

# Lecture critique et EBM : l'étude LIFE\*

## *Critical reading and EBM : the LIFE trial*

**M. Roland<sup>1</sup> et C. Mélot<sup>2</sup>**

<sup>1</sup>D.M.G.-U.L.B., <sup>2</sup>Service des Soins Intensifs, Hôpital Erasme

### RESUME

*Depuis une dizaine d'années, l'“ evidence-based medicine ” a progressivement envahi tout le champ médical : la clinique mais aussi la recherche et l'enseignement. Plus qu'un nouveau paradigme, elle constitue l'un des piliers de la bonne pratique médicale à côté de l'expérience, de l'éthique et des demandes du patient. Elle comporte des aspects quantitatifs collectifs (statistiques, santé publique, épidémiologie, recherche biomédicale) appliqués à la réalité individuelle de chaque patient en demande de soins. Une des méthodes utilisées en “ evidence-based medicine ” est la lecture critique : analyse en profondeur des recherches publiées pour en extraire et valider les éléments applicables à la clinique quotidienne. L'étude LIFE, une étude comparant un sartan et un bêtabloquant chez les patients hypertendus avec hypertrophie ventriculaire gauche, sert d'exemple pour expliciter concrètement les différentes étapes et méthodologies de la lecture critique.*

*Rev Med Brux 2007 ; 28 : 97-110*

### ABSTRACT

*For ten years, “ evidence-based medicine ” gradually invaded all the medical field : clinical practice but also research and teaching. More than being just a new paradigm, it constitutes one of the pillars of good medical practice beside experimentation, ethics and the requests of the patient. It gathers collective quantitative aspects (statistics, public health, epidemiology, biomedical research) applied to the individual reality of each patient in search of care. One of the methods used in “ evidence-based medicine ” is critical reading : in-depth analysis of the published trials in medical journals, to extract and validate the elements applicable to the daily clinical practice. The LIFE trial, designed to compare treatment with a sartan or a beta-blocker for hypertensive patients with left ventricular hypertrophy, is used as a model to concretely explicit the various steps and methodologies of critical reading.*

*Rev Med Brux 2007 ; 28 : 97-110*

*Key words : cardiovascular disease, human, mortality, morbidity, prevention, angiotensin receptor, angiotensin II, angiotensin antagonist, antihypertensive agent, losartan, atenolol, chemotherapy, treatment efficiency, hypertension, evidence-based medicine*

Depuis une dizaine d'années, aucun médecin ne peut plus ignorer l'“ *evidence-based medicine* ” (EBM), médecine basée sur la(les) preuve(s). Le concept, en tant que tel, n'est apparu que tout récemment mais, en des délais très courts, a progressivement envahi tout le champ médical.

Présentée souvent comme un nouveau mouvement de pensée (un changement paradigmatique), elle relève cependant d'une conception épistémologique qui remonte au 19<sup>ème</sup> siècle, le positivisme. A ce moment, le positivisme a permis de sortir la science de l'obscurantisme et de la pensée religieuse, et il a fourni les bases théoriques de toute

notre clinique et de toute notre recherche. Néanmoins, assimiler tout le non mesurable à du non tangible donc à de l'inexistant semble une hypothèse peu acceptable, un peu comme si les ultrasons n'existaient pas parce que l'oreille de l'homme ne les perçoit pas. Toute la science moderne (de la physique quantique à la philosophie, en passant par la médecine de l'humain qui doit intégrer les deux) souligne aujourd'hui l'importance du hasard (par rapport à la nécessité), de la complexité (par rapport au réductionnisme), de la circularité (par rapport à la linéarité), de la causalité

\* Conférence présentée lors du symposium du C.U.M.G.-U.L.B. à Chypre, avril 2006.

multiple et rétroactive (par rapport à la causalité déterministe). Karl Popper<sup>1</sup> soulignait déjà le fait que les théories scientifiques ne sont que des inventions humaines, comme des filets créés par nous pour appréhender le monde. Ces théories ne sont jamais des instruments parfaits, et elles ne doivent pas être confondues avec une représentation complète de tous les aspects de la réalité, pas même si elles sont très réussies, pas même si elles semblent donner d'excellentes approximations. Le monde vu par un œil d'homme et par un œil d'insecte n'est pas le même monde même si c'est le même monde ; et nous ne pouvons qu'imaginer le monde des insectes avec nos instruments d'hommes. Le positivisme, la pensée bio-médicale, l'“ *evidence-based medicine* ” doivent être pris pour nous, médecins, comme certains des éléments méthodologiques destinés à appréhender l'homme malade et ses traitements. Mais il y en a d'autres, ni meilleurs, ni moins bons, mais différents et complémentaires, et surtout indispensables tous ensemble pour une appréhension globale (holistique dans le sens noble du terme). L'“ *evidence-based medicine* ” a été effectivement fort insuffisamment prise en compte dans la définition des traitements ou des attitudes thérapeutiques. C'est très regrettable et il importe de lui rendre son importance justifiée et d'asseoir nos standards, *guidelines* et autres références médicales sur ses apports et à partir de sa méthodologie. Il importe aussi de former les étudiants aux méthodes rationnelles et critiques d'investigation et de connaissance, et de les (ré)apprendre aux médecins en exercice. Mais il ne s'agit pas de s'y arrêter. Le quantitatif ne donnera jamais que des chiffres, jamais des hypothèses explicatives, issues des recherches qualitatives (l'EBM n'est d'ailleurs pas limitée aux aspects quantitatifs). Celles-ci sont souvent d'ailleurs plus difficiles parce que, *a priori*, décriées par les chercheurs “ durs ” quantitatifs et parce que placées dans la nécessité de faire appel à des fondements théoriques complexes et transdisciplinaires. La recherche qualitative n'en est encore qu'à ses débuts en médecine expérimentale, mais ses apports promettent d'être très importants et des livres entiers lui sont aujourd'hui consacrés. Ce n'est qu'en rassemblant quantitatif et qualitatif que la connaissance sera plus globale, mais jamais entière parce que c'est impossible. De même, induction et déduction sont nécessairement complémentaires : l'induction théorise l'observation, et celle-ci cherche à falsifier les déductions de la théorie. Ni l'une ni l'autre n'est supérieure, il y a complémentarité épistémologique. Cette conception moderne (herméneutique de Ricoeur, triangle de Durkheim, cercle scientifique de De Groot) tord également le cou aux vieux débats dichotomiques : science hypothético-déductivo-nomologique (lois générales falsifiables) ou inductivo-statistique (crédibilité rationnelle dont se prévaut l'“ *evidence-based medicine* ”) ; les deux sont en effet nécessaires et complémentaires pour faire tourner le cercle dynamique de la connaissance. L'EBM ne relève pas intrinsèquement du second courant, mais elle pourrait en devenir une dérive. Certains auteurs le relèvent d'ailleurs très bien : Van Den Ende<sup>2</sup> (la stratégie

diagnostique du généraliste implique à la fois le probabilisme implicite, l'analyse contextuelle et la logique Bayésienne), Lamberts (diagnostiquer en médecine générale, c'est avant tout désigner et légitimer), Haynes<sup>3</sup> (modélisation de la décision médicale) et Sackett<sup>4</sup> (qui intègre sans les opposer “ *evidence-based medicine* ” et “ *experience-based medicine* ”). Tous sont chercheurs en science médicale, notamment Sackett dont le plus gros du travail d'épidémiologiste a été de traquer les biais sources d'erreur des études cliniques. Mais leurs conclusions sont avant tout synchrétiques, intégrant harmonieusement l'“ *evidence-based medicine* ” et d'autres approches plus qualitatives et/ou plus empiriques et/ou plus probabilistes.

L'“ *evidence-based medicine* ” a été définie par ses concepteurs comme l'utilisation consciencieuse et judicieuse des meilleures données actuelles de la recherche clinique dans la prise en charge personnalisée de chaque patient. Les quatre grandes étapes de la médecine basée sur la preuve sont :

- Le recueil des faits prouvés par la recherche clinique, en utilisant les méthodes scientifiques les plus adéquates dans chaque cas particulier pour donner à la preuve la plus grande validité possible.
- L'évaluation de la qualité méthodologique des publications.
- L'élaboration de stratégies combinant les données validées et les situations cliniques rencontrées par les médecins (importance fondamentale de l'expérience).
- L'application personnalisée de la preuve à chaque patient en tenant compte de ses caractéristiques particulières bio-médicales, psychologiques et sociales, de son adhésion à la proposition thérapeutique (“ *shared decision making*<sup>5</sup> ”) et de l'éthique professionnelle.

La lecture critique d'articles scientifiques rapportant les résultats d'études cliniques en constitue dès lors un pilier fondamental. Au cours de sa pratique, tout médecin peut y être confronté, dans le cadre d'une revue de littérature, lors d'une recherche sur un sujet déterminé, suite à la visite d'un délégué médical, pendant la lecture d'une revue scientifique, éventuellement via Internet ou lors de la consultation de la littérature tertiaire.

