumentales (voir détails plus bas). L'idée générale est de trouver une variable qui pourra être utilisée comme variable instrumentale pour mesurer l'effet causal de l'exposition d'intérêt sur un indicateur de santé.

Depuis une dizaine d'années, plusieurs études ont été publiées, principalement dans la littérature en économétrie, pour mesurer l'effet de polluants atmosphériques sur la santé *via* l'utilisation de variables instrumentales. Dans un travail récent (Benmarhnia, Bharadwaj *et al.*, à paraître), ces études ont été identifiées systématiquement et analysées. Le tableau 1 ci-dessous rapporte quelques-unes de ces études pour illustrer quelles variables instrumentales peuvent être utilisées en lien avec la pollution atmosphérique.

Dans les exemples présentés ci-dessus, on voit que plusieurs variables instrumentales peuvent être utilisées dans ce contexte, qui peuvent être résumées selon trois catégories : i) celles en lien avec un choc

Tableau 1. Illustration de différentes utilisations de variables instrumentales pour évaluer l'effet de la pollution atmosphérique sur la santé.

Summary table of studies that used IV to assess the air pollution effects on health.

Étude	Lieu	Indicateur de Santé	Polluant Atmosphérique*	Variable Instrumentale utilisée
Arceo <i>et al.</i> , 2015	Mexique	Mortalité infantile	PM ₁₀ & CO	Inversions météorologiques thermiques
Deryugina <i>et al.</i> , 2016	USA	Mortalité et hospitalisation	PM _{2.5}	Changements dans la direction du vent
He <i>et al.</i> , 2015	Chine	Mortalité totale	PM ₁₀	Organisation des Jeux Olympiques
Jayachandran, 2008	Indonésie	Mortalité infantile	O ₃	Feux de forêt
Luechinger, 2014	Allemagne	Mortalité infantile	SO ₂	Règlementation interdisant l'utilisation du soufre dans des activités industrielles
Moretti, Neidell, 2011	USA	Hospitalisations	O ₃ , NO ₂ et CO	Fluctuations du trafic portuaire de Los Angeles
Schlenker, Walker, 2015	USA	Hospitalisations	CO	Fluctuations dans les congestions à l'aéroport
Schwartz <i>et al.</i> , 2015	USA	Mortalité totale	PM _{2.5}	Inversions météorologiques thermiques

* PM : particule matter ; CO : monoxyde de carbone ; O₃ : Ozone ; NO₂ : dioxyde d'azote

économique ou réglementaire ou une variation géographique quant à l'un de ces chocs ; ii) celles en lien avec des fluctuations externes de trafic ; iii) celles en lien avec des événements environnementaux ou météorologiques externes. Ainsi, il est possible de profiter de ce type de répartition aléatoire non délibérée entre différents territoires pour utiliser une analyse avec variable instrumentale. Bien que la plupart des exemples présentés ci-dessus soient focalisés sur l'effet des polluants eux-mêmes plutôt que sur une politique publique en tant que telle, il est aisément possible d'étendre ces approches en incluant une étape supplémentaire dans l'analyse ou de trouver des variables instrumentales spécifiques à une politique publique.

L'approche de régression avec discontinuité

L'approche de Régression avec Discontinuité (RD) se prête particulièrement bien aux politiques publiques de lutte contre la pollution atmosphérique. L'approche de RD peut être utilisée lorsqu'une intervention se base sur un seuil précis d'une variable continue pour déterminer l'éligibilité quant à cette intervention (Bor, Moscoe *et al.*, 2014 ; Moscoe, Bor *et al.*, 2015). Tous les individus d'un côté du seuil recevront l'intervention, tandis que tous ceux de l'autre côté de ce seuil ne la recevront pas. L'idée générale de cette approche est de considérer les individus très proches d'une part et de l'autre de ce seuil, comme échangeables, considérant le choix du seuil à sa valeur précise comme aléatoire. Pour les détails de sa mise en œuvre analytique, plusieurs articles en anglais ont été publiés (Imbens, Lemieux, 2008 ; Lee, Lemieux, 2010 ; Moscoe, Bor *et al.*, 2015). Nous avons aussi proposé dans un ouvrage méthodologique récent en français (Benmarhnia, Fuller, 2017), les différentes étapes à mener et la syntaxe pour mener à bien les analyses avec les logiciels *R* et *Stata*, qui permettront au lecteur de reproduire ce type d'analyses dans un contexte d'évaluation donné.

