Comment attribuer une performance dans un pipeline socio-technique complexe ?

Qui attend quoi, et de quoi on parle

Une partie du discours public a fait de 2026 l'année du verdict. ***Une quinzaine de composés issus de plateformes d'intelligence artificielle entrent en Phase III pivotale, et l'on annonce que ces lectures diront enfin si l'IA sait, ou non, découvrir des médicaments***. La thèse de ce texte est que la question est doublement mal posée.

- Elle l'est sur le fond, parce qu'un essai pivotal mesure une molécule et non un procédé.
- Elle l'est sur la forme, parce qu'elle cherche un auteur (« l'IA a-t-elle découvert ? ») là où il faudrait chercher une répartition (« quelle part de la performance revient à quel composant ? »).

La découverte d'un médicament n'est pas un acte ; c'est une chaîne. On n'attribue pas une chaîne à l'un de ses maillons.

Par *capacité*, on entendra dans ce texte une chose précise et unique : la propriété reproductible d'un procédé produisant un avantage mesurable dans un domaine d'usage déclaré. Trois exigences, donc :

- Reproductible (pas un coup),
- Mesurable (pas un récit),
- Et déclarée quant au domaine (pas générale par défaut).

Cette définition tient tout l'article ; chaque fois que le mot reparaît, il faut y entendre ces trois conditions, jamais davantage.

Il faut aussi nommer qui porte l'attente, car ce n'est pas tout le monde. Les directions de R&D savent qu'un essai pivotal n'arbitre pas une méthode. Les investisseurs aguerris raisonnent en valeur actuelle nette ajustée du risque, pas en récit de validation.

Le « moment de vérité » est une production de la presse spécialisée, de quelques influenceurs et d'une fraction des fonds, qui comptent les programmes en clinique (de l'ordre de 173, dont une quinzaine en Phase III, selon des recensements sectoriels qu'il faut prendre pour ce qu'ils sont : des décomptes d'acteurs intéressés à la thèse haute) et lisent dans ce volume une promesse en passe d'être tenue. C'est à ce récit que ce texte s'adresse, avant de proposer ce par quoi le remplacer.

Trois niveaux d'indécidabilité, et celui qu'on défend

Quand on dit qu'on ne peut pas attribuer la performance à l'IA, on peut vouloir dire trois choses :

1. Niveau 1 : on ne dispose pas, aujourd'hui, des données,
2. Niveau 2 : ces données seraient obtenables, mais à un coût prohibitif,
3. Niveau 3 : l'attribution serait théoriquement impossible, quelles que soient les données.

Ce texte se tient aux niveaux 1 et 2. Il ne soutient pas que la contribution de l'IA soit inattribuable par essence : il soutient que les dispositifs actuels ne permettent pas de l'isoler, et il proposera plus loin ceux qui le permettraient. Une impossibilité de principe serait elle-même non falsifiable, donc indéfendable. Partout où ce texte écrit « ne permet pas », il faut entendre « pas avec ce qu'on observe en 2026 », et non pas « jamais ».

Le survivant, et l'intrication du signal

Le pipeline d'un médicament est une suite de filtres.

- Une plateforme propose des candidats,
- Des chimistes les retiennent ou les écartent,
- La chimie médicinale les optimise,
- L'ADME en élimine une part, la toxicologie une autre,
- La Phase I tranche sur la tolérance,
- La Phase II sur le premier signal d'efficacité…

Le taux de base le rappelle sans douceur : sur l'ensemble des indications, la probabilité de passer de Phase II à Phase III est de l'ordre de 30 %, et tombe sous 20 % en neurologie (Wong, Siah et Lo, *Biostatistics*, 2019) => ***Une molécule en Phase III n'est pas un échantillon : c'est un rescapé.***

On serait tenté de dire que le signal de la plateforme se dilue à chaque filtre. Ce serait faux, parce que les pipelines modernes ne sont pas séquentiels mais itératifs : la sortie de la plateforme informe une décision humaine, qui appelle une nouvelle requête, qui réoriente l'optimisation.

La bonne notion n'est donc pas la dilution, mais l'*intrication*, qu'il faut définir et non suggérer. L'intrication, ici, désigne la non-séparabilité : la contribution marginale d'un composant dépend des valeurs réalisées des autres, parce que sorties algorithmiques et décisions humaines se conditionnent mutuellement par itération. Deux quantités intriquées ne se soustraient pas. C'est cette propriété, pas l'évaporation d'un signal, qui fait obstacle à l'attribution naïve.

