

MÉDECINE FACTUELLE :

la hiérarchisation des preuves par le Centre for Evidence-Based Medicine d'Oxford

N. DURIEUX (1), S. VANDENPUT (2), F. PASLEAU (3)

RÉSUMÉ : La hiérarchisation des données issues de la recherche médicale est une notion essentielle en Evidence-Based Medicine (EBM). Le Centre for Evidence-Based Medicine d'Oxford a publié en 1998 un classement basé sur les schémas d'étude (niveaux de preuve) et l'a actualisé en 2011. La nouvelle version, plus conviviale, est focalisée sur les questions de prévalence, de diagnostic, de pronostic, de traitement et de dépistage. Cette approche est présentée et discutée dans cet article.

MOTS-CLÉS : Médecine factuelle - Compétence informationnelle en santé - Niveaux de preuve

OCEBM LEVELS OF EVIDENCE SYSTEM

SUMMARY : The hierarchy of evidence from medical research is a key concept in Evidence-Based Medicine (EBM). In 1998, the Oxford Centre for Evidence-Based Medicine published a levels table based on study designs. An updated and more user-friendly version, published in 2011, focuses on the issues of prevalence, diagnosis, prognosis, treatment and screening. This paper presents and discusses this approach.

Keywords : Evidence-Based Medicine - Health literacy - Levels of evidence

INTRODUCTION

L'*Evidence-Based Medicine* (EBM) - en français, médecine factuelle ou médecine basée sur les données probantes - est une démarche qui invite le praticien à exploiter de manière judicieuse les données les plus fiables issues de la recherche médicale («preuves») pour une prise en charge personnalisée des patients (1). Alors que les médecins sont de plus en plus souvent amenés à devoir justifier leurs pratiques (2), soit auprès de leurs patients, soit auprès des organismes de soins de santé, l'EBM apparaît comme une approche rigoureuse et utile (3). La pratique de l'EBM implique cependant d'avoir accès à la littérature scientifique, d'être capable d'utiliser efficacement les ressources disponibles et d'analyser l'information de manière critique (4).

Trouver les meilleures «preuves» du moment pour chacune des situations envisagées est un des obstacles majeurs à l'intégration de l'EBM dans la pratique quotidienne. Toutes les publications ne sont pas de qualité équivalente (5). Le degré de confiance à accorder aux résultats d'une étude (niveau de preuve) est conditionné par (i) le schéma de l'étude (*study design*) ainsi que (ii) la qualité de la méthodologie et de l'analyse des résultats (6, 7). Le premier critère est, en principe, repérable dès la lecture du titre

et du résumé. Le second nécessite une lecture approfondie et critique du document.

La notion de niveaux de preuve est fondamentale en médecine factuelle (8). Elle est classiquement représentée dans la littérature sous la forme d'une pyramide et est liée aux schémas d'étude (fig. 1) (6). Le plus haut niveau de confiance est accordé aux informations qui résultent de synthèses méthodiques (revues systématiques). Ces synthèses sont basées sur une recherche exhaustive d'études originales abordant une même question et sur une sélection des études les plus rigoureuses. Les études expérimentales, en particulier les essais contrôlés randomisés (RCT), conçues de manière à maîtriser un maximum de variables susceptibles d'influencer la qualité et l'interprétation des résultats, sont en principe plus fiables que les études d'observation. Les séries de cas ne permettent pas de mettre en évidence une relation causale entre différentes variables, mais elles sont intéressantes pour les hypothèses qu'elles suscitent. Les synthèses non méthodiques (revues narratives) se retrouvent dans le bas de la pyramide, car elles ne résultent pas d'une analyse systématique de l'existant.

Le principal défaut de la pyramide classique est qu'elle ne tient pas compte du fait que tous les schémas d'étude ne peuvent s'appliquer à toutes les situations. Un classement théorique plus nuancé a été proposé en 1998 par le *Centre for Evidence-Based Medicine* de l'Université d'Oxford (OCEBM). Il a été actualisé, tant sur la forme que sur le fond, en 2011 (7). L'objectif du présent article est de décrire et commenter cette dernière version.

(1) Assistante, (2) Conservateur, (3) Conservateur et Directeur de la Bibliothèque des Sciences de la Vie, Université de Liège.