L'analyse d'une étude très importante publiée il y a peu, l'étude LIFE, servira d'exemple illustratif.

#### **Abstract**

***Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE) : a randomised trial against atenolol. Bjorn Dahlof, Richard B Devereux, Sverre E Kjeldsen, Stevo Julius, Gareth Beevers, Ulf de Faire, Frej Fyhrquist, Hans Ibsen, Krister Kristiansson, Ole LederbaUe-Pedersen, Lars H Lindholm, Markku S Nieminen, Per Omvik, Suzanne Oparil, Hans Wedel, for the LIFE study group. Lancet 2002 ; 359 : 995-1003.***

**Background** : Blood pressure reduction achieved with  $\beta$ -blockers and diuretics is the best recorded intervention to date for prevention of cardiovascular morbidity and death in patients with hypertension. Left ventricular hypertrophy (LVH) is a strong independent indicator of risk of cardiovascular morbidity and death. We aimed to establish whether selective blocking of angiotensin II improves LVH beyond reducing blood pressure and, consequently, reduces cardiovascular morbidity and death.

**Methods** : We did a double-masked, randomised, parallel group trial in 9193 participants aged 55-80 years with essential hypertension (sitting blood pressure 160-200/95-115 mmHg) and LVH ascertained by electrocardiography (ECG). We assigned participants once daily losartan-based or atenolol-based antihypertensive treatment for at least 4 years and until 1040 patients had a primary cardiovascular event (death, myocardial infarction, or stroke). We used Cox regression analysis to compare regimens.

**Findings** : Blood pressure fell by 30.2/16.6 (SD 18.5/10.1) and 29.1/16.8 mmHg (19.2/10.1) in the losartan and atenolol groups, respectively. The primary composite endpoint occurred in 508 losartan (23.8 per 1000 patient-years) and 588 atenolol patients (27.9 per 1000 patient-years; relative risk 0.87, 95 % CI 0.77-0.98,  $p = 0.021$ ). 204 losartan and 234 atenolol patients died from cardiovascular disease (0.89, 0.73-1.07,  $p = 0.206$ ) ; 232 and 309, respectively, had fatal or non-fatal stroke (0.75, 0.63-0.89,  $p = 0.001$ ) ; and myocardial infarction (non-fatal and fatal) occurred in 198 and 188, respectively (1.07, 0.88-1.31,  $p = 0.491$ ). New-onset diabetes was less frequent with losartan.

**Interpretation** : Losartan prevents more cardiovascular morbidity and death than atenolol for a similar reduction in blood pressure and is better tolerated. Losartan seems to confer benefits beyond reduction in blood pressure.

La masse d'informations que reçoit le médecin généraliste est telle que la première question qu'il doit se poser devant un article scientifique est de savoir s'il le lit ou non.

En dehors des " big fives " (*The New England Journal of Medicine*, *The Lancet*, *British Medical Journal* BMJ, *Journal of American Association JAMA*, *Annals of Internal Medicine*), revues reconnues de très haute qualité scientifique qui gagnent à être lues *a priori*, il existe des critères qui permettent de classer une revue dans la catégorie supérieure : pertinence du périodique (intérêt pour la médecine générale et intérêt individuel), politique éditoriale décrite dans les critères de Vancouver (rigueur et sérieux scientifique, comité de rédaction, comité de lecture scientifique - avec révision par les pairs -, exigences de forme, pas de biais de publication, critique des mauvais articles), indexation dans une banque de données reconnue (littérature secondaire ou tertiaire), indépendance vis-à-vis du secteur marchand, variété des articles proposés

(éditorial, articles de recherche, articles de revue, méta-analyses et revues systématiques, cas cliniques (*case reports*), *guidelines*, courrier des lecteurs).

**L'étude LIFE est publiée dans le *Lancet*, une revue parmi les *big fives*.**

Il est possible par le balayage rapide d'un article et par son analyse critique et orientée (titre, auteur, résumé, contexte) de décider de sa lecture ou pas. En particulier, l'évaluation d'un bon *abstract* structuré permet de préciser l'intervention (" *what* "), la population (" *who* "), le type d'étude (" *how* "), les résultats et les conclusions, notamment si ces derniers éléments répondent bien à la question clinique posée.

**L'article rapportant l'étude LIFE répond à ces conditions de qualité *a priori*.**

Il s'agit d'un article de recherche dont le plan global doit répondre au format standardisé : présentation générale IMRAD (introduction, méthodes, résultats, analyse, discussion), définition du problème, formulation de l'objectif de recherche, description de l'hypothèse de recherche, description de la population et des méthodes (type d'étude, instruments de mesure (validité), contrôles de qualité), résultats, méthodes statistiques utilisées (analyse des résultats en " *intention to treat* ", description des tests utilisés, statistiquement significatif :  $p < 0.05$  (ou 0.01), intervalle de confiance suffisamment étroit), clarté de la présentation des résultats, discussion et conclusions.

**La structure de l'article rapportant l'étude LIFE, article de recherche, est conforme à la présentation recommandée.**

Il existe de nombreux types d'études, dont plusieurs classifications sont possibles selon les critères choisis : le temps (études transversales ou longitudinales), la nature (études qualitatives, quantitatives, ou semi-quantitatives), les objectifs et le protocole. C'est cette dernière qui est la plus utilisée et elle distingue :

- Les études observationnelles (épidémiologiques) comparatives (avec un groupe contrôle) :
  - transversales descriptives (*cross sectional studies*), individuelles ou de groupe (communautaires, écologiques) : études de prévalence.
  - analytiques : rétrospectives cas-témoins (*case-control studies*) ou prospectives de cohortes : études d'incidence.
- Les études interventionnelles : essais cliniques contrôlés randomisés (*randomized controlled trials* (RCT)), utilisant un mécanisme purement aléatoire de définition des groupes. Les études quasi-expérimentales sont des essais cliniques ou communautaires non randomisés.

L'existence d'un groupe contrôle permet de tenir compte du phénomène statistique bien connu de régression vers la moyenne lors de la répétition des

mesures ; la randomisation, quant à elle, valide l'inférence statistique (les tests statistiques) et permet de conclure à une relation causale entre l'intervention thérapeutique et le bénéfice observé.

Ces deux éléments (groupe contrôle et randomisation) expliquent les différences de niveaux de preuve des différentes études. De manière simplifiée, le niveau I concerne les études contrôlées randomisées, le niveau II les études comparatives sans randomisation et le niveau III les séries descriptives sans contrôle de même que les avis d'experts. Les études avec groupe contrôle historique et les études rétrospectives sont d'un niveau moindre que les études avec groupe contrôle contemporain. Parmi les multiples classifications existantes, celle de l'AHCP (Agency for Health Care Policy and Research) est une des plus utilisées, notamment par le CEBAM (Centre Belge d'Evidence-Based Medicine) ou le BAPCOC (Belgian Antibiotic Policy Coordination Committee) :

- Niveau de preuve Ia : méta-analyse d'études contrôlées randomisées<sup>a</sup> ou études randomisées contrôlées.
- Niveau de preuve Ib : au moins une étude randomisée contrôlée.
- Niveau de preuve IIa : au moins une étude contrôlée de bonne qualité, mais sans randomisation.
- Niveau de preuve IIb : au moins une étude quasi expérimentale de bonne qualité d'un autre type.
- Niveau de preuve III : études descriptives de bonne qualité, études rétrospectives cas-témoins, études de corrélation.
- Niveau de preuve IV : rapports de comités d'experts, opinions ou expérience clinique d'autorités en la matière.

Cette gradation dans les niveaux de preuve implique la gradation parallèle dans les niveaux de recommandation pour la pratique clinique :

- Niveau de recommandation A (niveaux de preuve Ia et Ib) : recommandation clinique spécifique étayée et validée.
- Niveau de recommandation B (niveaux de preuve IIa, IIb, et III) : recommandation clinique non validée par RCT.
- Niveau de recommandation C (niveau de preuve IV) : pas de recommandation directement applicable en clinique.

**L'étude LIFE est une étude d'intervention randomisée (RCT), un des plus hauts niveaux de preuve EBM (Ib).**

Le plus souvent, une étude analyse une relation de causalité potentielle, entre un phénomène, un événement, un résultat investigué et un ou différents facteurs potentiellement étiologiques, un traitement, une exposition, un facteur de risque, une intervention.

**L'étude LIFE étudie la relation de causalité entre l'amélioration d'une série de paramètres cardiovasculaires et l'utilisation du losartan par rapport à l'aténolol. Son caractère randomisé**

**contrôlé assure que la relation statistique est bien une relation causale.**

Il faut souligner que les études non randomisées ne peuvent mettre en évidence que des relations statistiques entre les différents facteurs étudiés, et que c'est alors la logique qui transforme une relation statistique en relation causale en vérifiant le respect des principes de causalité<sup>6</sup>. Cinq critères doivent être satisfaits pour établir un lien de causalité dans une association statistique d'une étude non randomisée :

- L'intensité de l'association statistique (par exemple : plus le risque relatif est élevé, plus la causalité est probable).
- La séquence temporelle (antériorité de l'exposition par rapport à la maladie).
- La spécificité et l'exclusivité de l'association (un facteur de risque donné produit toujours une maladie donnée).
- La consistance de l'association (permanence, reproductibilité). Des études différentes doivent trouver une association similaire.
- La plausibilité biologique (la causalité doit être biologiquement acceptable).