Elle en commande une autre, plus technique. Le récit critique invoque volontiers le biais du survivant comme facteur de confusion ***: la population de Phase III serait biaisée par la survie, donc non comparable***. Mais si la valeur propre d'une plateforme est précisément de mieux éliminer les mauvais candidats en amont, alors la surreprésentation de survivants n'est pas un artefact à corriger : c'est une *variable médiatrice*, un maillon sur le chemin causal de l'effet recherché. Conditionner sur elle bloquerait une partie de ce qu'on veut mesurer. Le problème n'est donc pas que la survie biaise la comparaison, mais qu'on ignore, en l'état, si elle est confondeur ou médiateur. Cette ambiguïté est l'obstacle d'identification, formulé proprement.

« Découvert par IA » ne nomme rien : une taxonomie à trois étages

Le récit suppose une catégorie : ***il existerait des « médicaments découverts par IA », qu'on pourrait dénombrer et dont on mesurerait le taux de succès***. Cette catégorie n'existe pas au sens opératoire. « Découvert par IA » est un terme de communication, et il agrège des contributions qui n'ont ni le même point d'application ni le même poids causal.

Ces contributions se rangent sur trois étages, et les confondre est la première erreur.

- Au stade de la *découverte* : identification de cible, validation de cible,
- Au stade de la *conception* : génération de novo de molécules, optimisation de tête de série,
- Au stade du *développement* : prédiction ADMET, découverte de biomarqueurs, stratification de patients, optimisation de protocole d'essai.

Le repositionnement de composés connus traverse les trois. Ces étages ne sont pas commensurables : une plateforme qui identifie une cible inédite et une plateforme qui réordonne une bibliothèque de têtes de série ne font pas le même métier, et leur réussite ne se compte pas dans la même monnaie.

S'y ajoute une dimension que le récit ignore : une plateforme excellente sur les kinases peut être médiocre ailleurs.

La question pertinente n'est donc pas « l'IA a-t-elle une capacité générale de découverte », mais « a-t-elle une capacité dans un domaine d'usage déclaré ». C'est la logique de substituabilité qu'on applique aux jumeaux numériques : un système ne vaut que dans son périmètre de validité énoncé, et l'extrapoler hors de ce périmètre est un biais, pas une généralisation. Une capacité sans domaine déclaré n'est pas une capacité : c'est une moyenne.

Dans quel cadre causal, et avec quelles précautions

Affirmer que l'IA « ne cause pas directement » une meilleure molécule n'a de sens que dans un cadre causal précisé, et il faut résister à la tentation de les invoquer comme des synonymes. Le cadre des effets potentiels, hérité de Rubin, définit la contribution d'un composant comme un contraste contrefactuel : la grandeur d'intérêt avec le composant, moins cette même grandeur sans lui, sur une unité comparable.

Le cadre des graphes orientés, hérité de Pearl, sert à savoir si ce contraste est seulement identifiable à partir des données observées, en distinguant confondeurs, médiateurs et variables de collision. Ces deux cadres ne répondent pas à la même question : l'un définit l'estimand, l'autre statue sur son identifiabilité. Les articuler dans un pipeline industriel réel est un sujet de recherche ouvert, pas une formalité ; ce texte les mobilise comme deux instruments complémentaires, sans prétendre que leur réconciliation soit acquise.

Dans ce cadre, une chose au moins est nette : la contribution causale d'un composant n'est pas son rang dans la chaîne, mais sa contribution marginale. L'IA en amont n'est pas l'IA responsable. La question n'est jamais « qui a agi en premier », mais « qu'est-ce qui aurait été différent sans ce composant ».

L'anomalie de Phase I, sans la trancher, et la question du temps

Le récit dispose d'un chiffre qui impressionne : les composés étiquetés IA franchiraient la Phase I à 80-90 %, contre environ 52 % en historique traditionnel. Ces ordres de grandeur proviennent des mêmes recensements sectoriels, à consolider, et l'argument n'en dépend pas. La Phase I teste la tolérance, pas l'efficacité. Un tel écart admet deux explications également compatibles avec les données : meilleure conception, ou sélection de candidats proches d'un espace chimique déjà connu comme sûr. Les deux produisent le même chiffre. Tant qu'on ne les sépare pas, l'avantage reste inattribuable, au sens de niveau 1.

