Thérapie 2013 Juillet-Août; 68 (4): 247–252
DOI: 10.2515/therapie/2013038

Pharmacoepidemiology Studies: what Levels of Evidence and how can They be Reached?

Maryse Lapeyre-Mestre[1], Claudine Sapède[2], Nicholas Moore[3] and participants of round table N° 5 of Giens Workshops XXVIII: Pascal Bilbault[4], Patrick Blin[5], Damien Chopy[6], David Evans[7], François Gueyffier[8], Laure Lacoin[9], Muriel Malbezin[10], Joëlle Micallef[11], Dana Morlet-Vigier[12], Sophie Muller[13], Emmanuel Oger[14], Yannick Plétan[15], Gérard Pons[16], Patrice Verpillat[17] and Caroline Vigneau[18] †

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

Text received March 15th, 2013; accepted May 7th, 2013

Keywords:
evaluation;
effectiveness;
risk;
treated population;
level of evidence

Abstract – In pharmacoepidemiology studies, the nature of the research question will dictate the choice of methodological approach and the conditions for optimizing the level of evidence. Thus, to document the treated population and the modes of use of a new drug in real-life prescribing conditions, a descriptive approach through cross-sectional or longitudinal studies conducted on databases, or else ad-hoc studies, will be preferred. On the other hand, evaluation of the real-life "effectiveness" of a new drug will be based on cohort, case-control or scientific modeling, depending on the drug and the disease of interest. For questions involving drug risks and safety, it is the adverse effects profile that will guide the choice of study design, both for identification of the effect (signal) and assessment of causation. In all cases, in the post-marketing authorization (MA) setting, the evidence acquired in pre-MA studies serves as the basis for generating hypotheses. Whatever the research question and the method chosen to address it, the potential biases and their impact on the results need to be identified. In certain cases, a combination of several complementary approaches may prove preferable to a single study.

Abbreviations: see end of article.

† Articles, analyzes and proposals from the Giens workshops are those of the authors and do not prejudice the position of their parent organization.

Article publié par EDP Sciences

Fig. 1. Comparison between hierarchies of studies according to their confidence value (level of evidence) for example as defined by Bradford Hill causation criteria[5] and *Haute autorité de santé*.[6]

1. Introduction

Pharmacoepidemiology studies and, more generally, studies evaluating the impact of drugs, have been the subject of numerous manuals, publications and guidelines. Over the last 10 years, this theme has been debated at the round tables of the Giens workshops. In 2002, round table No. 2 addressed the post-marketing evaluation of drugs and described the different possible methodologies as a function of the questions asked.[1] In 2004, round table No. 2 examined the respective roles of comparative clinical trials and cohort monitoring studies in the pre- and post-marketing assessment of drugs.[2] The 2010 workshops (round table No. 5) debated the role of post-marketing studies from a standpoint of risk assessment and pharmacovigilance.[3] In 2011, the impact of drugs in the real-life setting was analyzed through scientific modelling approaches.[4]

In 2012, the topic of pharmacoepidemiology studies was addressed, in terms of levels of evidence and how they can be reached. To propose a pragmatic approach, the working group decided to examine the subject from the perspective of the question being asked (usually by health authorities or institutions), since this constitutes the starting point for any reflection and therefore guides the choice of methodology. Furthermore, the nature of the question and possibly the time frames in which it must be answered define the conditions for optimizing the level of evidence (considerations of quality, biases and cost).

The questions were divided into three main categories:
- conditions of use/identification of target and treated populations;
- "effectiveness" or performance of the drug in real-life conditions of prescription and use;
- safety and risk assessment.

2. System approach

The working group reflected upon the best possible approach, according to the context of studies of use on the one hand, which aim to document the real-life conditions of use of a drug, or the context of association studies, pertaining either to effectiveness or safety, on the other. For each of these two domains, the working group discussed the most appropriate criteria by which to answer the question, ultimately based on the Bradford Hill guidelines for causation.[5]

Classically, studies are placed into hierarchies according to their confidence value underpinned by the design of the study itself (clinical trial, cohort or case-control study, etc.), and the general assumptions about potential biases. However, these hierarchies were developed primarily in a context of demonstrating efficacy and do not necessarily apply to other domains of post-marketing assessment (figure 1). For example, according to the French Health Authority (*Haute autorité de santé*, HAS) recommendations for elaborating clinical practice guidelines,[6] studies are classified into four levels of evidence, from level 1 (randomized clinical trials [RCT] with high power; meta-analyses) to level 4 (case series, retrospective studies). This leads to a ranking of recommendations from A (established scientific evidence conferred by data from level 1 studies) to C (low level of evidence, corresponding to low-quality retrospective studies or case series).

Other sources also propose categorizing the level of evidence according to study design, with the same hierarchy.

