

É

Troubles musculo-squelettiques

Études et recherches

RAPPORT R-736



Bilan des connaissances sur les guides de pratique en santé

**Enseignements clés et transférabilité
pour la santé et la sécurité au travail**

*Monique Lortie
Elena Laroche
Denys Denis
Iuliana Nastasia
Cheikh Faye
Sylvie Gravel
Laurent Giraud
Lise Desmarais*



Solidement implanté au Québec depuis 1980, l'Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) est un organisme de recherche scientifique reconnu internationalement pour la qualité de ses travaux.

NOS RECHERCHES

travaillent pour vous !

Mission

Contribuer, par la recherche, à la prévention des accidents du travail et des maladies professionnelles ainsi qu'à la réadaptation des travailleurs qui en sont victimes.

Offrir les services de laboratoires et l'expertise nécessaires à l'action du réseau public de prévention en santé et en sécurité du travail.

Assurer la diffusion des connaissances, jouer un rôle de référence scientifique et d'expert.

Doté d'un conseil d'administration paritaire où siègent en nombre égal des représentants des employeurs et des travailleurs, l'IRSST est financé par la Commission de la santé et de la sécurité du travail.

Pour en savoir plus

Visitez notre site Web ! Vous y trouverez une information complète et à jour. De plus, toutes les publications éditées par l'IRSST peuvent être téléchargées gratuitement. www.irsst.qc.ca

Pour connaître l'actualité de la recherche menée ou financée par l'IRSST, abonnez-vous gratuitement au magazine Prévention au travail, publié conjointement par l'Institut et la CSST. Abonnement : 1-877-221-7046

Dépôt légal

Bibliothèque et Archives nationales
2012
ISBN : 978-2-89631-607-6 (PDF)
ISSN : 0820-8395

IRSST - Direction des communications
et de la valorisation de la recherche
505, boul. De Maisonneuve Ouest
Montréal (Québec)
H3A 3C2
Téléphone : 514 288-1551
Télécopieur : 514 288-7636
publications@irsst.qc.ca
www.irsst.qc.ca
© Institut de recherche Robert-Sauvé
en santé et en sécurité du travail,
mai 2012



Troubles musculo-squelettiques

Études et recherches

■ RAPPORT R-736

Bilan des connaissances sur les guides de pratique en santé Enseignements clés et transférabilité pour la santé et la sécurité au travail

Avis de non-responsabilité

L'IRSST ne donne aucune garantie relative à l'exactitude, la fiabilité ou le caractère exhaustif de l'information contenue dans ce document. En aucun cas l'IRSST ne saurait être tenu responsable pour tout dommage corporel, moral ou matériel résultant de l'utilisation de cette information.

Notez que les contenus des documents sont protégés par les législations canadiennes applicables en matière de propriété intellectuelle.

*Monique Lortie,
Département des sciences biologiques, UQAM*

*Elena Laroche,
Télé-Université, UQAM*

*Denys Denis et Iuliana Nastasia,
Prévention des problématiques de SST et réadaptation, IRSST*

*Cheikh Faye et Sylvie Gravel,
École des sciences de la gestion, UQAM*

*Laurent Giraud,
Prévention des risques mécaniques et physiques, IRSST*

*Lise Desmarais,
Université de Sherbrooke*

Cliquez recherche
www.irsst.qc.ca



Cette publication est disponible
en version PDF
sur le site Web de l'IRSST.

Cette étude a été financée par l'IRSST. Les conclusions et recommandations sont celles des auteurs.

CONFORMÉMENT AUX POLITIQUES DE L'IRSST

Les résultats des travaux de recherche publiés dans ce document
ont fait l'objet d'une évaluation par des pairs.

SOMMAIRE

Les guides sont considérés comme un outil important au niveau du transfert des connaissances. Ils permettent une organisation et une mise en forme opérationnelle des connaissances pour que les milieux puissent se les approprier. Les guides sont un outil de relais, par ailleurs ciblés par l'axe en transfert de connaissances du Réseau de recherche en santé et sécurité au travail du Québec en tant qu'enjeu important voire central en SST. À cet égard, le sujet d'intérêt particulier est l'utilisation et l'appropriation des guides. D'après l'enquête de Laroche (2009), menée auprès de chercheurs canadiens en santé et sécurité au travail (SST), le tiers avait été impliqué au cours des cinq dernières années dans l'élaboration d'un guide.