Dans cette perspective, il est utile de distinguer les plateformes qui conçoivent des candidats de celles qui améliorent le fonctionnement du pipeline sans modifier les propriétés intrinsèques des molécules. Considérons, à titre d'illustration, une plateforme de filtrage toxicologique en amont, telle que ToxTwin V4. Son objectif n'est pas de produire de meilleures molécules, mais de réduire la probabilité qu'un candidat présentant un profil toxicologique défavorable poursuive son développement. Autrement dit, elle ne transforme pas une molécule médiocre en bonne molécule ; elle modifie la distribution des molécules effectivement engagées dans le pipeline expérimental.

Un tel système peut donc augmenter la probabilité de franchissement de la Phase I sans que les molécules retenues possèdent intrinsèquement de meilleures propriétés pharmacologiques. Le gain provient d'une amélioration de la fonction de sélection, non d'une amélioration de la fonction de conception. Les ressources expérimentales sont concentrées sur des candidats dont le profil toxicologique estimé est plus favorable, ce qui réduit les coûts d'évaluation, diminue l'attrition précoce et améliore la performance économique du pipeline, sans que l'algorithme ait, à lui seul, "découvert" une meilleure molécule.

Cette distinction est importante, car elle montre qu'une amélioration du taux de succès en Phase I ne permet pas, à elle seule, d'inférer une meilleure capacité de conception moléculaire. Deux mécanismes causalement distincts peuvent produire le même résultat observable : une plateforme peut générer de meilleurs candidats, ou simplement sélectionner plus efficacement ceux qui méritent d'être évalués. Un readout clinique ne permet pas, en lui-même, de départager ces deux hypothèses.

L'argument devient encore plus fort dans une architecture agentique. Un générateur de molécules, un système de filtrage toxicologique, des modules d'optimisation ADMET et un orchestrateur de décision forment alors un pipeline composé, dans lequel chaque composant modifie la distribution des candidats transmise au suivant. La performance finale n'est plus la propriété d'un algorithme isolé, mais celle de leur composition. Chercher à attribuer le succès d'un candidat clinique à un seul composant revient alors à confondre la performance d'un système avec celle de l'un de ses éléments. Ce qui augmente n'est pas nécessairement la qualité intrinsèque des molécules produites, mais la capacité collective du pipeline à transformer des ressources expérimentales en candidats viables.

Une difficulté redouble la première : l'objet n'est pas stationnaire. Une plateforme apprend, ses équipes apprennent, ses jeux de données et ses modèles évoluent, ses protocoles aussi. La capacité de 2024 n'est pas celle de 2026, qui ne sera pas celle de 2032. Demander « l'IA sait-elle découvrir des médicaments » sans dater la question revient à mesurer un mobile avec une règle fixe. La grandeur n'a pas seulement une valeur incertaine : elle a une dérivée.

La symétrie que le procès oublie

Le récit de validation est un procès mal instruit, parce qu'il croit que le verdict peut tomber dans les deux sens : un succès vaudrait acquittement de la méthode, un échec, condamnation. Les deux inférences butent sur le même fait.

L'attrition de Phase III n'a pas attendu l'IA. Depuis trente ans, des molécules conçues par les meilleurs chimistes humains franchissent la Phase I et la Phase II avant d'échouer en Phase III, le plus souvent sur l'efficacité.

Recursion a discontinué REC-994 en mai 2025 après non-confirmation des tendances observées plus tôt ; Insilico a publié une Phase IIa positive pour le rentosertib, un inhibiteur de TNIK, dans l'essai GENESIS-IPF (71 patients, 22 sites en Chine) : +98,4 mL de capacité vitale forcée à 60 mg par jour contre une baisse de 20,3 mL sous placebo (*Nature Medicine*, 2025), avant de poursuivre vers une Phase III pivotale.

Les deux trajectoires coexistent sous la même étiquette. Si l'échec d'une molécule humaine n'a jamais réfuté la chimie médicinale, l'échec d'une molécule IA ne réfute pas la méthode ; et un succès isolé ne la valide pas davantage. Un essai porte sur un cas ; en faire le juge d'une classe, c'est confondre l'issue et la propriété.

