

Études pharmaco-épidémiologiques : quels niveaux de preuve et comment les atteindre ?

Maryse Lapeyre-Mestre¹, Claudine Sapède², Nicholas Moore³ et les participants à la table ronde N° 5 de Giens XXVIII : Pascal Bilbault⁴, Patrick Blin⁵, Damien Chopy⁶, David Evans⁷, François Gueyffier⁸, Laure Lacoïn⁹, Muriel Malbezin¹⁰, Joëlle Micallef¹¹, Dana Morlet-Vigier¹², Sophie Muller¹³, Emmanuel Oger¹⁴, Yannick Plétan¹⁵, Gérard Pons¹⁶, Patrice Verpillat¹⁷ et Caroline Vigneau¹⁸ †

- 1 Service de Pharmacologie médicale et clinique, Université de Toulouse, INSERM 1027, Toulouse, France
- 2 Affaires économiques, Laboratoire Roche, Boulogne, France
- 3 Service hospitalo-universitaire de Pharmacologie, Université de Bordeaux, INSERM U657, Bordeaux, France
- 4 Laboratoires Boehringer Ingelheim, Reims, France
- 5 Université de Bordeaux, Bordeaux, France
- 6 Inserm Transfert, Paris, France
- 7 Laboratoires Lilly, Suresnes, France
- 8 Université Lyon 1, Hospices civils de Lyon, Lyon, France
- 9 Haute Autorité de Santé, Saint-Denis la Plaine, France
- 10 Laboratoires Janssen, Issy-les-Moulineaux, France
- 11 CIC-UPCET, Hôpital la Timone, Marseille, France
- 12 IMS Health, Puteaux, France
- 13 Laboratoires GlaxoSmithKline, Marly-le-Roi, France
- 14 Université de Rennes 1, Rennes, France
- 15 Laboratoires Roche, Boulogne-Billancourt, France
- 16 Université Paris Descartes, Assistance Publique – Hôpitaux de Paris, Paris, France
- 17 Laboratoires Lundbeck, Issy-les-Moulineaux, France
- 18 Centre Hospitalier Universitaire, Nantes, France

Texte reçu le 15 mars 2013 ; accepté le 7 mai 2013

Mots clés :
évaluation ;
effectiveness ;
risque ;
population rejointe ;
niveau de preuve

Résumé – En matière d'études pharmacoépidémiologiques, de la nature de la question posée dépendront l'approche méthodologique pour y répondre et les conditions d'optimisation du niveau de preuve. Une approche descriptive (études transversales ou longitudinales menées sur bases de données ou études *ad hoc*) sera privilégiée pour documenter la population rejointe et les modalités d'utilisation d'un nouveau médicament. L'évaluation de la performance en vie réelle *effectiveness* pourra se fonder sur des méthodes de type cohortes, cas-témoin ou sur la modélisation. En matière de sécurité médicamenteuse, la nature de l'effet indésirable orientera le choix de la méthode à la fois pour l'identification (signal) et l'évaluation de la causalité. Dans tous les cas, l'évidence acquise en pré-AMM servira de base pour la génération d'hypothèses. Quelle que soit la méthode retenue, il conviendra d'identifier les biais potentiels et d'analyser leur impact sur le résultat obtenu. La combinaison de plusieurs approches complémentaires peut parfois se révéler préférable à une seule étude.

Abréviations : voir en fin d'article.