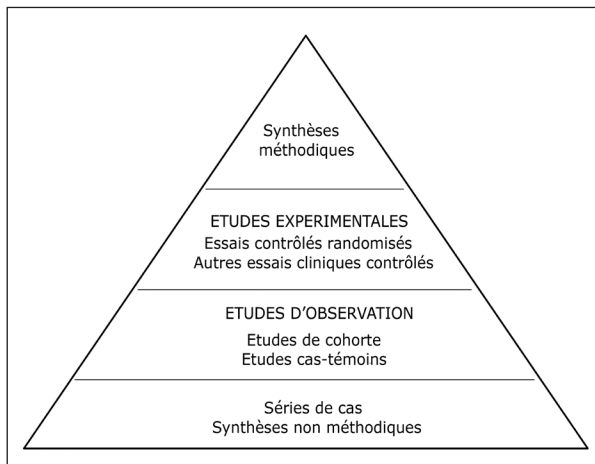


Figure 1. Hiérarchie simplifiée des niveaux de preuve en fonction des schémas d'étude (*study designs*). Adaptée de Greenhalgh T. - How to read a paper : the basics of evidence-based medicine. Fourth edition. Wiley-Blackwell, Chichester, 2010, 18.

LE CLASSEMENT DES PREUVES DU CENTRE D'EVIDENCE-BASED MEDICINE D'OXFORD (2011)

Le principe du classement de l'OCEBM (tableau I) est de proposer, pour les questions cliniques les plus courantes (prévalence, diagnostic, pronostic, traitement et dépistage), une hiérarchisation en cinq paliers des schémas d'étude possibles. Si, pour une question donnée, il n'existe pas de publication correspondant au schéma d'étude idéal (palier 1), le praticien est invité à considérer le palier suivant, avec les réserves que cela comporte.

QUESTIONS RELATIVES À LA PRÉVALENCE D'UNE PATHOLOGIE

Lorsqu'on souhaite connaître la fréquence ou la répartition d'un problème de santé à un moment donné (prévalence), l'idéal est de disposer d'enquêtes récentes portant sur la totalité de la population (recensement) ou sur un échantillon aléatoire représentatif d'une situation. À défaut, le praticien est invité à se tourner vers des revues systématiques d'enquêtes qui sont menées dans des conditions transposables à la situation étudiée (palier 2), vers des études d'échantillons non aléatoires (palier 3) et, en dernier recours, vers des séries de cas (palier 4).

QUESTIONS RELATIVES À L'INTÉRÊT D'UN TEST DIAGNOSTIQUE

Concernant la fiabilité diagnostique d'un test, plusieurs contrôles doivent avoir été effectués : (i) la comparaison du test à évaluer à un test de référence définitif et indépendant; (ii) l'appli-

cation indépendante des deux tests à l'ensemble de l'échantillon expérimental; (iii) l'application des deux tests dans un intervalle de temps très court; (iv) la procédure en aveugle (9, 10).

Le plus haut niveau de preuve est attribué à des revues systématiques d'études transversales, menées en aveugle, dans lesquelles deux tests diagnostiques (l'un à évaluer et l'autre de référence) ont été appliqués systématiquement à tous les patients sélectionnés. Si de telles revues systématiques ne sont pas disponibles, il est recommandé de se tourner vers des études transversales présentant toutes les caractéristiques énoncées ci-dessus. La qualité des études originales diminue si un standard de référence n'a pas été appliqué de manière systématique ou s'il y a un biais de sélection des patients, comme dans le cas d'un recrutement non consécutif (palier 3). Le risque de biais augmente encore (palier 4) : (i) si le test de référence a été utilisé pour distinguer les malades et les non-malades (études cas-témoins); (ii) si le test de référence pêche par défaut d'indépendance ou s'il est de faible qualité.

QUESTIONS RELATIVES À L'ÉVOLUTION D'UNE MALADIE OU À LA PROBABILITÉ DE SURVENUE DE CERTAINS ÉVÉNEMENTS

Pour les questions de pronostic qui orientent le choix d'une intervention, voire l'évaluation du risque en cas de non-intervention, l'idéal est de consulter des revues systématiques d'études de cohorte dans lesquelles les patients ont été inclus dès le début de la maladie («cohorte de départ» ou en anglais, *inception cohort study*) (palier 1). S'il n'existe pas de synthèse méthodique, il est recommandé de se tourner vers des études de cohorte, de préférence des «cohortes de départ» (palier 2). Considérer l'évolution du groupe contrôle d'un RCT est également envisageable (palier 3). Les autres types d'études (palier 4) apportent un niveau de preuve plus faible : séries de cas, études cas-témoins et études de cohorte de pauvre qualité pronostique.