De nombreuses politiques publiques de lutte contre la pollution atmosphérique peuvent se retrouver dans cette configuration. Premièrement, il sera possible d'utiliser la distance à une frontière administrative comme seuil pour évaluer l'effet d'une réglementation, par exemple. Considérant des populations très proches d'un côté et de l'autre de cette frontière comme unités échangeables, et que la réglementation n'a lieu que d'un côté de la frontière, une relation entre la distance à la frontière et un indicateur de santé pourra être attribuable à la réglementation. C'est précisément l'approche qui a été proposée par Chen *et al.* (Chen, Ebenstein *et al.*, 2013), qui ont utilisé une rivière (*i.e.* rivière Huai) comme frontière naturelle

pour évaluer l'effet d'une réglementation portant sur les émissions de charbon sur la santé des populations d'un côté et de l'autre de la rivière.

Deuxièmement, il sera possible d'utiliser le choix d'un seuil de déclenchement d'une alerte en cas de smog en milieu urbain. Ce type d'intervention, très répandu à travers le monde, vise à déclencher un ensemble d'actions (*e.g.* réduction de la vitesse de circulation, communication auprès des personnes vulnérables, etc.) lorsque les niveaux de polluants (un ou plusieurs polluants selon le contexte) dépassent un certain seuil qui définira les jours de smog. Ici, il s'agira de considérer le choix de ce seuil précis comme arbitraire, dans la mesure où il ne correspond à aucun effet physiologique connu des différents polluants atmosphériques. Ainsi, en considérant les jours étant juste au-dessus et juste en dessous de ce seuil, on pourra attribuer toute discontinuité suite à la mise en place de ce type d'intervention à son effet, dans un premier temps, sur les niveaux de polluants et, dans un second temps, sur la santé des populations. Cette approche a été utilisée pour évaluer les effets des alertes en cas de smog en Californie du Sud (Neidell, 2009) sur les activités extérieures. Dans un article récent, nous avons développé une approche de ce type pour évaluer l'effet des alertes en cas de smog dans la ville de Toronto sur la mortalité et les hospitalisations (Chen, Li, *et al.*, à paraître).

Conclusion

Dans cet article, nous avons pu présenter comment le modèle contrefactuel peut être utilisé pour inférer les effets sanitaires de politiques de lutte contre la pollution atmosphérique sans avoir recours aux ECR. Nous avons notamment expliqué pourquoi la randomisation pouvait être analysée comme une variable instrumentale, et qu'il est possible d'utiliser de nombreuses autres variables instrumentales dans le cadre d'expérimentations naturelles *via* le recours à des méthodes quasi expérimentales. Dans cet article, nous avons principalement abordé les variables instrumentales et un cas particulier de ces dernières, à savoir l'approche de régression avec discontinuité. Un autre type de méthode quasi expérimentale, appelé méthode de Différence dans les Différences (DD) a aussi été développé *via* le modèle contrefactuel et permet d'évaluer les effets des interventions de santé publique. Cette méthode DD sert d'un ou plusieurs groupes contrôle pour approximer la trajectoire d'un groupe contrefactuel n'ayant pas reçu d'intervention, qui serait interchangeable avec le groupe ayant reçu l'intervention. Plusieurs articles détaillent comment mettre en œuvre cette méthode (Donald, Lang, 2007 ; Dimick, Ryan, 2014 ; Grabich, Robinson *et al.*, 2015 ; Benmarhnia, Bailey

et al., 2016). Nous avons également proposé dans un ouvrage méthodologique récent en français (Benmarhnia, Fuller, 2017) comment la mettre en œuvre, ainsi que la syntaxe pour mener les analyses avec les logiciels *R* et *Stata*. Cette approche a également été récemment appliquée aux politiques de lutte contre la pollution atmosphérique, comme au Chili (Mullins, Bharadwaj, 2015) ou en Chine (Tanaka, 2015). En parallèle, des approches basées sur des analyses de séries temporelles interrompues avec un groupe contrôle ont également été menées comme, par exemple, au Japon pour évaluer l'effet d'une politique de lutte contre les émissions de diesel (Yorifuji *et al.* 2016).