C'est dans ce contexte qu'a surgi au sein de cette équipe une première question simple : Sait-on ce qu'est un bon guide et comment l'évaluer? Existe-t-il une grille ou des critères d'évaluation qui nous permettraient de poser un premier diagnostic sur un guide ? Les questions corollaires étaient : Que sait-on sur leur utilisation et leur appropriation ? Ont-ils les impacts escomptés ?

Les guides en SST ayant été l'objet de peu de recherches, notre bilan de connaissances s'est rapidement orienté vers d'autres domaines, où rapidement, l'importance des travaux menés en santé s'est imposée. En effet, tant la production de guides de pratique clinique que le nombre de recherches menées sur le sujet ont crû de façon spectaculaire au cours des deux dernières décennies. Nous nous y sommes donc arrêtés, avec pour objectif d'en saisir autant la logique que la cohérence, afin de mieux en comprendre les développements et les questionnements.

Cependant, faire une revue exhaustive des écrits, vu leur ampleur, était irréaliste. Nous avons plutôt opté pour une démarche visant à dégager des points de repère qui pourraient être utiles aux acteurs en SST quant au développement des guides, à leur évaluation et à leur implantation. Deux grands axes de recherche et de développement s'y côtoient et s'entremêlent. Le premier axe est centré sur l'objet lui-même, le guide : comment les développer, les évaluer, les implanter? Le second axe est centré sur son utilisateur : Quel usage en fait-il ? Qu'en pense-t-il? Que recherche-t-il ?

Suite à cette revue des écrits, un premier constat a été l'importance de l'implication des organisations publiques, voire des États, dans le développement des guides, ces derniers voyant les guides de pratique comme une pierre angulaire des politiques de santé. Les guides pourraient en améliorer la performance en favorisant l'implantation de ce qu'on appelle *evidence-based medicine*, une médecine fondée sur des preuves obtenues de façon rigoureuse. Dans les années qui ont suivi le début de l'élaboration de guides de pratique en santé, des groupes de travail nationaux ont été mis en place - essentiellement dans les pays anglophones - pour statuer et élaborer des règles de développement dont le cœur est le tri et l'évaluation des preuves. En fait, fondamentalement, les guides de pratique sont des regroupements d'énoncés, de directives. En anglais, on parle de *clinical practical guidelines*. Cela est compréhensible en santé, mais sans doute moins approprié en SST, où les guides de pratique (et pratiques) ont un sens plus large. Certains organismes internationaux, comme le Bureau International du Travail, commencent cependant à adopter cette perspective de «*guidelines*».

Globalement, le développement des guides de pratique est abordé de façon très méthodologique dans le domaine de la santé, le déroulement de chaque étape étant soigneusement défini.

Cependant, nous n'avons pas trouvé – c'est une critique qui leur est adressée aussi – d'études montrant que les guides de pratique ainsi conçus sont de meilleure qualité. Les études de suivi menées auprès des développeurs de guides de pratique montrent aussi que les règles préconisées ne sont que bien partiellement adoptées, les procédures étant considérées trop lourdes et trop coûteuses à mettre en place. Les guides sont aussi avant tout développés par des professionnels qui en font une spécialité. Cette approche professionnalisante ne nous apparaît pas convenir à la SST.

Quant aux outils d'évaluation des guides développés en santé, ils suivent en fait la même logique ; ils évaluent essentiellement le processus de développement du guide, ou du moins, sa description. L'application de ces outils montre que les guides développés n'y satisfont que bien partiellement. Le reproche précédent signalé demeure : les outils évaluent le processus de développement, ce qui en est décrit en fait, et non la qualité elle-même. Nous n'avons donc pas trouvé de réponse satisfaisante à notre première question : si on ouvre un guide, y a-t-il un instrument qui permette d'évaluer son contenu, sa structure, son organisation, etc ? Ceci dit, les questions posées par ces recherches et les résultats obtenus sont riches d'enseignements pour la SST ; ils nous ont aidés à situer nos propres enjeux et perspectives, qui sont progressivement exposés dans ce rapport.