QUESTIONS RELATIVES AUX BÉNÉFICES D'UN TRAITEMENT

Le plus haut niveau de preuve est accordé à des revues systématiques de RCT ou à des essais de taille 1 (en anglais, *n-of-1 trials*) où plusieurs traitements sont appliqués de façon séquentielle à un même patient, qui est son propre contrôle, de façon à déterminer l'intervention qui lui convient le mieux. À défaut, le praticien se tournera vers des RCT ou des études d'observation rapportant un effet majeur.

TABLEAU I. CLASSEMENT DES «PREUVES» PAR LE CENTRE D'EVIDENCE-BASED MEDICINE D'OXFORD

Question	Palier 1 (Niveau 1*)	Palier 2 (Niveau 2*)	Palier 3 (Niveau 3*)	Palier 4 (Niveau 4*)	Palier 5 (Niveau 5)
Le problème est-il courant ?	Enquête locale et récente sur des échantillons aléatoires (ou recensement)	Revue systématique d'enquêtes dont les conditions sont transposables aux conditions locales	Etude locale d'échantillons non aléatoires	Série de cas	/
Quelle est l'exactitude d'un test diagnostique ou de suivi ? (Diagnostic)	Revue systématique d'études transversales menées en aveugle et utilisant un standard de référence appliqué de manière systématique	Etude transversale menée en aveugle et utilisant un standard de référence appliqué de manière systématique	Etude à recrutement non consécutif; étude sans application systématique d'un standard de référence	Etude cas-témoins; étude utilisant un standard de référence non indépendant ou de faible qualité	Raisonnement déductif
Que se passera-t-il si aucun traitement n'est appliqué ? (Pronostic)	Revue systématique d'études de cohorte où les patients sont inclus au début de leur maladie (<i>inception cohort</i>)	Etude de cohorte où les patients sont inclus au début de leur maladie (<i>inception cohort</i>)	Etude de cohorte; groupe contrôlé d'un essai contrôlé randomisé	Série de cas; étude cas-témoins; étude de cohorte pronostique de pauvre qualité	/
Cette intervention est-elle bénéfique ? (Bénéfices du traitement)	Revue systématique d'essais contrôlés randomisés; Essai de taille 1 (<i>n-of-1 trial</i>)	Essai contrôlé randomisé; étude d'observation rapportant un effet majeur	Essai contrôlé non randomisé; étude de cohorte prospective	Série de cas; étude cas-témoins; étude de cohorte historique	Raisonnement déductif
Quels sont les effets indésirables fréquents ? (Effets indésirables du traitement)	Revue systématique d'essais contrôlés randomisés; revue systématique d'études cas-témoins recrutés dans la population d'une étude de cohorte; essai de taille 1 (<i>n-of-1 trial</i>); étude d'observation rapportant un effet majeur	Essai contrôlé randomisé; (exceptionnellement) étude d'observation rapportant un effet majeur	Etude de cohorte prospective (surveillance post-commercialisation) à condition qu'il y ait un nombre suffisant de patients par rapport à la fréquence de l'événement (pour les effets à long terme, la durée du suivi doit être suffisante)	Série de cas; étude cas-témoins; étude de cohorte historique	Raisonnement déductif
Quels sont les effets indésirables rares ? (Effets indésirables du traitement)	Revue systématique d'essais contrôlés randomisés; essai de taille 1 (<i>n-of-1 trial</i>)	Essai contrôlé randomisé; (exceptionnellement) étude d'observation avec effet majeur			
Ce test (de détection précoce) vaut-il la peine ? (Dépistage)	Revue systématique d'essais contrôlés randomisés	Essai contrôlé randomisé	Etude de cohorte prospective	Série de cas; étude cas-témoins; étude de cohorte historique	Raisonnement déductif

Traduction française (effectuée par Durieux N, Pasleau F, Howick J) du 2011 *Oxford CEBM Levels of Evidence* rédigé par l'OCEBM Levels of Evidence Working Group (Howick J, Chalmers I, Glasziou P, et al.) et disponible à l'adresse <http://www.cebm.net/index.aspx?o=5653> (reproduit avec l'accord de Howick, J).