† Les articles, analyses et propositions issus des Ateliers de Giens n'engagent que leurs auteurs et ne préjugent pas de la position de leur organisme de tutelle

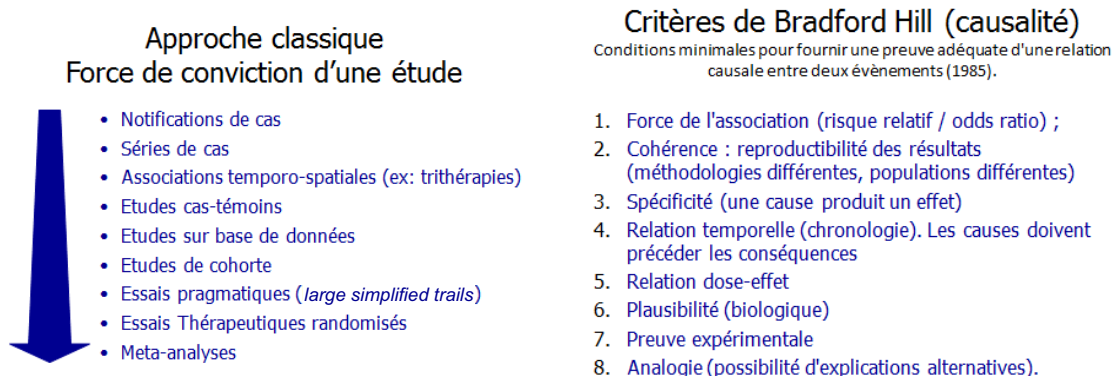


Fig. 1. Comparaison entre hiérarchisation des études selon leur force de conviction (niveau de preuve) par exemple comme défini par les critères de causalité de Bradford Hill^[5] et par la Haute autorité de santé.^[6]

1. Introduction

Les études pharmacoépidémiologiques et plus généralement les études d'évaluation de l'impact des médicaments ont fait l'objet de nombreux manuels, publications et recommandations. Au cours des 10 dernières années, ce thème a également été débattu aux Ateliers de Giens lors de plusieurs tables rondes. En 2002, la table ronde n° 2 a discuté de l'évaluation des médicaments post-commercialisation et décrit les différentes méthodologies possibles selon les questions posées.^[1] En 2004, la table ronde n° 2 a débattu de la place relative des essais cliniques comparatifs et des suivis de cohortes dans l'évaluation pré et post-autorisation de mise sur le marché (AMM) des médicaments.^[2] Les Ateliers de 2010 (table ronde n° 5) ont abordé la place des études post-AMM sous l'angle de l'évaluation du risque et de la pharmacovigilance.^[3] En 2011 enfin, l'étude de l'impact des médicaments en vie réelle a été analysée au travers des techniques de modélisation.^[4]

En 2012, le thème concernant les études pharmacoépidémiologiques était posé en termes de niveaux de preuve et des conditions pour les atteindre. De manière à proposer une approche pragmatique, le groupe de travail a pris le parti d'aborder le sujet sous l'angle de la question posée (le plus souvent par les autorités de santé ou les institutions) dans la mesure où celle-ci constitue le point de départ de toute réflexion donc du choix méthodologique. De la nature de la question et éventuellement des délais donnés pour y répondre, dépendent aussi les conditions d'optimisation du niveau de preuve (fonction de la qualité, des biais, des coûts).

Les questions ont été regroupées en 3 domaines principaux :

- les conditions d'utilisation/identification de la population cible/rejointe ;
- l'*effectiveness* ou performance du médicament en condition réelles de prescription/utilisation ;
- l'évaluation des risques/sécurité.

2. Approche systémique

Le groupe de travail a réfléchi à la meilleure approche possible selon qu'il s'agisse d'une part du contexte des études d'utilisation, dont l'objectif consiste à documenter la réalité de l'utilisation du produit en condition réelle d'utilisation et d'autre part, du contexte des études d'association, soit dans le cadre de la performance ou bien de la sécurité du produit. Pour chacun de ces deux domaines, le groupe de travail a discuté des critères d'études les plus appropriés pour répondre à la question, en s'appuyant *in fine* sur les critères de causalité de Bradford Hill.^[5]