Disons-le sans excès de rigueur poppérienne : la science ne procède pas par preuves absolues, mais par faisceaux d'indices, convergence et vraisemblance. Le reproche fait au readout de 2026 n'est pas de manquer d'une preuve définitive, que rien n'exige. Il est plus précis : dans l'état actuel du dispositif, l'indice qu'il fournit ne discrimine pas les hypothèses qu'on prétend lui faire départager.

Capacité n'est pas P(succès) : l'eNPV au centre

Le débat réduit la capacité d'une plateforme à une probabilité de succès clinique. C'est l'erreur économique centrale, et la corriger suppose le bon vocabulaire. La grandeur que maximise une direction pharmaceutique n'est pas P(succès), mais la valeur actuelle nette espérée, ajustée du risque (eNPV) : l'espérance des flux futurs d'un programme, pondérée par les probabilités de transition et actualisée pour le délai. Trois leviers y entrent, pas un. La probabilité de succès, certes. Mais aussi le coût, qui grève la valeur directement. Et le délai, qui la grève par l'actualisation : dix-huit mois gagnés sur la recherche déplacent les flux vers le présent et augmentent l'eNPV, à probabilité inchangée.

La conséquence est qu'une plateforme peut transformer l'économie d'un portefeuille sans toucher à la biologie. Découvrir 20 % plus vite et 50 % moins cher, fût-ce au prix de quelques points d'échec supplémentaires, peut relever l'eNPV agrégé. Les investisseurs sérieux ne cherchent d'ailleurs pas une meilleure IA : ils cherchent un meilleur portefeuille, et un meilleur portefeuille se construit autant par la structure de coût et la vitesse d'itération que par le taux terminal. Réduire la capacité à la découverte de molécules efficaces, c'est ignorer que l'essentiel de la valeur d'une plateforme peut résider dans ce qu'elle fait à l'eNPV, pas dans le readout d'efficacité.

Ce qu'il faudrait mesurer, et avec quelle unité

La question correcte est causale : quel est l'effet du procédé sur l'eNPV, et non « cette molécule fonctionne-t-elle ». Encore faut-il une unité. Sans grandeur à attribuer, une théorie de l'attribution reste philosophie. On en retient une, principale : la contribution

marginale d'un composant à l'eNPV, elle-même décomposable en trois effets interprétables, sur la probabilité de transition, sur le coût engagé, sur le délai.

En variante, lorsqu'on raisonne en amont du chiffrage, une unité informationnelle convient : la réduction d'incertitude qu'un composant apporte sur la distribution des candidats viables, mesurable comme baisse d'entropie. Le choix de l'unité n'est pas cosmétique : il fixe ce que « contribuer » veut dire, et conditionne tout estimateur en aval.

Le dispositif idéal serait alors un bras apparié : même cible, même indication, même fenêtre, un pipeline assisté par IA contre un pipeline classique, comparés sur la diversité chimique, la distance aux squelettes connus, les taux de passage, l'attrition, le délai, le coût. C'est une ablation : on retire le composant et on observe la variation de l'unité retenue. Il faut admettre aussitôt qu'il est inexécutable : nul ne financera deux programmes concurrents à plusieurs centaines de millions pour le seul plaisir de l'inférence. La voie réaliste passe par les quasi-expériences, familières à l'économétrie : appariement historique sur cibles et indications comparables, score de propension, variables instrumentales, contrôles synthétiques, émulation d'essais cibles, inférence causale bayésienne. Aucune ne vaut un essai randomisé ; leur convergence formerait le faisceau d'indices que le readout isolé ne fournit pas.

On objectera que le brevet offre déjà une trace causale objectivable : un scaffold inédit, sa distance à l'art antérieur, son caractère inventif sont documentés et opposables. C'est exact, et utile. Mais le brevet atteste la nouveauté d'une structure, pas la contribution d'un composant à la performance. Une molécule peut être brevetable et son avantage thérapeutique nul ; l'IA peut avoir produit la nouveauté sans produire l'efficacité. La trace existe, mais elle répond à une autre question.