*Le niveau de preuve d'une étude peut être rétrogradé sur base des faiblesses intrinsèques de l'étude, d'imprécisions, du caractère indirect de la preuve, à cause de l'incohérence entre études, ou à cause de la taille de l'effet absolu qui est très petit; le niveau de preuve peut être mieux classé si la taille de l'effet est grande ou très grande.

Les études de cohorte prospectives ou les essais contrôlés non randomisés se retrouvent au troisième palier. Les séries de cas, les études cas-témoins, les études de cohorte historiques (rétrospectives) offrent des niveaux de preuve plus faibles.

QUESTIONS RELATIVES AUX EFFETS INDÉSIRABLES D'UN TRAITEMENT

Pour ces questions, le même raisonnement que précédemment peut être tenu à quelques nuances près (paliers 1 et 3).

Si les effets indésirables sont fréquents, les études d'observation mettant en évidence un effet majeur peuvent apporter un haut niveau de preuve (palier 1), de même que des revues systématiques un peu particulières désignées en anglais par *systematic reviews of nested case-control studies*. Il s'agit de synthèses méthodiques d'études cas-témoins dont les sujets ont été recrutés au sein d'une population qui a fait l'objet d'une étude de cohorte. L'intérêt de cette approche est que l'observateur dispose, dès le départ, de données précises sur les sujets, ce qui lui permet de réduire, voire d'éliminer, les effets de certaines variables de confusion (9).

Au troisième palier, apparaissent les études d'observation menées dans le cadre d'une surveillance post-commercialisation. Ces dernières doivent porter sur un nombre suffisant de patients et durer assez longtemps pour révéler les effets rares ou ceux qui ne se dévoilent qu'à long terme.

QUESTIONS RELATIVES À L'UTILITÉ D'UN TEST DE DÉPISTAGE

Dans cette dernière ligne du tableau réapparaît le modèle classique de hiérarchisation des preuves où les synthèses méthodiques ont plus de poids que les études individuelles et où les études expérimentales ont plus de poids que les études d'observation ou les études descriptives.

Pour finir, quelle que soit la question posée et en l'absence de toute donnée probante, le recours ultime (palier 5) est le raisonnement basé sur les connaissances disponibles.

DISCUSSION

La profusion de publications scientifiques, leur hétérogénéité et leur qualité très inégale (11) constituent autant d'obstacles à la pratique de l'EBM. Le principe de hiérarchisation des preuves a été conçu pour aider le praticien à sélectionner les données les plus solides issues de la recherche médicale. Un classement comme celui de l'OCEBM, qui détaille avec finesse les niveaux de preuve en fonction de la qualité attendue des divers types d'études pour chaque type de question envisagée (prévalence, diagnostic, pronostic, intervention et dépistage), permet d'affiner les stratégies de recherche et d'opérer une présélection des publications potentiellement intéressantes. Les logiciels de recherche spécialisés tels que PubMed ou OvidSP ont intégré la notion de niveaux de preuve en ce sens qu'ils permettent

de trier en quelques clics les références de Medline en fonction du schéma d'étude (filtre) : revues méthodiques (systématiques), RCT et autres essais cliniques contrôlés, études de cohorte et études cas-témoins, séries de cas et revues narratives. Une maîtrise suffisante des technologies de l'information apportera un gain de temps considérable tout au long du processus. À partir de cette présélection, le repérage des études qui répondent à des critères de qualité bien définis est sous la responsabilité du praticien qui travaille d'abord sur base des informations disponibles dans les notices bibliographiques.

La sélection finale des données probantes repose sur une analyse approfondie de la méthodologie et de la validité des résultats présentés dans chacune des études identifiées. Pour ce faire, des grilles facilitant l'évaluation critique de l'information existent : comme par exemple, la grille CONSORT pour les essais randomisés (12) ou le système GRADE pour les recommandations de bonnes pratiques (13). Cette analyse est essentielle et a une influence prépondérante sur la hiérarchisation des études, qui peuvent descendre ou monter d'un niveau en fonction de leurs qualités intrinsèques (7) : ainsi, il est préférable de tenir compte des données d'une bonne étude de cohorte plutôt que de celles d'un RCT biaisé.