Classiquement, la hiérarchisation des études repose sur leur force de conviction sous tendue par le schéma de l'étude elle-même (essai clinique, cohorte, cas témoins...), avec des présomptions d'ordre général concernant les biais éventuels. Cependant, cette hiérarchisation a été développée principalement dans un contexte de démonstration de l'efficacité et ne s'applique pas nécessairement aux autres domaines concernés par les études post-inscription (figure 1). Par exemple, selon les recommandations de la Haute autorité de santé (HAS) pour la réalisation de recommandations pour la pratique clinique,^[6] les études sont classées en 4 niveaux de preuve, du niveau 1 (essais comparatifs randomisés de forte puissance [*randomized clinical trials*, RCT] ; méta-analyse) au niveau 4 (séries de cas, études rétrospectives). Ceci permet d'obtenir une gradation des recommandations de A (preuve scientifique établie, conférée par l'information issue d'études de niveau 1) à C (faible niveau de preuve, correspondant aux études rétrospectives de qualité insuffisante ou de séries de cas).

D'autres sources proposent également de catégoriser le niveau de preuve selon le type d'étude, avec la même hiérarchisation.

Selon le centre d'*Evidence based medicine* d'Oxford (CEBM), le niveau de preuve d'une étude peut être rétrogradé sur la base des faiblesses intrinsèques de l'étude, d'imprécisions, du caractère indirect

de la preuve, à cause de l'incohérence des résultats entre les études, ou à cause de la taille de l'effet absolu jugée trop faible.^[7] Indépendamment du type d'étude, le niveau de preuve peut à l'inverse être supérieur si la taille de l'effet est grande ou très grande. De plus, selon le CEBM, le niveau de preuve d'une revue systématique est toujours meilleur que celui d'une étude individuelle. Les propositions d'évaluation de la qualité selon les recommandations du *grading of recommendations assessment, development and evaluation* (GRADE)^[8] reposent sur une approche qui s'appuie sur la conception de l'étude (essais ou études observationnelles) et proposent ensuite cinq raisons pour éventuellement diminuer la qualité de la preuve (biais, incohérence, caractère indirect de la relation, manque de précision et biais de publication) et trois raisons (importance de l'effet, relation dose effet, confusion résiduelle) pour éventuellement augmenter la qualité. Ceci bien sûr ne concerne que les études d'association, celles qui relient une exposition à un événement, positif ou négatif, mais ne sauraient s'appliquer aux études descriptives, qui relèvent d'une hiérarchie d'études bien différente.

Il a semblé au groupe que la hiérarchisation classique ne s'appliquait pas forcément à l'ensemble des situations rencontrées lors de la demande d'études pharmacoépidémiologiques. Pour se rapprocher du contexte des recommandations GRADE, le groupe a repris les critères de causalité de Bradford Hill, classiquement et historiquement utilisés pour apprécier le niveau de causalité évalué dans le contexte d'études observationnelles. Ces critères sont anciens puisque proposés dans les années soixante, avec une adaptation dans les années quatre-vingts, mais ont été repris récemment par d'autres auteurs pour hiérarchiser le niveau de preuve en fonction de la question posée et des études disponibles.^[9, 10] Comme ces critères utilisent l'ensemble de l'information pour apprécier la causalité, le groupe a retenu comme fil conducteur de la réflexion la prise en compte d'un faisceau de preuves.

Howick *et al.* proposent une simplification des lignes directrices originales de Bradford Hill, regroupées en trois catégories : preuves directes, mécanistiques et parallèles. Selon ces auteurs, la taille de l'effet revêt une importance supérieure au niveau de preuve individuel d'une étude donnée.^[9] Une étude observationnelle identifiant une taille d'effet suffisamment importante pour ne pas être potentiellement remise en question par des facteurs de confusion, peut ainsi être traitée sur un pied d'égalité avec les RCT. Ceci est également valable pour les RCT eux-mêmes. Ces auteurs suggèrent également de compléter ce niveau de preuve, par des éléments de preuve mécanistique (à partir d'études *in vitro* ou chez l'animal) et par des éléments de preuve parallèles (par la convergence des résultats d'autres études). Ceci rejoint l'intérêt des revues systématiques ou de la téléanalyse.^[11] Selon Wald et Morris, la téléanalyse peut être définie comme une synthèse de différentes catégories de preuves permettant d'obtenir une approche globale de la relation causale entre un facteur d'exposition et le risque d'une maladie. C'est en cela qu'elle diffère de la méta-analyse, puisqu'elle s'appuie sur des données issues de types d'études différents (cas rapportés, séries de cas, essais cliniques ou méta-analyses...) avec un niveau de preuve dif-