According to the Oxford Centre for Evidence Based Medicine (CEBM), the level of evidence of a study can be graded down due to intrinsic weaknesses, imprecision, indirect nature of evidence, inconsistency between studies, or an absolute effect size that is too small.[7] Independently of study design, the level of evidence can conversely be graded up if the effect size is large or very large. Also, according to the CEBM, the level of evidence of a systematic review is always higher than that of an individual study. The grading of recommendations assessment, development and evaluation (GRADE) recommendations for rating the quality of evidence[8] are based on study design (trials or observational studies) and propose five reasons to possibly rate down the quality of evidence (bias, inconsistency, indirectness, imprecision and publication bias) and three reasons to possibly rate up the quality of evidence (size of effect,

dose-response, residual confounding). Of course, this only concerns association studies which relate exposure to a positive or negative outcome, and does not apply to descriptive studies which fall under a quite different hierarchy.

The working group considered that the classical hierarchies do not necessarily apply to all types of situations encountered when pharmacoepidemiology studies are requested. To focus on the context of the Grade recommendations, the group revisited the Bradford Hill guidelines for causation, classically and historically used to assess causation in the context of observational studies. These guidelines, first proposed in the 1960s and modified in the 1980s, were recently revised by other authors to rank the level of evidence according to the research question and the available studies.[9, 10] As these criteria make use of the entire set of information to assess causation, the guiding principle in the group's discussions was the consideration of a bundle of evidence.

Howick *et al*. propose a simplification of the original Bradford Hill guidelines into three categories: direct evidence, mechanistic evidence and parallel evidence. These authors state that the size of effect has more importance than the individual level of evidence of a given study.[9] An observational study identifying an effect large enough to outweigh the combined effects of plausible confounders can thus be considered on a par with RCTs. This is also the case for RCTs themselves. These authors also propose supplementing this level of evidence by mechanistic evidence (from *in vitro* or animal studies) and parallel evidence (from related studies with convergent results), in addition to systematic reviews or teleoanalysis.[11] According to Wald and Morris, teleoanalysis can be defined as the synthesis of different categories of evidence to obtain a quantitative general summary of (*a*) the relation between a cause of a disease and the risk of the disease and (*b*) the extent to which the disease can be prevented.[11] Teleoanalysis is different from meta-analysis because it relies on combining data from different study designs (case reports or case series, randomized clinical trials, meta-analyses...) across all grades of evidence rather than one type of study.[12, 13] This could be especially useful when RCTs are unfeasible.

3. Treated population/modalities of use

In this domain, the questions are usually aimed at documenting the real-world conditions of use of a drug. This defines the extent to which the conclusions from clinical trials and other studies with high confidence value are applicable. For a drug in a defined disease, the questions relate to diagnosis and/or disease stage (related to the indication), characteristics of treated patients, medical history, comorbidities, concomitant treatments, dosage, dispensing frequency and adherence to treatment, and the frequency and reasons for discontinuing treatment, to give a few examples.

In this context, sample representativeness is the essential point on which the level of evidence is based. Sample selection therefore requires suitable survey techniques that are clearly described and justified in the sampling plan, followed by a revision of certain criteria of representativeness when necessary. The approach by which to address this type of question is mainly of a descriptive nature through cross-sectional and longitudinal studies on data from:

- databases (eg. for France: representation sample of health insurance beneficiaries [*échantillon généraliste de bénéficiaires*, EGB], National inter-scheme Health Insurance Information system [*système national d'information inter-régimes*, SNIIR-AM]; Disease Analyzer/Thales; *ad hoc* registry, eg. cancer registry, etc.) and/or
- *ad hoc* field studies, ensuring that the study will capture the representativeness of the included populations.

Generally speaking, these studies should comply with good epidemiology practices and sampling/survey methods. The epidemiologist can and should collaborate with the investigators in order to achieve the suitable level of evidence.

- In the case of a new drug, it is important to make sure that there is sufficient market penetration to be able to conduct a descriptive study before commencing the study. This is because when a new drug is launched, prescribers initially "test" it in specific subsets of patient who are refractory or have failed other drugs already on the market, in order to acquaint themselves with the new drug. Between six and twelve months after product launch is considered a useful lag time to reach "stable" prescribing practices. *This interval depends on the drug in question because practices vary according to the disease of interest, the alternative treatments, and the efficacy and perceived risks of the new drug*;
- furthermore, so as to guarantee an acceptable level of evidence, sample representativeness should be taken into account in the protocol and the effects of selection and information biases should be analyzed and discussed. This may require an analysis of different source data which, if convergent, will confirm in advance the absence of selection bias.

4. Real-life effectiveness/performance

In this domain, studies are aimed at evaluating and comparing the effectiveness of a drug in the post-MA setting by comparing it with efficacy data from the pre-MA phase. The question can also be asked for a new drug, in terms of effectiveness relative to alternative treatments not used as comparator in pre-MA studies.