Il est à noter que le niveau de preuve à accorder aux méta-analyses n'est pas abordé en tant que tel dans le tableau de l'OCEBM. La méthodologie de la méta-analyse n'est pas en elle-même un gage de qualité dans la mesure où il s'agit d'un procédé mathématique de synthèse des résultats de plusieurs études indépendantes partageant la même hypothèse de travail et le même schéma expérimental (14). Cependant, la méta-analyse prend tout son sens si elle vient compléter une revue systématique valide et fiable de la littérature afin d'augmenter la précision du résultat (15). Elle est alors considérée au même titre que les synthèses méthodiques de la littérature (J. Howick, communication personnelle, 1^{er} octobre 2012).

La notion de niveaux de preuve a ses détracteurs. Certains auteurs jugent que les hiérarchies sont rigides, qu'elles sous-estiment les limites des essais contrôlés randomisés et surestiment les limites des études d'observation (16), qu'elles ne doivent pas être utilisées de manière radicale (17), et que le bon sens, qui consiste à tenir compte des risques potentiels et des bénéfices d'une intervention, a également sa place dans le processus de décision clinique (18).

Pourtant, dès 1996, au moment où ils définissent l'approche EBM, Sackett et al. (1) ont bien expliqué que l'EBM ne se restreint pas à la recherche des seuls RCT et méta-analyses mais qu'il importe de trouver les meilleures preuves pour chacune des questions posées. Les auteurs insistent aussi sur le rôle du clinicien et sur les valeurs et préférences des patients. Il est vrai que ces deux piliers sont souvent oubliés dans les discussions (19, 20) alors que les données probantes doivent servir à informer et à guider les intervenants et non à imposer le choix d'une intervention (19).

Le classement de l'OCEBM tient compte de ces réflexions en faisant appel au bon sens du praticien (21) et en soulignant l'importance d'évaluer la qualité intrinsèque des études (méthodologie, taille de l'effet...) et leur cohérence par rapport à d'autres résultats (7).

L'OCEBM Levels of Evidence Working Group (21) considère que le classement de l'OCEBM peut également être un outil pour les patients. Cependant, il faut reconnaître que les publications concernées ne s'adressent pas au grand public et que comprendre la philosophie d'un tel classement demande une bonne connaissance de la méthodologie de la recherche scientifique. Bien que contestable, la prise de position de l'OCEBM a néanmoins le mérite d'insister sur la nécessité d'informer le patient en lui fournissant une information fiable présentée de manière adaptée.

L'utilité du classement de l'OCEBM pour le praticien est incontestable. Néanmoins, au quotidien, lorsqu'il est face à un patient et qu'il a besoin d'une réponse rapide, ce dernier préférerait sans doute avoir à sa disposition des ressources «pré-évaluées», pour autant qu'il en existe et qu'il y ait accès. Peuvent être citées, à titre d'exemple, les recommandations de bonne pratique fondée sur les données probantes (en anglais, *evidence-based, current clinical guidelines*) et les synopses de revues systématiques ou d'études individuelles (22). Toutefois, dans ce cas de figure, le clinicien délègue à d'autres le soin d'évaluer les données issues de la littérature.

CONCLUSION

Les trois premières étapes de l'EBM requièrent de localiser et d'évaluer les meilleures données disponibles pour répondre à une question posée *a priori* (23). Les systèmes de classement des preuves sont des indicateurs théoriques utiles à tous ceux qui sont ame-

nés à exploiter les connaissances issues de la recherche médicale. Quelle que soit leur présentation, ces systèmes véhiculent le même message : toutes les preuves, toutes les informations ne sont pas équivalentes; il est, dès lors, essentiel de s'assurer de la crédibilité de l'information trouvée, peu importe d'où elle provient (6). Ce message s'adresse à tous les professionnels de la santé, quelles que soient les disciplines dans lesquelles ils exercent. Dans un contexte dominé par l'explosion des publications et des outils de communication, l'exercice de la critique objective est plus que jamais indispensable au maintien de la qualité des soins de santé.