férent, plutôt que sur un seul type d'études.^[12,13] Ceci pourrait être particulièrement pertinent lorsque les RCT ne sont pas réalisables.

3. Population rejointe/modalités d'utilisation

Dans ce domaine, les questions posées cherchent le plus souvent à documenter la réalité de l'utilisation du produit en condition réelle. Ceci est la base de l'applicabilité des conclusions tirées des essais cliniques et autres études de haut grade de conviction. Pour un médicament dans une pathologie déterminée, elles portent par exemple sur le diagnostic et/ou le stade de la maladie (en lien avec l'indication), les caractéristiques des patients traités, leurs antécédents médicaux, les comorbidités, les traitements associés au médicament étudié, la posologie, la fréquence de la dispensation et l'observance du produit, la fréquence et les motifs des arrêts de traitement...

Dans ce contexte, la représentativité de l'échantillon de sujets utilisée dans l'étude constitue le point essentiel sur lequel reposera le niveau de preuves. La constitution de l'échantillon nécessitera donc des techniques de sondage adéquates, décrites de façon claire avec une justification des choix opérés dans le plan d'échantillonnage, suivie d'un redressement si nécessaire sur certains critères de représentativité. L'approche permettant de répondre à ce type de question est essentiellement de nature descriptive au travers d'études transversales ou longitudinales menées à partir de données issues de :

- bases de données (ex : pour la France : échantillon généraliste de bénéficiaires [EGB], Système national d'information inter-régimes [SNIIR-AM] ; base *Disease Analyzer/Thales* ; registre *ad hoc* par exemple registre des cancers...) et/ou
- des études de terrain *ad hoc*, en essayant de s'assurer de la représentativité des populations incluses.

De manière générale, pour ces études, il convient de respecter les bonnes pratiques d'épidémiologie, et des techniques d'enquêtes et d'échantillonnages. L'épidémiologie peut/doit collaborer avec les enquêteurs pour atteindre le niveau de preuve approprié

Dans le cas d'un nouveau médicament, il conviendra également de s'assurer avant le démarrage de l'étude que le produit soit suffisamment représenté dans le marché pour pouvoir conduire une étude descriptive. En effet, au moment de la mise sur le marché d'un nouveau produit, les prescripteurs vont d'abord le « tester » sur des patients particuliers, résistants ou en échec des autres produits déjà sur le marché, afin de se familiariser avec le nouveau médicament. Un délai de 6 à 12 mois après le lancement semblerait utile pour pouvoir analyser des pratiques de prescription « stables ». Ce délai est à proposer selon le produit spécifiquement concerné, car les pratiques varient selon la pathologie concernée, les alternatives thérapeutiques, l'efficacité et le risque perçus pour le nouveau produit.

Par ailleurs et de manière à garantir un niveau de preuve acceptable, il conviendra de s'assurer de la prise en compte dans le

protocole de la représentativité de l'échantillon sur lequel porte l'étude ainsi que d'analyser et discuter de l'impact des biais de sélection et d'information. Ceci peut imposer l'analyse de sources de données différentes dont la convergence confirme l'absence *a priori* de biais de sélection.

4. Effectiveness/performance en vie réelle

Dans ce domaine les études peuvent chercher à évaluer et comparer la performance (*effectiveness*) d'un produit en situation post-AMM par rapport à l'efficacité documentée lors du programme de phase clinique (données pré-AMM). La question est parfois également posée pour un nouveau produit, en termes de performance relativement à des alternatives thérapeutiques qui n'auraient pas été utilisées comme comparateur en pré-AMM.