Les méthodes quasi expérimentales sont jusqu'à présent très peu utilisées en ce qui concerne les politiques de lutte contre la pollution atmosphérique en comparaison à d'autres types de politiques publiques, mais leur utilisation est aujourd'hui largement recommandée (Dominici, Greenstone *et al.* 2014). Ces approches permettent de se rapprocher des résultats qui seraient obtenus *via* un ECR, à

condition de respecter certaines hypothèses et de mener les analyses statistiques appropriées. Comme nous l'avons vu, cela permet également de poser les questions d'évaluation en termes causaux dans le cadre d'un modèle contrefactuel bien défini. Ainsi, les approches quasi expérimentales peuvent permettre l'émergence d'une nouvelle génération d'études quant à l'évaluation des effets sanitaires de politiques de lutte contre la pollution atmosphérique et ainsi contribuer à améliorer les connaissances sur l'efficacité de ce type de politique publique. Savoir si une politique publique précise a directement contribué à améliorer la santé des populations est une priorité vis-à-vis des mandats de la santé publique, étant donné le fardeau de la pollution atmosphérique sur la santé des populations, et le fait que nous ne disposons que de très peu de connaissances à ce sujet. Ainsi, le développement des méthodes quasi expérimentales peut contribuer à répondre à cette exigence dans les prochaines années pour mieux comprendre comment lutter contre les effets sanitaires de la pollution atmosphérique.

Références bibliographiques

Angrist J, Krueger AB, 2001: Instrumental variables and the search for identification: From supply and demand to natural experiments, National Bureau of Economic Research.

Angrist JD, Imbens GW, 1995: Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430), 431-442.

Angrist JD, Pischke JS, 2008: Mostly harmless econometrics: An empiricist's companion, Princeton university press.

Benmarhnia T, Bailey Z, Auger N *et al.*, 2016: A difference-in-differences approach to assess the effect of a heat action plan on mortality and equity in Montreal, Quebec. *Environmental Health Perspectives*.

Benmarhnia T, Bharadwaj P, Romero M, à paraître: Using Instrumental Variables under Partial Observability of Endogenous Variables for Assessing the Effects of Air Pollution on Health.

Benmarhnia T, Fuller D, 2017 : Les méthodes quasi expérimentales pour évaluer les interventions de santé publique. Méthodes avancées d'Évaluation d'Interventions de Santé Mondiale, R. V. D. C., Presses de l'Université de Montréal.

Bonell CP, Hargreaves J, Cousens S *et al.*, 2011: Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology and Community Health*, 65(7), 582-587.

Bor J, 2016: Capitalizing on Natural Experiments to Improve Our Understanding of Population Health, *American journal of public health*, 106(8), 1388.

Bor J, Moscoe E, Mutevedzi P *et al.*, 2014: Regression discontinuity designs in epidemiology: causal inference without randomized trials, *Epidemiology*, 25(5), 729-737.

Champagne F, Contandriopoulos AP, Brousselle A *et al.*, 2009 : L'évaluation dans le domaine de la santé : concepts et méthodes, *L'évaluation : concepts et méthodes*, 35-56.

Chen H, Li Q, Wang J *et al.*, à paraître : Do air pollution alerts influence public health ? A regression discontinuity analysis.

Chen Y, Ebenstein A, Greenstone M, Li H, 2013: Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy, *Proceedings of the National Academy of Sciences*, 110(32), 12936-12941.

Cook TD, Campbell DT, Day A, 1979: Quasi-experimentation: Design & analysis issues for field settings, Houghton Mifflin Boston.