A ces critères peuvent s'ajouter une preuve expérimentale (introduction/suppression de l'agent causal).

Par exemple, en ce qui concerne l'association entre un agent causal présumé (le tabagisme) et une maladie (le cancer du poumon), les conclusions d'une étude rétrospective cas-témoins pourraient être schématisées (Figure 1).

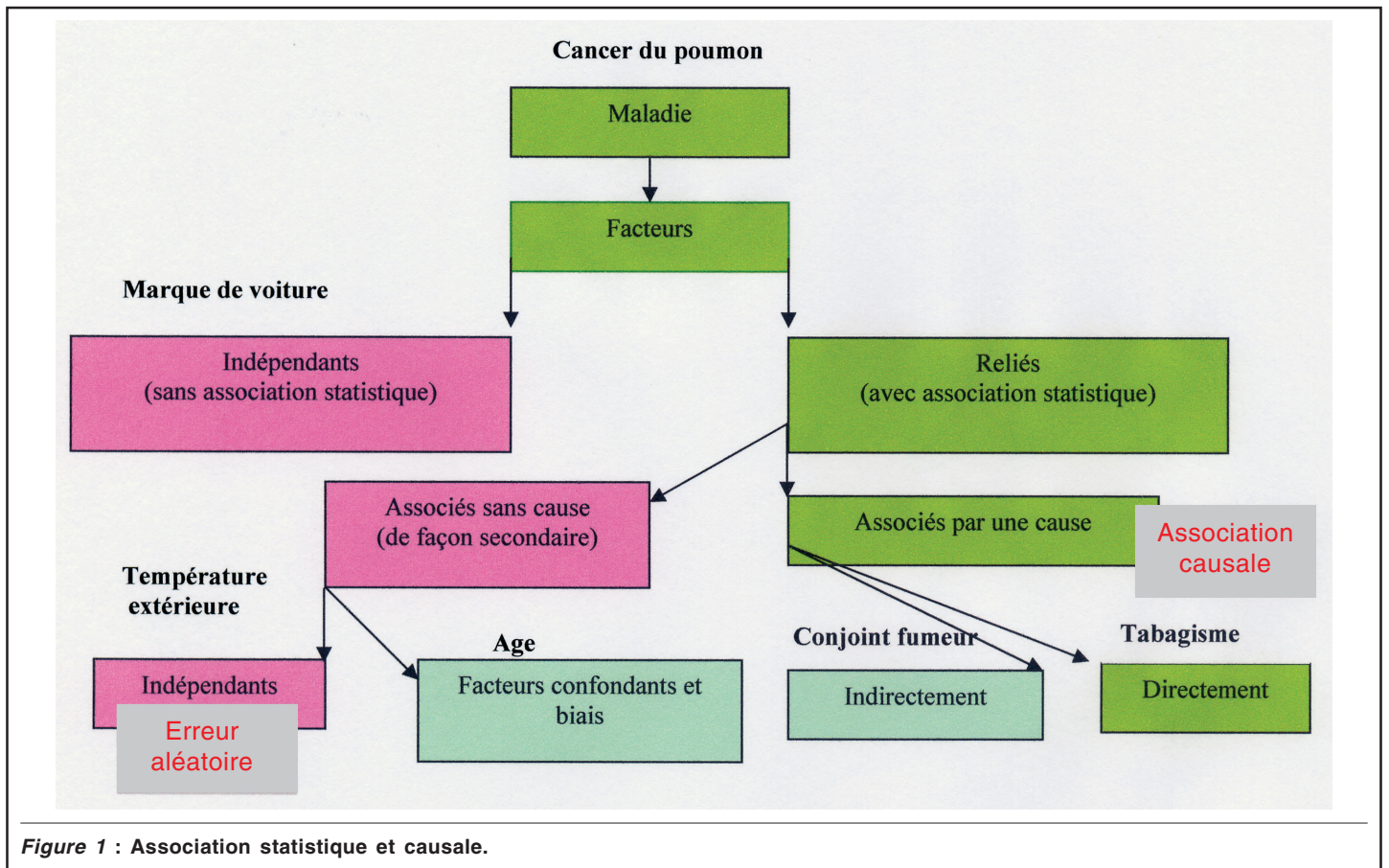
Une analyse multivariable fictive (Figure 1) montre les relations possibles entre la variable étudiée (variable expliquée, placée sur l'axe y), le cancer pulmonaire, et les variables potentiellement causales (variables explicatives, placées sur l'axe des x), la marque de voiture, la température extérieure moyenne, l'âge, le tabagisme passif et le tabagisme actif.

**Dans l'étude LIFE, les variables étudiées sont la fréquence des événements cardiovasculaires apparus, les variables potentiellement causales, soit le traitement aténolol, soit le traitement losartan, et son caractère randomisé en assure l'inférence statistique.**

Les Tableaux 1 et 2 schématisent les études rétrospectives cas-témoins, prospectives de cohorte et transversales et les études d'intervention expérimentale.

<sup>a</sup> La méta-analyse est une démarche plus qu'une simple technique, qui a pour but de combiner les résultats de plusieurs essais thérapeutiques pour en faire une synthèse reproductible et quantifiée. Cette synthèse produit un gain de puissance statistique dans la recherche de l'effet d'un traitement, une précision optimale dans l'estimation de la taille de l'effet et permet en cas de résultats apparemment discordants d'obtenir une vue globale de la situation. Une méta-analyse est précédée d'une revue systématique qualitative de la littérature selon une méthodologie bien précise de sélection des articles pertinents.





**Tableau 1 : Etude rétrospective cas-témoins, prospective de cohorte et transversale.**

	Maladie présente M+	Maladie absente M-	
Facteur de risque présent FR+	A	B	A+B
Facteur de risque absent FR-	C	D	C+D
	A+C	B+D	A+B+C+D

**Tableau 2 : Etude d'intervention expérimentale.**

	Maladie présente M+	Maladie absente M-	
Intervention à investiguer T+	A	B	A+B
Intervention contrôle T-	C	D	C+D
	A+C	B+D	A+B+C+D

Les trois types d'étude sont schématisés par le même tableau, mais différent par la direction et le point de départ de la flèche du temps dans le Tableau. Pour les études rétrospectives, on étudie des cas (M+) appariés à des témoins (M-) dont on retrace l'histoire vers le passé par rapport à l'exposition à un ou à des facteurs de risque ; pour les études prospectives, on compare vers le futur l'apparition de cas dans deux cohortes, l'une exposée, l'autre pas, à un ou à des facteurs de risque ; pour les études transversales on met en relation à un moment donné le facteur de risque et la présence de la maladie.

sont randomisées, sont toujours prospectives comme l'étude LIFE.