In this context, the methodological designs to address the question can be of the type:

- cohort: disease/therapeutic class/drug under study (on databases-EGB/SNIIRAM, programme for medicalization of information systems [*programme de médicalisation des systèmes d'information*, PMSI], Disease Analyzer, Thalès - or *ad hoc*, depending on the main outcome measure chosen). Modern methods of adjustment and matching, in particular

based on instrumental variables or propensity scores including high dimensional, in conjunction with very careful analyses of potential biases, make it possible to minimize bias and yield highly pertinent results;
– pragmatic trials, with the difficulties of clarifying the still very uncertain notion of a pragmatic trial *versus* simple, very large clinical trials;
– case-control study (*e.g.* in the domain of prevention or vaccination where the event of interest is very rare): studies on databases or *ad hoc* studies. Such studies require very rigorous attention and care to methodological issues and control of biases, which must be clearly explained and very accurately analyzed;
– scientific modelling: this is of particular interest when seeking to evaluate the potential long-term impact of a new drug or to simulate its impact on morbidity and mortality (*e.g.* vaccination). Accurate data must form the basis of any modelling approach and specific studies may be necessary to obtain such data.

In terms of recommendations and choice of study design, it is important to:
– first formulate hypotheses about the projected effectiveness based on data from clinical trials;
– check the information available in existing databases and registries (see previous section) and the feasibility of a database study, identifying any potential biases;
– otherwise plan and set up an *ad hoc* study, making sure that the modalities of study execution can actually contribute useful information.

It should be recalled that observation of an unexpected benefit will always require a randomized trial in order to obtain a MA.

5. Safety

The approach to questions relating to drug safety depends on the type of undesirable effect(s) concerned.

Several adverse effect classifications exist (World Health Organization [WHO], medical dictionnary for regulatory activities [MedDRA], etc.), according to nature, severity, frequency, etc. There is also a classification based on mechanism, containing up to six categories which take into account withdrawal effects or time-dependent effects, for example.[14] In the round table discussions, the working group retained the first three of the six proposed categories because they account for the majority of situations for which pharmacoepidemiology studies are requested.

These three types of effects are as follows:
– type A effect: a reaction related to the pharmacological action of the drug and/or a predictable reaction,[14] explained by the drug's pharmacodynamic and/or pharmacokinetic characteristics. This type of effect is common, predictable and easily identifiable during clinical trials, even though it accounts for the majority of serious adverse effects (AEs) and drug-related deaths. Examples include hypoglycaemia and oral antidiabetics, haemorrhage during anticoagulant therapy, diarrhoea caused by antibiotics, falls in patients taking benzodiazepines;
– type B effect: a reaction not related to the pharmacological action of the drug; the reaction occurs after taking the drug (clear time-relation), the reaction is specific, generally serious and rare; it typically manifests in an identical manner and is usually unpredictable except in case of a previous exposure which provoked an initial reaction. Examples include immuno-allergic reactions (maculopapular rash and macrolide antibiotics, Lyell syndrome and antiepileptics, etc.);
– type C effect: the working group considered this to be a reaction spontaneously present in the general population but suspected to increase in frequency in relation to the drug with a chronological relationship that is difficult to establish. Examples include cancer and diabetes medications, myocardial infarction and NSAIDs, etc.

The nature of these three major types of effects requires different approaches for identification (signal) and evaluation of causation, and hence for defining the study design suited to each category. Figure 2 illustrates, for each of these categories, a hierarchy of studies according to their level of confidence to best answer the question.

- For type A effects, by definition, the data are available from pre-MA studies. The studies yielding the highest level of evidence would be randomized comparative clinical trials or cohort studies in a database, possibly supplemented by an ad hoc field study (in particular for populations excluded from pre-MA studies such as pregnant patients). If the event is rare, a case-control study can also provide an optimal level of evidence in light of the rarity of the event;
- for type B effects, non-identifiable in pre-MA studies and unpredictable, the signal often emanates from spontaneous notifications. The signal intensity (frequency of cases) will influence the desired level of evidence. Spontaneous notification may itself be sufficient if the intrinsic causation assessment of the report(s) is high, especially if the event is rare and typical which confers the highest level in the evidence hierarchy. Next come pharmacoepidemiology studies (case-control and cohort), either on databases or disease or patient registries, finally followed by *ad hoc* studies, whose size may be prohibitive due to the often very low frequency of the event. Whatever approach is chosen, the availability of explanatory arguments observed in specific patient groups or derived from pharmacogenetic investigations gives each study a higher level of evidence than the data from the study alone;
- for type C effects, the signal generally comes from various sources, be it spontaneous notifications, case series, a pharmacological hypothesis, clinical trial data or datamining on medical databases. The prerequisite for implementing a study is to document the basic risk of the disease causing the adverse