Crippa M, Janssens-Maenhout G, Dentener F, *et al.*, 2015: Forty years of improvements in European air quality: the role of EU policy-industry interplay, *Atmospheric Chemistry & Physics Discussions*, 15(15).

Dimick JB, Ryan AM, 2014: Methods for evaluating changes in health care policy: the difference-in-differences approach, *Jama*, 312(22), 2401-2402.

Dominici F, Greenstone M, Sunstein CR, 2014: Particulate matter matters, *Science*, 344(6181), 257-259.

Donald SG, Lang K, 2007: Inference with difference-in-differences and other panel data, *The review of Economics and Statistics*, 89(2), 221-233.

Fann N, Fulcher CM, Hubbell BJ, 2009: The influence of location, source, and emission type in estimates of the human health benefits of reducing a ton of air pollution, *Air Quality, Atmosphere & Health*, 2(3), 169-176.

Gertler P, Martinez S, Premand P *et al.*, 2016: Impact Evaluation in Practice, World Bank Group.

Grabich SC, Robinson WR, Engel SM *et al.*, 2015: County-level hurricane exposure and birth rates: application of difference-in-differences analysis for confounding control. *Emerging themes in epidemiology*, 12(1), 1.

Hawe P, Shiell A, Riley T, 2004: Complex interventions: how" out of control. *Penelope*, 328, 1561-1563.

Henschel S, Atkinson R, Zeka A *et al.*, 2012: Air pollution interventions and their impact on public health, *International journal of public health*, 57(5), 757-768.

Holland, P. W., 1986. Statistics and causal inference, *Journal of the American statistical Association*, 81(396), 945-960.

Imbens GW, Lemieux T, 2008: Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.

Jackson JW, Swanson SA, 2015: Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology (Cambridge, Mass.)*, 26(4), 498.

Kaufman JS, 2007: Making causal inferences about macrosocial factors as a basis for public health policies. *Macrosocial Determinants of Population Health*, Springer, 355-373.

Lee DS, Lemieux T, 2010: Regression discontinuity designs in economics, *Journal of economic literature*, 48(2), 281-355.

Moore L, Moore GF, 2011: Public health evaluation: which designs work, for whom and under what circumstances?, *Journal of epidemiology and community health*, 65(7), 596-597.

Moscoe E, Bor J, Bärnighausen T, 2015: Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice, *Journal of clinical epidemiology*, 68(2), 132-143.

Mullins J, Bharadwaj P, 2015: Effects of short-term measures to curb air pollution: Evidence from Santiago, Chile, *American Journal of Agricultural Economics*, 97(4), 1107-1134.

Neidell M, 2009: Air quality warnings and outdoor activities: evidence from Southern California using a regression discontinuity design, *Journal of epidemiology and community health*, jech. 2008.081489.

Petticrew M, Chalabi Z, Jones DR, 2011: To RCT or not to RCT: deciding when 'more evidence is needed' for public health policy and practice, *Journal of epidemiology and community health*, jech. 2010.116483.

Ravindra K, Sidhu MK, Mor S *et al.*, 2015: Air Pollution in India: Bridging the Gap between Science and Policy, *Journal of Hazardous, Toxic, and Radioactive Waste*, A4015003.

Shadish WR, Cook TD, 2009: The renaissance of field experimentation in evaluating interventions, *Annual review of psychology*, 60, 607-629.

Shrier I, Platt RW, 2008: Reducing bias through directed acyclic graphs, *BMC medical research methodology*, 8(1), 70.

Tanaka S, (2015): Environmental regulations on air pollution in China and their impact on infant mortality, *Journal of health economics*, 42, 90-103.

Yorifuji T, Kashima S, Doi H, 2016: Fine-particulate Air Pollution from Diesel Emission Control and Mortality Rates in Tokyo: A Quasi-experimental Study, *Epidemiology*, 27(6), 769-778.

Zigler CM, Dominici F, 2014: Point: clarifying policy evidence with potential-outcomes thinking—beyond exposure-response estimation in air pollution epidemiology, *American Journal of Epidemiology*, kwu263.