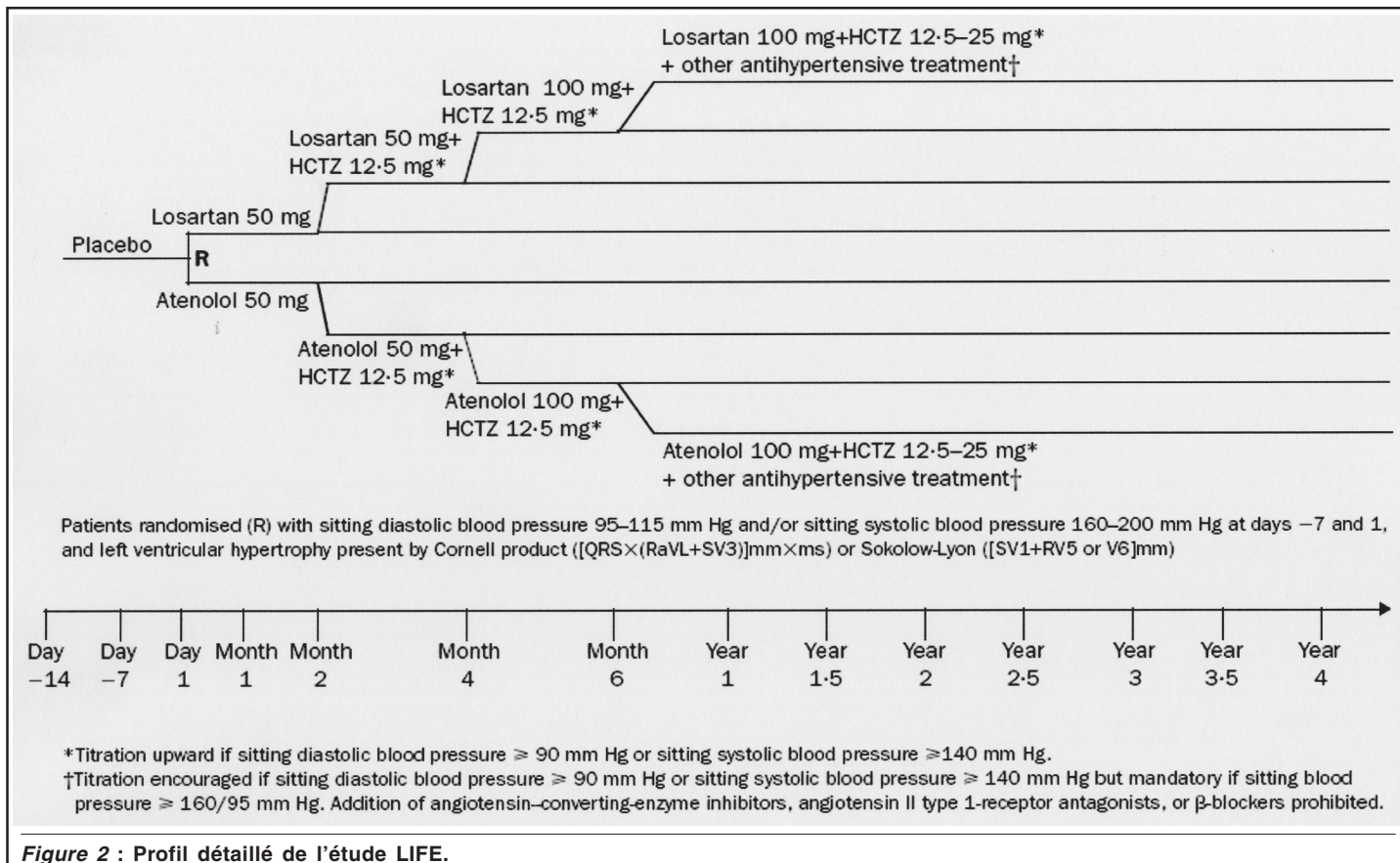
Une étude randomisée contrôlée (RCT *randomized controlled trial*) est une étude prospective qui compare un groupe (une cohorte) intervention et un groupe (une cohorte) contrôle de référence (éventuellement placebo), dont la constitution s'est faite aléatoirement. Elle est dite en double aveugle (*double blind*) si l'appartenance d'un patient à l'une ou l'autre des deux cohortes n'est connue ni du patient lui-même, ni de l'investigateur.

Les études interventionnelles, *a fortiori* si elles

Le profil de l'étude LIFE, étude d'intervention

**Tableau 3 : Profil global de l'étude LIFE.**

	Evénements cardiovasculaires présents	Evénements cardiovasculaires absents	
Intervention à investiguer (losartan)	A	B	A+B
Traitement de contrôle (aténolol)	C	D	C+D
	A+C	B+D	A+B+C+D



**Figure 2 : Profil détaillé de l'étude LIFE.**

randomisée contrôlée, est repris au Tableau 3 (Figure 2).

De plus en plus, les études sont soumises à des conditions éthiques extrêmement strictes. L'étude LIFE les respecte : un contrôle (l'aténolol) déjà démontré efficace (dans ce cas, comparer le losartan à un placebo serait éthiquement inacceptable), la sollicitation de l'avis d'un comité d'éthique et la notification des conflits d'intérêt (financement par la firme Merck).

Cependant le choix du comparateur est discutable : pourquoi l'aténolol et pas un inhibiteur de l'enzyme de conversion (IEC) ? En effet, les IEC ont une efficacité démontrée sur l'évolution péjorative de l'hypertrophie ventriculaire gauche, ils sont plus efficaces que les comparateurs classiques<sup>7,8</sup> dans les situations de post-infarctus, d'insuffisance cardiaque, d'hypertension ou d'antécédent d'accident vasculaire cérébral, et leurs associations à d'autres médicaments sont sûres et bien connues (dans LIFE, moins de 15 % des patients recevaient une monothérapie). Mais les résultats attendus étaient certainement meilleurs en comparaison avec l'aténolol qu'avec un IEC.

**L'étude LIFE est bien construite, son profil (design) est rigoureux, et elle est éthiquement correcte ; le choix du comparateur aténolol est discutable.**

Les critères d'inclusion et d'exclusion sont fondamentaux. C'est leur définition et leur prise en compte précise (proportion de patients respectant le protocole) qui permettent ou non la généralisation des résultats de l'étude, c'est-à-dire leur possible application à un patient de notre population, rencontré dans notre pratique clinique quotidienne. La dernière partie de la définition de l'EBM consiste d'ailleurs en : " L'application personnalisée de la preuve à chaque patient en tenant compte de ses caractéristiques ". Dans l'étude LIFE :

- Critères d'inclusion : 55-80 ans, HTA traitée ou non traitée, HVG (score de Cornell > 2.440 mm\*ms).
- Critère d'exclusion : HTA secondaire, infarctus myocardique ou AVC endéans les 6 mois, angor traité par  $\beta$ -bloquant ou anti-calcique, insuffisance cardiaque ou FEV  $\leq 40$  %, situation clinique nécessitant  $\beta$ -bloquant, hydrochlorothiazide, antagoniste angiotensine, IEC.



En analysant plus avant le protocole d'étude, on remarque les faits significatifs suivants :

- Les patients sont inclus uniquement à partir de pratiques hospitalières ; ils sont fort âgés (âge moyen 67 ans), ils présentent une HTA sévère (moyenne = 175/95 mmHg) et un score élevé d'HVG. Ils représentent 25 % des hypertendus dans LIFE, moins de 10 % en médecine générale (6 % dans une étude de De Cort<sup>9</sup>), impliquant que seulement 11 % des patients dans LIFE reçoivent une monothérapie.
- Les critères d'inclusion ont été modifiés en cours d'étude (majoration du score de Cornell et introduction d'un indice de Sokolov > 38 mm), sans aucune explication, mais ce changement visait très probablement à accentuer l'importance (l'impact) de la différence observée en augmentant le risque.

**Etudier une population très sélectionnée et à haut risque est un des moyens classiques pour atteindre plus rapidement des résultats significatifs (mais évidemment limités à cette seule population). Modifier les critères d'inclusion ou d'exclusion en cours d'étude est strictement interdit.**

Un des biais les plus importants en recherche clinique est le biais de transfert, qui consiste à ne tenir compte, pour l'analyse des résultats, que des patients arrivés au bout de l'étude, en ignorant ceux qui, après la randomisation, refusent de continuer l'étude à un moment ou à un autre (" *drop out* "), les perdus de vue (" *lost to follow-up* "). Dans cette situation, la moins mauvaise manière de rapporter les résultats est de calculer les taux de réponse en divisant le nombre d'événements observés par le nombre de patients initialement randomisés : il s'agit de l'analyse en intention de traiter (" *intention to treat* ") qui donne évidemment de moins bons résultats que l'analyse par protocole (" *per protocol* ") qui, elle, ne tient compte au dénominateur que des patients qui ont terminé l'étude.

Un autre biais important et difficile à évaluer est le biais de performance, c'est-à-dire de l'observance de la posologie administrée (compliance du patient à son traitement).

Ces deux biais (transfert et performance) sont parfois repris sous le terme de biais de suivi. La randomisation a pour but de minimiser l'influence de ces biais sur la comparaison des deux traitements en les répartissant de manière équilibrée entre les deux groupes.

D'autres biais peuvent introduire une distorsion dans le recrutement des patients avant la randomisation : patients éligibles mais non inclus, patients refusant de participer. Pour évaluer l'importance de cette distorsion, il est important de disposer d'un diagramme (" *flow chart* ") reprenant la séquence de recrutement et de suivi des patients tout au long de l'étude (Figure 3).