REMERCIEMENTS

Nous souhaitons remercier le Dr. Jeremy Howick de nous avoir fait l'honneur de nous associer à son travail de valorisation des recommandations du OCEBM Levels of Evidence Working Group et pour les discussions enrichissantes que cette collaboration a suscitées.

BIBLIOGRAPHIE

1. Sackett DL, Rosenberg WM, Gray JA, et al.— Evidence based medicine : what it is and what it isn't. *BMJ*, 1996, **312**, 71-72.
2. De Cort P.— Evidence-based practice. *Minerva*, 2006, **5**, 49.
3. Delvenne C, Pasleau F.— Comment résoudre en pratique un problème diagnostique ou thérapeutique en suivant une démarche EBM ? *Rev Med Liège*, 2000, **55**, 226-232.
4. Delvenne C, Pasleau F.— Sélection de ressources EBM sur l'Internet. *Rev Med Liège*, 2000, **55**, 220-225.
5. Evidence-Based Medicine Working Group.— Evidence-based medicine : a new approach to teaching the practice of medicine. *JAMA*, 1992, **268**, 2420-2425.
6. Greenhalgh T.— *How to read a paper : the basics of evidence-based medicine*. Quatrième édition. Wiley-Blackwell, Chichester, 2010, 31-60, 18-19.
7. OCEBM Levels of Evidence Working Group.— *The 2011 Oxford Levels of Evidence*. <http://www.cebm.net/index.aspx?o=5653> - Consultation du 8 août 2013
8. Guyatt G, Haynes B, Jaeschke R, et al.— The philosophy of evidence-based medicine. In Guyatt G, Drummond R, Meade MO, Cook DJ Ed., *Users' guides to the medical literature : a manual for evidence-based clinical practice*. McGraw Hill Medical, New York, 2008, 9-16.
9. Salmi LR.— *Lecture critique et communication médicale scientifique : comment lire, présenter, rédiger et publier une étude clinique ou épidémiologique*. Troisième édition. Elsevier Masson, Paris, 2012, 166-188, 210-213, 231.

10. Clinical Evidence : diagnostic test studies : assessment and critical appraisal.— <http://clinicalevidence.bmj.com/x/set/static/ebm/toolbox/665061.html> - Consultation du 8 juillet 2013.
11. Smith R.— *The trouble with medical journal*. Royal Society of Medicine, London, 2006.
12. Consort.— <http://www.consort-statement.org/home/>.- Consultation du 8 août 2013
13. GRADE.— <http://www.gradeworkinggroup.org/index.htm>.- Consultation du 8 août 2013
14. Scheen AJ.— Comment j'explore ... les secrets d'une méta-analyse. *Rev Med Liège*, 2003, **58**, 41-46.
15. Greenhalgh T.— How to read a paper : papers that summarise other papers (systematic reviews and meta-analyses). *BMJ*, 1997, **315**, 672.-675.
16. Concato J.— Observational versus experimental studies : what's the evidence for a hierarchy? *NeuroRx*, 2004, **1**, 341-347.
17. La Caze A.— Evidence-based medicine must be. *J Med Philos*, 2009, **34**, 509-527.
18. Smith GCS, Pell JP.— Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ*, 2003, **327**, 1459-1461.
19. Glasziou P.— Evidence based medicine : does it make a difference? Make it evidence informed practice with a little wisdom. *BMJ*, 2005, **330**, 92.
20. Rosner AL.— Evidence-based medicine : revisiting the pyramid of priorities. *J Bodyw Mov Ther*, 2012, **16**, 42-49.
21. OCEBM Levels of Evidence Working Group.— *Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence : background document*. <http://www.cebm.net/index.aspx?o=5653> - Consultation du 9 août 2013
22. DiCenso A, Bayley L, Haynes RB.— Editorial : accessing preappraised evidence : fine-tuning the 5S model into a 6S model. *ACP J Club*, 2009, **151**, JC3-2.
23. Straus SE, Glasziou P, Richardson WS, et al.— *Evidence-based medicine : how to practice and teach it*. Quatrième édition. Churchill Livingstone Elsevier, Edinburgh, 2011, 3.

Les demandes de tirés à part sont à adresser à Mme N. Durieux, Assistante, Bibliothèque des Sciences de la Vie, Université de Liège, CHU de Liège, Belgique.
Email : Nancy.Durieux@ulg.ac.be