Dans ce contexte, les méthodes permettant de répondre à la question peuvent être de type :

- cohorte : de pathologie/classe thérapeutique/produit étudié (sur base de données - EGB/SNIIRAM, programme de médicalisation des systèmes d'information [PMSI], *disease analyzer*, Thalès - ou *ad hoc* terrain, en fonction du critère principal retenu). Les techniques modernes d'ajustement ou d'appariement, en particulier basées sur des variables instrumentales ou des scores de propension y compris à hautes dimensions, en conjonction avec une analyse très attentive des biais potentiels permettent de réduire ceux-ci et d'avoir des résultats très pertinents ;
- essais pragmatiques, avec les difficultés de faire comprendre la notion encore très incertaine d'essai pragmatique par rapport aux très grands essais cliniques simples ;
- étude cas-témoin (exemple dans le domaine de la prévention ou de la vaccination où l'évènement d'intérêt est très rare) : études sur bases de données ou *ad hoc*. Ces études cas-témoin imposent une attention et un soin tout particulier aux questions méthodologiques et de contrôle des biais, qui méritent des explications et une analyse très précise ;
- modélisation : celle-ci présente un intérêt particulier dans les cas où l'on cherche à évaluer l'impact potentiel à long terme d'un nouveau produit ou simuler son impact en termes de morbi-mortalité (exemple : vaccination). La modélisation suppose d'avoir des données précises de base, qui peuvent imposer des études spécifiques pour obtenir les éléments de base de la modélisation.

En termes de recommandations et de manière à déterminer le type d'étude à mettre en œuvre, il importe de :

- définir *a priori* l'hypothèse de performances attendues à partir des résultats d'essais cliniques ;
- vérifier les informations disponibles dans les bases de données ou registres existants (cf. champ précédent) et donc la faisabilité d'une étude sur base de données en identifiant les possibles biais ;

- à défaut prévoir et mettre en place une étude *ad hoc*, à condition que ses modalités de réalisation soient effectivement susceptibles d'apporter des informations utiles.

Il convient de rappeler que l'observation d'un bénéfice inattendu nécessitera toujours un essai randomisé pour obtenir une AMM.

5. Sécurité

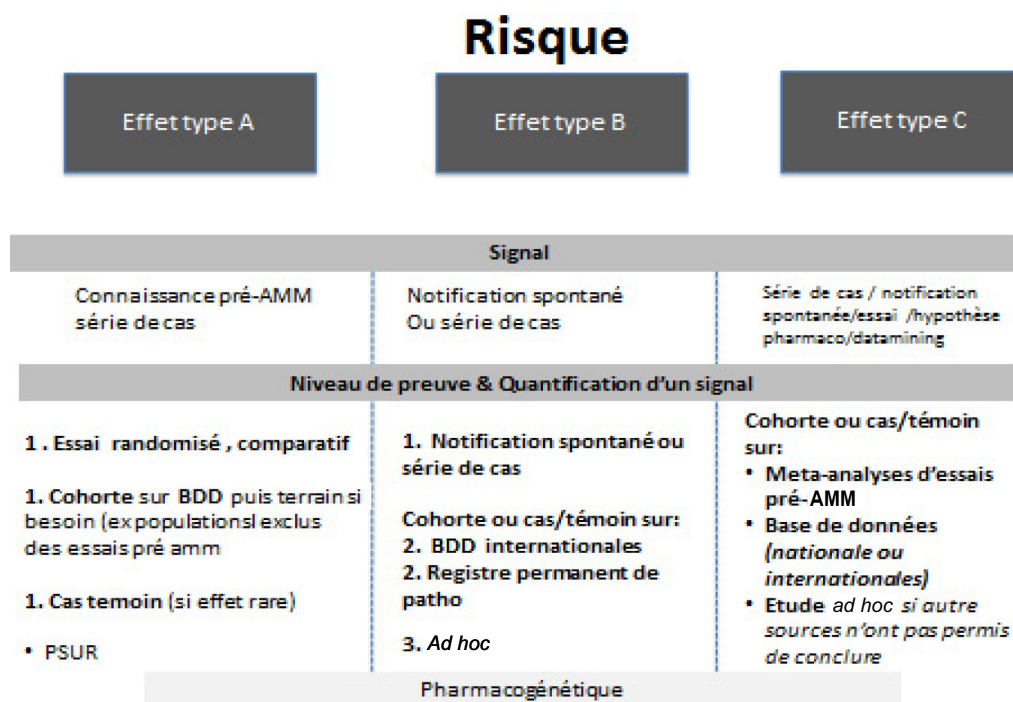
L'approche permettant de répondre à une question portant sur la sécurité d'un médicament, dépend du type de ou des effets indésirables concernés.