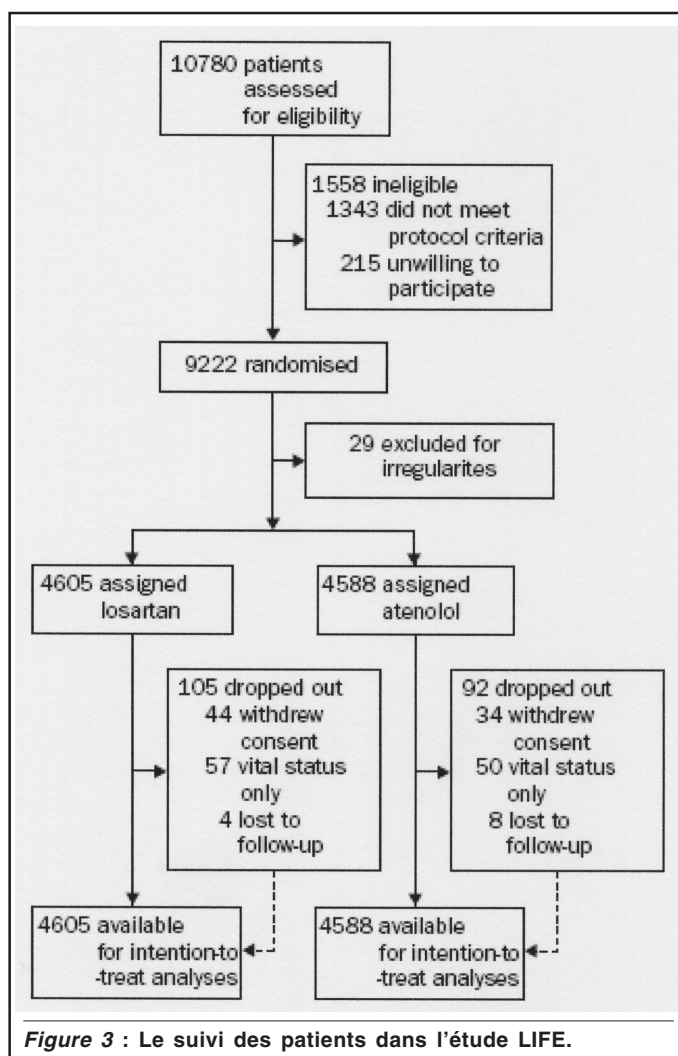


Figure 3 : Le suivi des patients dans l'étude LIFE.

**L'étude LIFE respecte bien les conditions de l'analyse en " intention de traiter ".**

D'autres biais fréquents dans les études cliniques sont le biais de sélection (le choix du traitement n'est pas aléatoire) et le biais de susceptibilité (les groupes constitués par randomisation ne présentent pas le même degré de sévérité). Ces biais sont évités par une bonne technique de randomisation (méthode non prédictible). Néanmoins, le biais de sélection peut être le résultat d'une erreur dans la randomisation (le traitement effectivement donné n'est pas celui attribué par la randomisation).

La randomisation a pour but d'équilibrer les groupes pour tous les facteurs connus et inconnus et rend les groupes similaires au départ mais généralement non représentatifs de la population des patients présentant la maladie étudiée. Elle permet de répartir uniformément les facteurs confondants entre les groupes (voir ci-dessous).

Parfois, on observe un biais accidentel, c'est-à-dire un déséquilibre entre les groupes pour un facteur important pour le pronostic des patients lié au hasard seul. Ce biais peut être corrigé *a posteriori* en utilisant des techniques d'ajustement.

La manière de vérifier l'absence de biais de sélection, de susceptibilité et accidentel est de présenter un tableau comparatif des groupes constitués par randomisation avant l'administration des traitements (Figure 4).

L'évaluation des résultats en double insu (" *double blind* ") permet une mesure objective des effets observés sans distorsion liée à la subjectivité du médecin et du patient : elle évite le biais de détection. Le double aveugle permet également de garantir la stricte observance de la randomisation : dissimulation ou secret de l'attribution (" *concealment of allocation* ").

En conclusion : la validité interne est garantie lorsque tous les biais sont contrôlés par la randomisation, que l'évaluation des résultats est en double aveugle et l'analyse en intention de traiter ; les critères d'inclusion et d'exclusion sont définis et précis pour contribuer à assurer la validité externe, c'est-à-dire la généralisabilité des résultats ; le consentement éclairé (" *informed consent* ") du patient est impératif, la taille des cohortes calculée avant le début de l'étude, les méthodes statistiques précisées, de même que les seuils de signification, et que, plus globalement, tout le protocole de recherche est publié préalablement, sans analyse ou stratification *a posteriori* (" *post hoc* " ou " *ex post* ") ou " essorage " des données.

**Dans l'étude LIFE, étude de supériorité, la randomisation est tout à fait correcte, les méthodes statistiques sont précisées *a priori* et clairement définies<sup>b</sup>.**

Il existe en fait différents types d'erreurs qui réduisent la qualité des études cliniques :

- L'erreur aléatoire (voir plus loin), inévitable mais quantifiable, liée aux fluctuations d'échantillonnage : deux échantillons successifs ne sont jamais identiques et peuvent donc induire des conclusions différentes et même contradictoires. Elle diminue avec l'augmentation de la taille des groupes étudiés. Pour les études non significatives, elle diminue avec l'augmentation de la puissance de l'étude.
- Les variables confondantes (facteurs de confusion), associées aussi bien à la variable indépendante (facteur causal vrai) qu'à la variable dépendante (maladie). Les variables confondantes ne sont pas des biais, elles représentent une bonne caractéristique du risque de la maladie, et sont présentes dans l'étude par la nature même des unités expérimentales. Par exemple, le cancer pulmonaire est associé causalement au tabagisme ; mais le nombre de cigarettes fumées augmente évidemment avec l'âge du fumeur, induisant une relation statistique secondaire dans laquelle l'âge est une variable confondante<sup>c</sup>.

Il est évident que le problème des variables confondantes concerne essentiellement les études épidémiologiques, la présence d'un groupe contrôle et la randomisation dans les études d'intervention permettant théoriquement d'éviter leur influence sur les résultats (mais pas toujours : cf. ci-dessous

l'exemple de l'étude CHARM).

- Les biais ou erreurs systématiques (dont la plupart viennent d'être décrits) : biais de sélection (" *selection bias* "), biais de susceptibilité (" *susceptibility bias* "), biais de détection (" *detection* ", " *information* ", " *measurement* ", ou " *assessment bias* "), biais de performance (" *performance bias* ", " *compliance bias* "), biais de transfert (" *transfert bias* ") et biais accidentel (" *accidental bias* ").

Dans les méta-analyses, s'ajoute le biais de sélection des études qui vont être regroupées ainsi que le biais de publication. Le biais de publication (40 % des études non significatives ou négatives ne sont jamais publiées) est aujourd'hui recherché systématiquement. Plusieurs méthodes quantitatives ont été mises au point pour le mettre en évidence (diagramme en entonnoir ou " *funnel plot* "). Récemment aussi, des registres des RCT en cours ont été mis en place, pour assurer leur publication, de manière complète et exclusive (comme par exemple le registre régional des tumeurs en Monterégie au Canada<sup>10</sup>).

Le biais de performance (ou biais d'observance) suggère que l'observance est souvent tellement variable qu'elle peut devenir facteur de confusion. L'étude CHARM<sup>11</sup>, qui évalue l'effet du candesartan dans l'insuffisance cardiaque, le met particulièrement bien en évidence : globalement les résultats sont en faveur du candesartan (réduction du risque de 10 %), mais tous les patients (candesartan ou pas) dont la compliance définie selon des critères objectifs dépasse 80 %, montrent une réduction de risque de 35 %.

Le biais d'observation est lié à l'observateur et est surtout présent dans les études observationnelles non randomisées : contamination précoce et tardive des témoins (" effet Hawthorne " : cet effet est inclus dans l'effet placebo), confusion incidence-prévalence (paralogisme de Neyman ou étude trop tardive des

<sup>b</sup> Du point de vue statistique, la détection d'une différence d'au moins 15 % entre la cohorte traitement (losartan) et la cohorte contrôle (aténolol), avec une puissance de 80 %, et un niveau significatif de 5 % bilatéral, nécessite l'inclusion de 8.300 patients, sur la projection d'un taux d'incidence de l'*endpoint* primaire de 15 % dans le groupe aténolol et de 12,75 % dans le groupe losartan, c'est-à-dire de 1.040 en 4 ans dans chacune des cohortes.

<sup>c</sup> Il existe plusieurs méthodes pour " neutraliser " les facteurs confondants :

- Appariement des données : grouper les sujets étudiés de telle façon que ceux d'un même groupe possèdent les mêmes facteurs de confusion (répartition égale dans les différents bras de l'étude).
- Stratification des données : choisir un échantillon de malades et un échantillon de témoins, puis former des classes de sujets des 2 échantillons (stratifier) par rapport aux facteurs de confusion (méthode statistique classique de Mantel-Haentzel).
- Standardisation des taux : rendre artificiellement comparables des séries d'observations, soit par standardisation directe (appliquer à une population de référence les taux, par groupe d'âge par exemple, de chacune des populations concernées : taux standardisés), soit par standardisation indirecte (appliquer à chacune des populations à comparer, par groupes d'âge par exemple, les taux d'une population de référence).
- Contrôle de ceux-ci par ajustement statistique.