Un cas d'application, en clair

Soit deux programmes appariés sur une même cible et une même indication, l'un mené sans plateforme (pipeline A), l'autre avec (pipeline B). Le pipeline B atteint la Phase III et la franchit. Que peut-on conclure ? Que la molécule B fonctionne : sa performance est établie.

Rien de plus. On ne peut pas conclure que la plateforme en est la cause, parce qu'on n'observe pas le contrefactuel (ce qu'aurait donné B sans la plateforme), parce que la sélection humaine a pu, à chaque étape, corriger ou même produire l'avantage que l'on prête à l'IA, et parce que A et B ne sont presque jamais appariés dans la réalité.

Comment l'attribution opérerait-elle ? Par ablation, on retirerait la plateforme à un point précis du pipeline B, par exemple l'optimisation de tête de série, en rejouant cette étape par la voie classique, et l'on mesurerait l'écart d'eNPV imputable à ce seul retrait. Par contribution marginale de type Shapley, on irait plus loin : on calculerait l'apport de chaque composant en moyennant son effet sur tous les ordres d'entrée possibles dans

la coalition, de façon que la somme des contributions reconstitue l'eNPV total. Le terrain ToxTwin opère déjà cette discipline à petite échelle : y prédire une toxicité revient à estimer sous priors dans un espace contraint, et la valeur du système s'y juge par ablation de ses composants, jamais par un cas favorable. La même logique vaut, en plus grand, pour un pipeline de découverte.

Pourquoi Shapley, et à quelles conditions

Citer Shapley ne suffit pas ; il faut dire pourquoi lui, et où il achoppe. L'attrait de la valeur de Shapley tient à deux propriétés qu'aucune intuition ne remplace : l'efficience (la somme des contributions égale exactement la performance totale, sans reste inexpliqué) et la symétrie (deux composants interchangeables reçoivent la même part). Ce sont précisément les garanties qu'on veut quand les contributions sont intriquées et qu'aucune décomposition naïve ne tient. D'autres outils existent et ne visent pas la même chose : les indices de Sobol et l'ANOVA fonctionnelle décomposent la *variance* d'une sortie, donc répondent à une question de sensibilité, non de répartition d'une valeur scalaire ; les gradients intégrés attribuent une prédiction de modèle à ses entrées, à l'échelle d'un modèle, non d'un pipeline socio-technique. Shapley est le bon candidat quand la question est « comment partager équitablement une performance totale entre des composants en interaction ».

Reste l'objection, qui est sérieuse : Shapley exige une coalition bien définie, une fonction de valeur, et un coût combinatoire qui explose avec le nombre de composants. Dans un pipeline pharmaceutique réel, aucune de ces conditions n'est donnée d'avance ; les définir est précisément le travail. La construction effective de cette fonction de valeur, le découpage opérationnel des composants et la calibration du protocole sur des programmes réels relèvent d'un savoir-faire d'implémentation que ce texte ne détaille pas. C'est pourquoi il parle d'esquisse, et non de méthode : la valeur de Shapley indique la forme que prendrait une attribution rigoureuse, elle ne dispense pas du chantier. L'outil est juste ; son application est un travail en propre.

Performance, provenance, capacité : la généralisation

Le débat se tient parce qu'il confond trois questions. La performance demande : cette molécule fonctionne-t-elle. La provenance demande : comment est-elle apparue, et quelle part revient à l'IA. La capacité, au sens fixé en ouverture, demande : l'avantage est-il une propriété reproductible du procédé, dans un domaine déclaré, à un coût connu. Une approbation répond à la première et reste muette sur la troisième. Cette trilogie n'est pas une nouveauté ex nihilo : elle prolonge, dans le corpus de l'Institut, l'analyse de la provenance de la performance déjà conduite ailleurs, dont elle est l'extension au registre de la capacité reproductible.

On objectera une dernière fois que la pharma se juge à ses approbations, et que la loi des grands nombres dispense d'attribuer cas par cas : si les composés IA s'approuvent en masse plus souvent et moins cher, cet écart agrégé *est* la capacité. L'objection serait recevable si la population étiquetée IA était comparable à la population de référence. Elle ne l'est pas : l'étiquette agrège trois étages disjoints, la survie filtre avant le comptage et joue tantôt comme confondeur tantôt comme médiateur, et nul appariement de cible ou de période ne corrige ces biais. La statistique de masse ne sauve l'inférence que sur des populations homogènes.