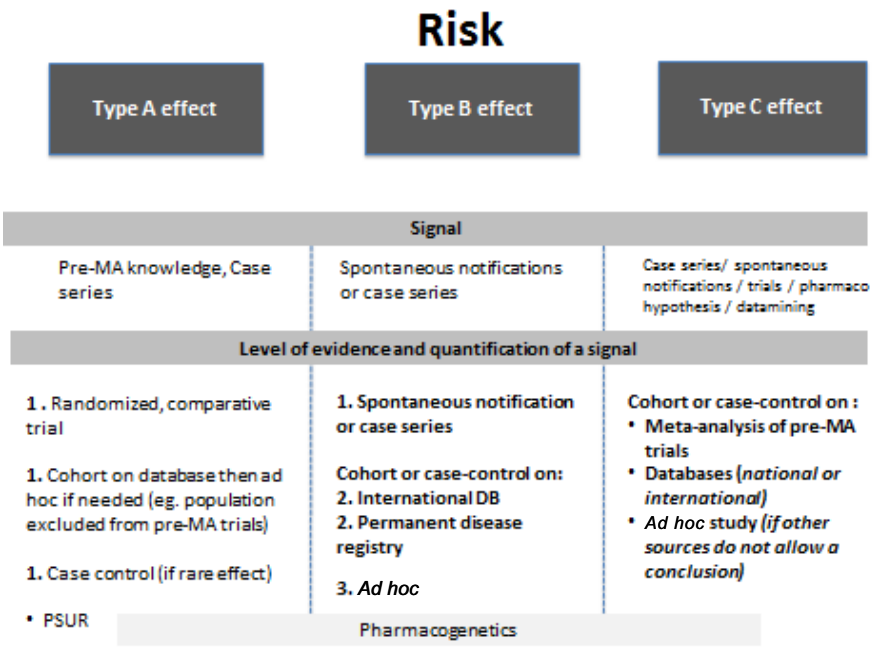

Fig. 2. Schematic representation of the most suitable study designs to assess risk according to type of adverse effect.
DB: databases; **MA**: marketing authorization; **PSUR**: periodic safety update reports.

effect in the population, which then allows an interpretation of the absolute risk and the relative risk. Without truly being able to establish a hierarchy of studies, the optimal level of evidence can be obtained from a meta-analysis of pre-MA trials, cohort or case-control studies on large (national or international) databases, or from *ad hoc* studies if no other sources allow a conclusion.

6. Conclusion

– In the post-MA setting, the evidence acquired in pre-MA studies (including class effects) serves as reference for generating hypotheses.
– Good practice guidelines should be followed: criteria to obtain the seal of European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) studies[15-17] (auditable, reproducible, data transparency, etc.), HAS methodological guidance.[18]

Identification of potential biases and analysis of their positive or negative impact on the results is an essential element. Anything that might compromise confidence in the result should be taken into account, possibly in a sensitivity analysis and/or scientific modelling approach.

– The qualitative and/or quantitative incorporation of all data (basic pharmacology, clinical trials, case series, etc.) through a teleoanalysis approach is essential. Where necessary, a combination of several complementary approaches (from different sources), either qualitatively by combining the Bradford Hill guidelines, or by quantitative modelling, accompanied by sensitivity analyses, might be preferable to a single study. This should be discussed and optimized according to feasibility, timelines (required for the study and necessary for measurements of the "surrogate endpoint" or "final outcome"), and costs. The next step in our reflection should concern the integration of all this information in models to quantify effect or perception of effect, in particular through the use of bayesian approaches.

In summary, it is a bundle of evidence that informs decision-making, rather than a level of evidence.

Conflicts of interests. None.

Abbreviations. AEs: adverse effects; CEBM: center for evidence-based medicine; DB: databases; EGB: representative sample of Health insurance beneficiaries (*échantillon généraliste de bénéficiaires*); ENCePP: European Network of Centres for Pharmacoepidemiolgy and Pharmacovigilance; GRADE: grading of recommendations assessment, development and evaluation; HAS: French Health Authority (*Haute autorité de santé*); SNIIRAM: national inter scheme health insurance information system (*système national d'information inter régimes*); MA: marketing authorization.

References

1. Le Gales C, el Hasnaoui A, Goehrs JM. Postmarketing evaluation of drugs. Actual efficacy, population exposed and impact on public health. Therapie 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. Therapie 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. Therapie 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. Therapie 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. Teleoanalysis: 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 teleoanalysis. 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. Therapie 2012; 67(5): 409-21

Correspondence and 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.
Email: maryse.lapeyre-mestre@univ-tlse3.fr