	Losartan (n=4605)	Atenolol (n=4588)	All (n=9193)
<b>Demographic and clinical characteristics</b>			
Age (years)*	66.9 (7.0)	66.9 (7.0)	66.9 (7.0)
Women	2487 (54%)	2476 (54%)	4963 (54%)
Ethnic origin			
White	4258 (92%)	4245 (93%)	8503 (92%)
Black	270 (6%)	263 (6%)	533 (6%)
Hispanic	47 (1%)	53 (1%)	100 (1%)
Asian	25 (0.5%)	18 (0.4%)	43 (0.5%)
Other	5 (0.1%)	9 (0.2%)	14 (0.2%)
Blood pressure (mm Hg)*			
Systolic	174.3 (14.2)	174.5 (14.4)	174.4 (14.3)
Diastolic	97.9 (8.8)	97.7 (9.0)	97.8 (8.9)
Heart rate (bpm)*	73.9 (11.0)	73.7 (11.2)	73.8 (11.1)
BMI (kg/m <sup>2</sup> )*	28.0 (4.8)	28.0 (4.8)	28.0 (4.8)
Cornell voltage-duration product (mm×ms)*	2834.4 (1065.4)	2824.1 (1033.3)	2828.8 (1049.5)
Sokolow-Lyon (mm)*	30.0 (10.6)	30.1 (10.4)	30.0 (10.5)
Framingham risk score*	0.223 (0.095)	0.225 (0.096)	0.224 (0.096)
Current smokers	729 (16%)	770 (17%)	1499 (16%)
<b>Medical history</b>			
Any vascular disease	1203 (26%)	1104 (24%)	2307 (25%)
Coronary heart disease	771 (17%)	698 (15%)	1469 (16%)
Cerebrovascular disease	369 (8%)	359 (8%)	728 (8%)
Peripheral vascular disease	276 (6%)	244 (5%)	520 (6%)
Atrial fibrillation	150 (3%)	174 (4%)	324 (4%)
Isolated systolic hypertension†	660 (14%)	666 (15%)	1326 (14%)
Diabetes	586 (13%)	609 (13%)	1195 (13%)

Bpm=beats per minute. BMI=body mass index. Data are number (%) unless otherwise indicated. \*Data are mean (SD). †Definition  $\geq 160 / < 90$  mm Hg.

Table 1: Baseline characteristics

Figure 4 : La répartition des patients dans l'étude LIFE.

sujets exposés), biais du " voyant ", méfiance ou servilité des participants, problèmes dans la codification, l'enregistrement (le double encodage des données permet de minimiser ce problème) ou l'interprétation des données.

L'étude LIFE prévoit toute une série de procédures destinées à diminuer au maximum le biais d'observation :

- Le *follow-up* régulier des patients est assuré pendant au moins 4 ans (consultations régulières, adaptation du traitement, relevé des événements, etc.).
- Les ECG (*screening*, départ, suivi, fin de l'étude) sont tous rassemblés et lus dans une structure centrale de référence.
- Tous les événements cardiovasculaires sont revus et évalués par un comité de classification des critères de jugement (" *endpoints* ") composé de deux cliniciens indépendants, et ré-analysés en cas de désaccord.

**La rigueur des procédures dans l'étude LIFE tend à minimiser le plus possible les erreurs d'observation et d'évaluation des critères de jugement.**

Le biais d'information, lié aux tests pratiqués ou aux instruments de mesure, hypothèque la validité interne et la reproductibilité (les mêmes mesures donnent les mêmes résultats). Il concerne le choix des variables à étudier (les critères de jugement ou " *endpoints* ") et les tests ou instruments utilisés pour le faire.

Toutes les études nécessitent en effet l'utilisation de tests ou d'instruments pour mesurer les critères de

jugement (Tableau 4). Aucun de ces moyens n'est totalement fiable, et plusieurs erreurs différentes sont possibles : il s'agit d'une décision ou d'une évaluation optimale dans l'incertitude. Se pose aussi l'existence d'une référence absolue, d'une mesure parfaite, virtuelle par nature (le " *gold standard* ").

Les caractéristiques classiques d'un test sont sa sensibilité (proportion de tests positifs parmi la sous-population atteinte) et sa spécificité (proportion de tests négatifs parmi la sous-population indemne). Un test à haute sensibilité ne laissera échapper que peu de faux négatifs ; un test à haute spécificité ne diagnostiquera que peu de faux positifs.

Deux autres paramètres intéressent aussi beaucoup le clinicien : la valeur prédictive positive VPP ou VPP+ (proportion de patients atteints parmi la sous-population des tests positifs : combien de chance un patient a-t-il d'être vraiment malade si son test est positif ?) et la valeur prédictive négative VPN ou VPP- (proportion de patients indemnes parmi la sous-population des tests négatifs : combien de chance un patient a-t-il d'être vraiment indemne si son test est négatif ?).

La force probante correspond au rapport de vraisemblance positif (*positive likelihood ratio*, LR+), la force excluante au rapport de vraisemblance négatif (*negative likelihood ratio*, LR-). Certains appellent force excluante l'inverse du LR- pour des raisons pédagogiques : l'ordre de valeur est alors le même que pour la force probante LR+ (plus grande est la valeur, plus grande est la force). Les rapports de vraisemblance estiment le rapport entre la probabilité

**Tableau 4 : Les différents paramètres d'un test.**

	Maladie présente M+	Maladie absente M-	Total
Test positif T+	A	B	A+B
Test négatif T-	C	D	C+D
Total	A+C	B+D	A+B+C+D
Prévalence P = A+C/A+B+C+D			
Vrais positifs VP = A Vrais négatifs VN = D Sensibilité Se = A/A+C Valeur prédictive positive VPP = A/A+B Likelihood ratio pos. LR+ = (A/A+C)/(B/B+D) = Se/1 - Sp (rapport de vraisemblance positif)		Faux positifs FP = B Faux négatifs FN = C Spécificité Sp = D/B+D Valeur prédictive négative VPN = D/C+D	
Likelihood ratio neg. LR- = (D/B+D)/(C/A+C) = 1 - Se/Sp (rapport de vraisemblance négatif)			
Diagnostic odds ratio (rapport de cotes) DOR = AD/BC = (VP/FN)/(FP/VN)			
DOR = [Se/(1 - Sp)]/[(1 - Se)/Sp] = LR+/LR-			
Exactitude (proportion de classifications correctes) = (A+D)/(A+B+C+D)			

d'avoir un test positif (ou négatif) chez les sujets malades à celle d'avoir un test positif (ou négatif) chez les sujets sains.

Le rapport des cotes diagnostiques (" *diagnosis odds ratio* ") cherche à quantifier la performance d'un test par une seule valeur qui ne soit pas influencée par la prévalence. Il correspond au ratio entre la cote (*odds*) d'être malade (probabilité d'être malade divisée par la probabilité de ne pas être malade) lorsque le test est positif et la cote de ne pas être malade (probabilité de ne pas être malade divisée par la probabilité d'être malade) lorsque le test est négatif. L'*odds ratio* est connu comme un indice statistique dans les études épidémiologiques où il représente la force de l'association entre le facteur de risque et la maladie. Ici, il pourrait être utilisé pour montrer la force de l'association entre le résultat d'un test et la présence de la maladie.

Enfin, l'exactitude est définie comme le rapport entre le nombre de résultats exacts et l'ensemble des résultats ; de son côté, la précision dépend de la variance.

**Dans l'étude LIFE, les tests et les moyens de mesure utilisés pour identifier l'incidence des critères de jugement (décès, infarctus myocardiques, AVC, hospitalisations, etc.) sont simples, bien validés, sensibles et spécifiques (leurs VPP, VPN, LR+ et 1/LR- sont élevés).**

Les critères de jugements, toujours exprimés en intention de traiter, se divisent en primaires ou secondaires, directs ou intermédiaires (" *surrogate* "), univoques ou composites ; ce sont des indicateurs de résultats, définis complètement et préalablement à l'étude dans le protocole.

Dans l'étude LIFE, les critères de jugement sont les suivants :

- Critère de jugement primaire composite : mortalité cardiovasculaire + accident vasculaire non fatal + infarctus myocardique non fatal.

- Critères de jugement secondaires univoques : mortalité totale, admission à l'hôpital pour angor ou insuffisance cardiaque, revascularisation coronaire, arrêt cardiaque réanimé, apparition d'un diabète.

**L'étude LIFE définit précisément et préalablement un ensemble de critères de jugement primaire et secondaires.**

Les études d'intervention sont évaluées à l'aide d'une série de paramètres (Tableau 3) :

- Risque absolu : RA (losartan) = A/A+B, RA (aténolol) = C/C+D (nombre d'événements survenus sur le total des patients traités). A noter que, *stricto sensu*, le risque absolu est le risque de base non réductible par le traitement, et correspond en général au groupe témoin lorsqu'il n'y a aucune intervention thérapeutique. Dans l'étude LIFE, le groupe contrôle n'est pas " placebo " mais " aténolol ".
- Risque relatif (*risk ratio* RR, *hazard ratio* HR) :  $RR = (A/A+B)/(C/C+D)$  (rapport entre le risque de la cohorte intervention et celui de la cohorte contrôle).
- Cote, rapport de cotes (*odds ratio*) :  $OR = ((A/A+B)/(B/A+B))/((C/C+D)/(D/C+D)) = (A/B)/(C/D) = AD/BC$  (approximation du RR).
- Réduction du risque absolu (bénéfice attribuable) :  $RRA = (A/A+B)-(C/C+D)$ . Dans la mesure où, dans l'étude LIFE, il n'y a pas de groupe placebo, on ne peut pas vraiment parler de réduction du risque absolu mais plutôt de différence de risque. Néanmoins les chiffres obtenus dans un cas ou dans l'autre sont identiques.
- *Number needed to treat*, NNT (ou nombre de sujets à traiter, NST) =  $1/RRA$  (nombre de patients à traiter dans la cohorte intervention pour éviter un événement en plus pendant la durée de l'étude par rapport à la cohorte contrôle).
- *Number needed to harm*, NNH (ou nombre nécessaire pour nuire NNN).
- Le rapport NNH/NNT exprime un rapport bénéfice/risque pour le traitement par rapport au contrôle.
- Réduction du risque relatif :  $RRR = 1-RR$ .