Plusieurs classifications des effets indésirables médicamenteux sont disponibles (*World Health Organisation* [WHO], *medical dictionary for regulatory activities* [MedDRA]...), selon la nature, la gravité, la fréquence... Il existe également une classification des types d'effets indésirables selon leur mécanisme, distinguant jusqu'à 6 catégories par la prise en compte par exemple des effets de sevrage ou des effets temps-dépendants.^[14] Dans le cadre de la table ronde, le groupe a retenu les 3 premières des 6 catégories proposées, car elles représentent la majorité des situations pour lesquelles des études pharmacoépidémiologiques sont demandées.

Ces 3 types d'effets sont les suivants :

- effet de type A : effet indésirable (EI) pharmacologiquement déterminé et/ou attendu du produit expliqué par ses caractéristiques pharmacodynamiques et/ou pharmacocinétiques.^[14] Ce type d'effet est fréquent, prévisible et facilement identifiable lors des essais cliniques, bien que représentant l'immense majorité des EI graves et des décès pour cause médicamenteuse. Exemples : hypoglycémie et médicaments oraux du diabète, hémorragie sous anticoagulants, diarrhée sous antibiotiques, chutes sous benzodiazépines etc. ;
- effet de type B : effet indésirable non pharmacologiquement déterminé ; l'effet indésirable apparaît à la suite de la prise du médicament (chronologie d'apparition évidente), l'effet est spécifique, généralement grave et rare ; il se manifeste typiquement de manière identique et est en général non prévisible sauf en cas d'exposition précédente ayant entraîné une première réaction. Exemples : réactions immuno-allergiques (rash maculo-papuleux et macrolides, syndrome de Lyell et antiépiléptiques...)
- effet de type C : dans le cadre de la réflexion du groupe, il s'agit d'un évènement indésirable spontanément présent dans la population générale, mais dont on suspecte une augmentation de la fréquence d'apparition en lien avec le médicament avec une relation chronologique difficile à établir. Exemples : cancers et médicaments du diabète, infarctus du myocarde et anti-inflammatoires non stéroïdiens.

La nature de ces 3 grands types d'effets implique des approches différentes à la fois pour leur identification (signal) et pour l'évaluation de la causalité, et donc des schémas d'étude appropriés à



AMM : autorisation de mise sur le marché ; BDD : bases de données ; PSUR : *periodic safety update reports*

Fig. 2. Représentation synthétique des schémas d'étude du risque les plus adaptés selon le type d'effet indésirable.

chacune de ces catégories. La figure 2 propose de façon synthétique pour chaque catégorie d'effet, une hiérarchisation des études selon le niveau de conviction pour explorer au mieux la question.