Cette question n'est pas neuve, et il serait malhonnête de la présenter comme une intuition isolée : l'analyse de médiation causale, l'IA explicable au niveau du système plutôt que du modèle, et la littérature d'accountability algorithmique posent déjà, ailleurs, la répartition d'un résultat entre composants humains et techniques. Le présent texte ne l'invente pas ; il l'importe dans la découverte de médicaments, où le récit médiatique l'avait recouverte d'une fausse simplicité.

Il faut alors énoncer ses propres conditions de réfutation, faute de quoi la critique serait, elle aussi, infalsifiable. Voici la mienne. Si, sur plusieurs centaines de cas, des quasi-expériences indépendantes convergeaient pour montrer des réductions robustes de délai et de coût, une diversité chimique supérieure et de meilleurs taux d'approbation à cible et indication appariées, et si ces effets résistaient à plusieurs spécifications de confondeurs, alors l'hypothèse de capacité deviendrait raisonnablement crédible, même sans bras randomisé. La convergence de quasi-expériences tiendrait lieu de la randomisation qu'on ne peut pas s'offrir. Tant que ce faisceau n'existe pas, la capacité reste à estimer, pas à proclamer.

La portée dépasse le médicament, mais il faut en borner le domaine. Dans tout système où la production résulte d'une chaîne d'agents hétérogènes en interaction, humains et algorithmiques, la question pertinente n'est pas de savoir qui a produit un résultat, mais comment répartir quantitativement la contribution marginale de chaque composant à la performance finale. La chimie est un cas. La génération de protéines, la robotique de laboratoire, la conception de matériaux, l'optimisation industrielle satisfont cette condition et posent la même question sous d'autres habits. Là où elle est réunie, attribuer un résultat collectif à un seul agent est une erreur de catégorie, et le répartir est un programme de mesure.

Reste le survivant, par où nous avons commencé. Il faut lui rendre sa juste fonction. Un composé qui atteint la Phase III est un excellent indicateur de performance clinique : il a survécu à tout ce qui tue les molécules, et cela se respecte. Mais il est un mauvais estimateur de provenance : sa survie même a effacé la trace de ce qui l'a produit. Une approbation démontre qu'un médicament fonctionne. Elle ne démontre pas que le procédé qui l'a produit constitue une nouvelle capacité industrielle. 2026 ne dira pas si

l'IA sait découvrir des médicaments. Il dira si tel survivant tient, et il offrira, à qui veut bien la construire, la première occasion d'attribuer le reste.

Limites de cet argument

Six réserves, pour ne pas reconduire l'erreur qu'on dénonce :

1. Les ordres de grandeur sur la volumétrie du pipeline et le taux de Phase I proviennent de recensements sectoriels non consolidés, produits par des acteurs intéressés à la thèse haute ; ils sont mobilisés comme indices, pas comme preuves, et l'argument n'en dépend pas.
2. L'indécidabilité défendue est de niveau 1-2, conditionnelle aux dispositifs : un faisceau de quasi-expériences convergentes la lèverait, selon les conditions de réfutation énoncées plus haut.
3. Le cadre d'attribution est une esquisse, non une méthode validée : la fonction de valeur, le découpage en composants et le coût combinatoire de Shapley restent à construire par programme.
4. L'articulation Rubin-Pearl est mobilisée comme complémentarité, alors qu'elle constitue un problème de recherche à part entière.
5. La distinction confondeur / médiateur appliquée à la survie suppose un graphe causal qui reste à spécifier cas par cas.
6. Enfin, ce texte ne traite pas la valorisation de ces actifs (licences, portefeuille, revendications de brevet au sens stratégique) : ces questions relèvent d'un conseil spécialisé et débordent le périmètre doctrinal assumé ici.

Références

1. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;20(2):273-286. doi:10.1093/biostatistics/kxx069.
2. Xu Z, Ren F, Wang P, Cao J, Tan C, et al. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial. Nat Med. 2025;31(8):2602-2610. doi:10.1038/s41591-025-03743-2.
3. Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, editors. Contributions to the Theory of Games II. Princeton (NJ): Princeton University Press; 1953. p. 307-317.
4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66(5):688-701.
5. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge: Cambridge University Press; 2009.