Les paramètres choisis pour présenter les



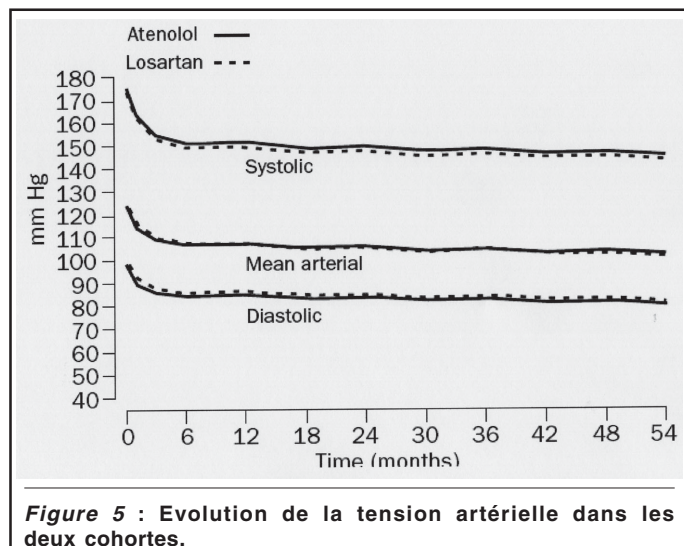
résultats et la manière de le faire ne sont pas indifférents, montrant bien que le quantitatif peut aussi se teinter de qualitatif. Un éditorial déjà ancien de Sackett<sup>12</sup> dans le BMJ le démontre bien, de même qu'une enquête<sup>13</sup> pratiquée chez 309 patients anglais, 102 hypertendus, 207 normotendus, entre 35 et 65 ans. Les réponses suivantes (identiques pour les hyper- et les normotendus, sans influence de l'âge ni du sexe) sont données à des questions identiques mais posées de manière différente. Prendriez-vous un médicament qui :

- réduit le risque d'AVC de 45 % ? (réduction risque relatif RRR) : oui pour 92 % des répondants.
- ramène le risque d'AVC de 1/400 à 1/700 ? (réduction risque absolu) : oui pour 75 %.
- permet d'éviter 1 AVC chez 1 patient parmi 35 qui prennent le médicament pendant 25 ans ? (nombre de sujets à traiter NST) : oui pour 68 %.
- permet, au cours des 25 prochaines années, de vous éviter 1 AVC dans 3 % des cas, et qui n'a pas d'effet dans 97 % ? (bénéfice personnel probable) : oui pour 44 %.

**L'étude LIFE présente les résultats en réduction des risques relatifs plutôt qu'en réduction des risques absolus ou en nombres nécessaires à traiter, ce qui est une méthode classique pour augmenter l'impression d'efficacité du traitement investigué.**

Durant tout le déroulement de l'étude, il n'y a aucune différence de tension artérielle entre les 2 cohortes, d'intervention (losartan) et de contrôle (aténolol) (Figure 5).

Les études d'intervention observent généralement la même méthodologie globale (en tout cas pour les études de supériorité, non pour les études d'équivalence ou de non infériorité, de plus en plus nombreuses) : utiliser un test statistique approprié pour valider l'hypothèse que l'intervention étudiée (ici le losartan) est plus efficace que l'intervention de contrôle (ici l'aténolol) sur une série de critères de jugement prédéfinis.



Il s'agit donc de tester une hypothèse de base appelée hypothèse nulle (H0) et de quantifier l'intervention du hasard pour expliquer une différence égale ou supérieure à celle observée, la règle de décision consistant à rejeter l'hypothèse nulle lorsque le hasard intervient peu ( $p < 0.05$ ) (Tableau 5).

Dans toutes les statistiques, l'hypothèse nulle est celle pour laquelle on contrôle la probabilité de rejet à tort. Cette probabilité est appelée risque  $\alpha$  ou risque de 1<sup>ère</sup> espèce (le seuil du test). Par exemple dans ce cas de l'étude LIFE, l'hypothèse nulle H0 est qu'il n'y a pas de différence entre les 2 groupes (losartan et aténolol). Rejeter H0 (c'est-à-dire, accepter l'hypothèse alternative H1) consiste à accepter qu'il y a une réelle différence (donc que le losartan est plus efficace que l'aténolol), avec une chance minimum (valeur de p) que cette différence soit due au hasard. L'intervalle de confiance à 95 % est une autre manière de présenter les résultats en y ajoutant la précision de la mesure. Il délimite la fourchette de valeurs dans laquelle se situe le résultat.

A l'inverse, le risque  $\beta$  ou risque de 2<sup>ème</sup> espèce correspond au non rejet, à tort, de l'hypothèse nulle, c'est-à-dire à la probabilité de ne pas mettre en évidence une différence qui existe.

La puissance de l'étude, enfin, est égale à  $1-\beta$ .

**Les statistiques choisies, utilisées et calculées dans l'étude LIFE sont tout à fait correctes.**

Les résultats de l'étude LIFE sont présentés dans l'article du *Lancet* de la manière reprise au Tableau 6 (NB : le cas particulier du diabète n'est pas envisagé dans le présent article).

Comme signalé plus haut, ces résultats sont présentés en réduction des risques relatifs. Recalculés en réduction des risques absolus et en nombres nécessaires à traiter, ils se présenteraient alors de la manière reprise au Tableau 7.

L'analyse des résultats montre qu'il n'y a pas de différence statistiquement significative entre les 2 cohortes losartan et aténolol en ce qui concerne la mortalité cardiovasculaire, l'infarctus myocardique non fatal, la mortalité totale, la fréquence d'hospitalisation et les revascularisations. Par contre, les 2 bras de l'étude sont différents en ce qui concerne l'AVC, et surtout le critère combiné primaire, la réduction de risque relatif pour ce critère atteignant 13 %. C'est ce résultat (Tableau 6) qui a été rapporté largement dans le public médical, les congrès et les revues scientifiques, la conclusion de l'article étant la suivante : " *our results are directly applicable in clinical practice and should affect future guidelines* ".

Les mêmes résultats, mais présentés selon d'autres paramètres d'évaluation (Tableau 7 : Présentation alternative des résultats de l'étude LIFE), montrent que la réduction du risque relatif d'AVC de 25 % correspond à un nombre de patients à traiter



**Tableau 5 : La statistique de test.**

	Hypothèse H0 fautive : groupes différents (H1)	Hypothèse H0 vraie : groupes identiques (H0)
Rejet de l'hypothèse nulle H0	1 - $\beta$ (puissance)	Erreur $\alpha$ (type 1)
Non rejet de l'hypothèse nulle H0	Erreur $\beta$ (type 2)	1 - $\alpha$

**Tableau 6 : Résultats de l'étude LIFE.**

	Losartan (n = 4.605)	Aténolol (n = 4.588)	HR corrigé (IC 95 % et valeur p)
<b>Critères primaires</b>			
Critère combiné composite (utilisé précédemment)	508 (11 %)	588 (13 %)	0.87 (0.77-0.98, p = 0.021)
- dont mortalité cardiovasculaire	204 (4 %)	234 (5 %)	0.89 (0.73-1.07, p = 0.206)
- dont AVC non fatal	232 (5 %)	309 (7 %)	0.75 (0.63-0.89, p = 0.001)
- dont infarctus myocardique non fatal	198 (4 %)	188 (4 %)	1.07 (0.88-1.31, p = 0.491)
<b>Critères secondaires</b>			
Mortalité totale	383 (8 %)	431 (9 %)	0.90 (0.78-1.03, p = 0.128)
Survenue de diabète (non envisagé dans l'article)	241 (6 %)	319 (8 %)	0.75 (0.63-0.88, p = 0.001)

**Tableau 7 : Présentation alternative des résultats de l'étude LIFE.**

<b>Calcul du risque d'incidence d'AVC non fatal</b>		
RA losartan	232/4.605 = 0.05	5 %
RA atémolol	309/4.588 = 0.067	6,7 %
RR losartan/atémolol	0.05/0.067 = 0.75	75 % (RRR = 25 %)
RRA losartan/atémolol	0.067-0.05 = 0.017	1,7 %
NNT losartan/atémolol	1/0.017 = 59	59 patients pendant 4 ans
<b>Calcul du risque d'incidence du critère combiné primaire</b>		
RA losartan	508/4.605 = 0.11	11 %
RA atémolol	580/4.588 = 0.126	12,6 %
RR losartan/atémolol	0.11/0.126 = 0.87	87 % (RRR = 13 %)
RRA losartan/atémolol	0.126-0.11 = 0.016	1,6 %
NNT losartan/atémolol	1/0.016 = 2	62 patients pendant 4 ans

(NNT) avec du losartan plutôt qu'avec de l'atémolol de 59 pendant 4 ans pour éviter 1 AVC. Pour le critère combiné primaire, la RRR est de 13 % et le NNT pendant 4 ans de 62. Et ces résultats concernent, comme souligné plus haut, moins de 10 % des patients hypertendus en médecine générale.