Pour un effet de type A, par définition, des données sont disponibles en pré-AMM. Les études permettant d'obtenir le meilleur niveau de preuve seront une étude clinique comparative randomisée, ou une étude de cohorte dans une base de données, éventuellement complétée par une étude de terrain *ad hoc* (notamment pour étudier le médicament dans une population exclue des études pré-AMM, exemple de l'exposition pendant la grossesse). Si l'évènement est rare, une étude cas-témoins peut également fournir des résultats d'un niveau de preuve optimal, compte-tenu de la rareté de l'évènement.

Pour un effet de type B, non identifiable en pré-AMM et imprévisible, le signal émane souvent de la notification spontanée. L'intensité du signal (fréquence des cas) va influencer la recherche du niveau de preuve. La notification spontanée peut en elle-même s'avérer suffisante si l'évaluation intrinsèque de l'imputabilité de la ou des observations recueillies est importante, surtout si l'évènement est rare et typique, ce qui lui confère le niveau le plus élevé dans la hiérarchisation du niveau de preuve. Viennent ensuite les études pharmacoépidémiologiques de type cohorte ou cas-témoins, soit sur base de données ou à partir de registres de pathologies ou de patients traités, et enfin, en dernier lieu, une étude *ad hoc*, dont la taille sera souvent réhibitoire au vu de la fréquence souvent rare de ces événements. Quelque soit le type d'approche retenu, la mise

à disposition d'arguments explicatifs de l'effet observé dans des groupes de patients particuliers, ou issus d'exploration de type pharmacogénétique, confère à chaque étude un niveau de preuve plus important que l'information issue de l'étude seule.

Pour un effet de type C, le signal émane en général de sources variées, qu'il s'agisse de la notification spontanée, d'une série de cas, d'une hypothèse pharmacologique, des observations issues d'essais cliniques ou de *datamining* sur bases de données médicales. Le préalable à la mise en place d'études est de documenter le risque de base de la pathologie représentant l'effet indésirable dans la population, ce qui permettra ensuite l'interprétation du risque absolu et du risque relatif. Sans pouvoir véritablement hiérarchiser les études, le niveau de preuve optimal pourra être obtenu à partir de méta-analyse des essais pré-AMM, d'études de cohorte ou d'étude cas témoins sur base de données de grande taille (nationale ou internationale), ou à partir d'étude *ad hoc* si aucune autre source n'a permis de conclure.

6. Conclusion

En post-AMM, l'évidence acquise dans les essais en pré-AMM (y compris effet de classe) sert de référence pour la génération des hypothèses.

Les recommandations de bonnes pratiques d'étude doivent être suivies : critères pour obtenir le sceau des études *European Network of Centres for Pharmacoepidemiology and Pharmacovigilance* (ENCePP)^[15-17] (auditable, reproductible, transparence des données...), guide méthodologique de la HAS.^[18]

L'identification des biais potentiels et l'analyse de l'impact positif ou négatif de ces biais sur le résultat est un des éléments essentiels. Tout ce qui serait de nature à modifier la confiance dans le résultat doit être pris en compte éventuellement dans une analyse de sensibilité ou une approche de modélisation.

L'intégration qualitative et/ou quantitative de l'ensemble des données (pharmacologie fondamentale, essais cliniques, série de cas...), à travers une approche de téléanalyse est un élément essentiel. Le cas échéant, la combinaison de plusieurs approches complémentaires (issues de sources différentes), soit qualitativement selon l'approche de la combinaison des critères de Bradford-Hill, soit à partir d'une modélisation quantitative, accompagnée par des analyses de sensibilité, pourrait se révéler préférable à une seule étude. Ceci est à discuter et optimiser en fonction de la faisabilité, des délais (imposé pour l'étude et nécessaire pour la mesure du *surrogate endpoint* ou *outcome* final) et du coût. L'étape suivante de la réflexion devra porter sur l'intégration de toutes ces informations dans des modèles de quantification de l'effet ou de sa perception, utilisant en particulier des techniques bayésiennes.

En résumé, plus qu'un niveau de preuve, c'est un faisceau de preuves qui permettra de prendre la décision.