Construire un critère de jugement (*endpoint*) composite potentiellement significatif par la somme de critères de jugement univoques potentiellement non significatifs est une méthode classique pour augmenter l'impression d'efficacité des résultats. Evidemment, il est important de connaître le bénéfice clinique global

d'une intervention (évalué par un critère composite), et c'est le cas ici dans l'étude LIFE. Mais trop souvent, le critère composite additionne des événements et des procédures interventionnelles (parfois arbitraires ou laissés à l'appréciation du patient), ce qui est beaucoup plus discutable.

**Les critères de jugement de l'étude LIFE ont un intérêt clinique évident.**

Dernier élément important dans l'évaluation d'une étude clinique : les effets indésirables. Dans l'étude LIFE, moins d'effets secondaires sont rapportés avec

le losartan qu'avec l'aténolol. Mais les chiffres doivent être relativisés :

- les différences sont peu importantes et vont dans les 2 sens,
- il s'agit essentiellement d'effets de classe,
- et, fait difficilement explicable, il y a plus de sorties d'étude ("dropped out") avec le losartan (105/4.605) qu'avec l'aténolol (92/4.588), surtout en ce qui concerne le retrait de consentement.

**Il ne semble pas y avoir de différence significative sur le plan clinique entre les 2 cohortes losartan et aténolol dans l'étude LIFE sur le plan des effets indésirables.**

## CONCLUSION

L'étude LIFE est une étude d'intervention, randomisée, contrôlée, multicentrique, dont le design est très bien construit. Toutes les précautions et garanties méthodologiques nécessaires ont été prises pour éviter un maximum de biais et d'interférence de variables confondantes, en dehors d'une modification inexplicée des critères d'inclusion en cours d'étude, mais qui visait très probablement à accentuer l'importance (l'impact) de la différence observée en augmentant le risque. Les conditions éthiques exigées ont été respectées. L'analyse statistique est correcte. Quelques méthodes classiques ont été utilisées pour atteindre plus rapidement des résultats significatifs (sélection d'une population à haut risque) et pour augmenter l'impression d'efficacité du traitement investigué (présentation des résultats en RRR plutôt qu'en RRA ou NNT, et construction d'un critère de jugement primaire composite mais adéquat).

Il y a donc 2 manières différentes de lire les résultats :

- L'étude LIFE montre que le losartan réduit de 13 % un critère de jugement composite (décès cardiovasculaires, AVC et infarctus myocardique) et de 25 % les AVC chez les patients hypertendus avec hypertrophie ventriculaire gauche par rapport à l'aténolol. Ces résultats sont directement applicables dans la pratique clinique et devraient conduire à une modification des *guidelines*.
- L'étude LIFE concerne des patients âgés, avec une HTA sévère et une hypertrophie ventriculaire gauche majeure (moins de 10 % des HTA en médecine générale). Chez ces patients, le NNT pour le critère primaire composite est de 62 pendant 4 ans ; pour l'AVC, il est de 59, et il serait utile de procéder à une étude comparant sartans et inhibiteurs de l'enzyme de conversion (IEC). Pour la majorité des patients en médecine générale, il n'est pas indiqué de modifier les *guidelines* actuels concernant l'HTA et l'aténolol reste en première ligne.

## Remerciements

Nous tenons à remercier notre collègue, le Dr P. Chevalier, pour la relecture minutieuse qu'il a faite

de cet article ainsi que pour ses critiques et commentaires judicieux.

## BIBLIOGRAPHIE

### - Article princeps :

- Dahlöf B, Devereux RB, Kjeldsen SE : Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE) : a randomised trial against atenolol. *Lancet* 2002 ; 359 : 995-1003

### - Articles critiques :

- Decort P : Losartan ou aténolol pour l'hypertension : résultats de l'étude LIFE. *Minerva F* 2003 ; 2 : 20-3
- Scheen AJ : L'étude clinique du mois. L'étude LIFE : protection cardiovasculaire du patient hypertendu par le losartan. *Rev Med Liège* 2002 ; 57 : 240-4
- Julius S : Bringing VALUE to LIFE. *Eur Heart J* 2003 ; 5 (Suppl C) : C5-8

### - Livres et articles de référence :

- Davidoff F, Haynes B, Sackett D *et al* : Evidence-based medicine. A new approach for teaching the practice of medicine. *JAMA* 1992 ; 268 : 24-5
- Greenhalgh T : Savoir lire un article médical pour décider. Meudon, Ran D, 2000
- Haynes RB, McKibbon KA, Fitzgerald D *et al* : How to keep up with medical literature. *Ann Intern Med* 1986 ; 105 : 149-54, 309-12, 474-8, 636-40, 810-6, 978-84
- Jenicek M, Clérout R : Epidémiologie : principes, techniques, applications. Paris, Maloine, 1984
- Khan KS, Kunz R, Kleijnen J *et al* : Systematic reviews to support evidence-based medicine. How to apply findings of health care research. London, Royal Society of Medicine Press, 2003
- Morton RF, Hebel JR : Epidémiologie et biostatistiques. Paris, Doin, 1983
- Pestiaux D, Bouilliez JD (coordination) : L'information médicale, une jungle à défricher. Ottignies Louvain-la-Neuve, Quorum, 1997
- Rosenberg W, Donald A : Evidence-based medicine : an approach to clinical problem-solving. *BMJ* 1995 ; 122 : 727
- Sackett DL : How to read clinical journals. *Can Med Assoc J* 1981 ; 124 : 555-63, 703-10, 869-90, 1156-62
- Schuman SH : Practice-based epidemiology. Amsterdam, Gordon and Breach Science Publishers, 1986

### - Bibliographie proprement dite :

1. Popper K : The logic of scientific discovery. London, Hutchinson, 3<sup>rd</sup> edition, 1968
2. Van Den Ende J, Derese A, Debaene : Medische beslissonde, een nieuw accent in de vlaamse huisartsgeneeskunde. *Huisarts Nu* 1996 ; 25 : 281-340
3. Haynes RB, Sackett D, Gray JMA *et al* : Transferring evidence form research into practice. The role of clinical care research evidence in clinical decisions. *ACP Journal Club* 1996 ; 125 : 14-5
4. Sackett DL, Rosenberg WMC, Muir Gray JA *et al* : Evidence-based medicine : what it is and what it isn't. *BMJ* 1996 ; 312 : 72-3

5. Butler CC, Kinnersley P, Prout H *et al* : Antibiotics and shared decision-making in primary care.  
J Antimicrob Chemother 2001 ; 48 : 435-40
6. Jenicek M, Cléroux R : Epidémiologie : principes, techniques, applications. Paris, Maloine, 1984
7. Wing LM, Reid CM, Ryan P *et al* : A comparison of outcomes with angiotensin-converting-enzyme inhibitors and diuretics for hypertension in the elderly. N Engl J Med 2003 ; 348 : 583-92
8. Di Napoli M, Papa F : Angiotensin-converting enzyme inhibitor use is associated with reduced plasma concentration of C-reactive protein in patients with first-ever ischemic stroke. Stroke 2003 ; 34 : 2922-9
9. De Cort P : Vergelijkende studie van de bloeddruk, verkregen met verschillende methoden, bij personen boven de 60 jaar (proefschrift). Leuven, Katholieke Universiteit Leuven, 1998
10. [http://www.hclm.qc.ca/wmfichiers/projet\\_actif\\_2003-2004.pdf](http://www.hclm.qc.ca/wmfichiers/projet_actif_2003-2004.pdf)
11. James B, Young MD, Mark E *et al* for the Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity (CHARM) Investigators and Committees : Mortality and Morbidity Reduction With Candesartan in Patients With Chronic Heart Failure and Left Ventricular Systolic Dysfunction. Circulation 2004 ; 110 : 2618-26
12. Sackett DL, Cook RJ : Understanding clinical trials. BMJ 1994 ; 309 : 755-6
13. Misselbrook D, Armstrong D : Patients' responses to risk information about the benefits of treating hypertension. Br J Gen Pract 2001 ; 51 : 276-9

**Correspondance et tirés à part :**

M. ROLAND  
Avenue Van Volxem 45  
1190 Bruxelles

Travail reçu le 11 mai 2006 ; accepté dans sa version définitive le 1<sup>er</sup> décembre 2006.