Conflits d'intérêts. Aucun.

Abréviations. AMM : autorisation de mise sur le marché ; BDD : bases de données ; CEBM : *evidence based medicine* (Oxford) ; EI : effet indésirable ; EGB : échantillon généraliste de bénéficiaires ; ENCePP : *European Network of Centres for Pharmacoepidemiology and Pharmacovigilance* ; GRADE : *grading of recommendations assessment, development and evaluation* ; HAS : Haute autorité de santé ; PMSI : programme de médicalisation des systèmes d'information ; RCT : essais comparatifs randomisés de forte puissance ; SNIIRAM : Système national d'information inter-régimes.

Références

1. Le Gales C, el Hasnaoui A, Goehrs JM. Postmarketing evaluation of drugs. Actual efficacy, population exposed and impact on public health. *Thérapie* 2003; 58(3): 209-19
2. Vray M, Hamelin B, Jaillon P. The respective roles of controlled clinical trials and cohort monitoring studies in the pre- and postmarketing assessment of drugs. *Thérapie* 2005; 60(4): 339-44, 345-9
3. Tubach F, Lamarque-Garnier V, Castot A. Role of the post-marketing authorisation studies in drug risk surveillance: specifications and methodologies. *Thérapie* 2011; 66(4): 355-62, 347-54
4. Gueyffier F, Strang CB, Berdaux G, et al. Contribution of modeling approaches and virtual populations in transposing the results of clinical trials into real life and in enlightening public health decisions. *Thérapie* 2012; 67(4): 367-74
5. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58: 295-300
6. Haute autorité de santé. Élaboration de recommandations de bonne pratique : Méthode « Recommandations pour la pratique clinique » in Guide méthodologique. 2010. http://www.has-sante.fr/portail/upload/docs/application/pdf/2011-01/guide_methodologique_recommandations_pour_la_pratique_clinique.pdf
7. Trans Durieux N, Pasleau F, Howick J. Table des niveaux de preuves du Centre d'Evidence Based Medicine d'Oxford (OCEBM Levels of Evidence Working Group). The Oxford 2011 Levels of Evidence, 2011. <http://www.cebm.net>
8. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64(4): 401-6
9. Howick J, Glasziou P, Aronson JK. The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? *J R Soc Med* 2009; 102(5): 186-94
10. Howick JP, Glasziou P, Aronson JK. Evidence-based mechanistic reasoning. *J R Soc Med* 2010; 103(11): 433-41
11. Wald NJ, Morris JK. Teleanalysis: combining data from different types of study. *BMJ* 2003; 327 (7415): 616-8
12. Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007; 7: 32
13. Singh S, Loke YK, Furberg CD. Thiazolidinediones and heart failure: a teleanalysis. *Diabetes Care* 2007; 30(8): 2148-53
14. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 2000; 356 (9237): 1255-9
15. Blake KV, Devries CS, Arlett P, et al. Increasing scientific standards, independence and transparency in post-authorisation studies: the role of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2012; 21 (7): 690-6
16. Blake KV, Prilla S, Accadebled S, et al. European Medicines Agency review of post-authorisation studies with implications for the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2011; 20(10): 1021-9
17. Berger ML, Dreyer L, Anderson F, et al. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health* 2012; 15(2): 217-30
18. Stamenkovic S, Solesse A, Zanetti L, et al. French Health Authority (HAS) Guidance: the post-registration studies on health technologies (drugs, medical devices and instruments): principles and methods. *Thérapie* 2012; 67(5): 409-21

Correspondance et offprints : Maryse Lapeyre-Mestre, Équipe de Pharmacoépidémiologie, INSERM 1027, Université Paul Sabatier – Toulouse 3, Service de Pharmacologie clinique, Faculté de Médecine, 37 allées Jules Guesde, 31000 Toulouse, France.
E-mail : maryse.lapeyre-mestre@univ-tlse3.fr