De façon générale, nos travaux ont révélé que trois grandes critiques sont adressées aux guides (directives) : ils ne prennent pas suffisamment en compte l'ensemble des éléments (ne sont pas suffisamment systémiques) ; ils ne collent pas suffisamment au processus de décision du praticien (ils sont construits sur une logique de résolution de problèmes); ils ne cadrent pas suffisamment avec le fait que chaque patient est une entité spécifique (et non un ensemble ou un patient moyen). Ces critiques remettent en partie en question le processus d'élaboration des connaissances. Les questions soulevées et leurs débats sont tout à fait pertinents au domaine de la SST.

D'autre part, les études menées auprès des utilisateurs montrent que ces derniers souscrivent aux efforts de rationalisation, mais les suivis d'utilisation montrent que les recommandations ne sont pas appliquées au niveau attendu. La raison première n'est pas que les gens n'y adhèrent pas, ni qu'ils les méconnaissent. Ce serait plutôt que les conditions d'application, les contextes et les efforts à investir y font obstacle. Les utilisateurs redoutent aussi toute approche coercitive qui pourrait se substituer à leur jugement. Au vu des efforts investis, ces résultats sont considérés comme décevants. Les études d'impact confirment par ailleurs que les effets attendus n'ont pas été au rendez-vous.

Ceci a amené les chercheurs à se pencher plus sur la question de l'implantation et à chercher à mieux comprendre ce qui fait obstacle ou au contraire favorise l'utilisation des guides de pratique, reconnaissant que trop peu d'efforts avaient été consacrés à cette question. Nous en retenons trois résultats clés: la possibilité de voir (ex. quelqu'un utiliser le guide, ce que cela donne, les bénéfices), de disposer de conditions adéquates (ex. ressources humaines et matérielles, organisation, support) et l'effort à consentir.

Enfin, au niveau de l'implantation, l'ensemble des études menées montre que les moyens passifs comme la diffusion de matériel sont peu efficaces et qu'il faut élaborer un plan d'implantation combinant des moyens. Cependant, aucune des synthèses effectuées sur la question ne met en

évidence de moyens ou combinaisons foncièrement efficaces. Les impacts demeurent au mieux modestes. Nous en concluons qu'aucune stratégie d'implantation n'aura de succès si le contenu des guides ne correspond pas suffisamment à leurs conditions d'application.

Il nous apparaît ainsi central d'intégrer les questions d'implantation au processus de développement lui-même, non pas comme une étape finale, mais tout au long du processus. Et bien que l'ensemble des résultats mentionnés montre que statuer sur une méthode d'élaboration ne soit pas un gage de succès, le processus de développement demeure quant à nous une clé importante. Dans ce processus, le cœur devrait cependant en être moins celui de l'identification des preuves à traduire en directives que celui de l'identification des personnes porteuses d'un savoir pertinent aux questions d'implantation. Le rôle déterminant des contextes qui ressort des études nous convainc qu'il doit demeurer au cœur des développements.

TABLE DES MATIÈRES

SOMMAIRE.....	I
INTRODUCTION.....	1
1 HISTORIQUE, CONCEPTS ET APPROCHES RELIÉS À LA PRODUCTION DES GUIDES	5
1.1 Survol historique et mise en contexte.....	5
1.2 Terminologie de base	6
1.3 Approches reliées au développement des guides.....	9
1.3.1 Modèles pour colliger les données.....	10
1.3.2 Méthodes d'identification des données probantes : revue systématique et méta-analyse.....	11
1.3.3 Classification des méthodes qualitatives de recherche	13
1.3.4 Les principales méthodes de consensus utilisées.....	14
1.4 Mise en parallèle avec la SST.....	15
1.4.1 Les données alimentant les contenus des guides	15
1.4.2 Les guides produits en SST.....	17
2 LES GUIDES SUR LES GUIDES, LEUR UTILISATION ET LES INSTRUMENTS D'ÉVALUATION	19
2.1 Les étapes d'élaboration des guides	20
2.1.1 La sélection du sujet	20
2.1.2 La délimitation du but et du public cible anticipé (scope).....	21
2.1.3 L'adaptation du guide.....	21
2.1.4 La formation du groupe de développement.....	23
2.1.5 La participation des patients	24
2.1.6 L'élaboration des questions.....	24
2.1.7 La revue systématique	24
2.1.8 Les critères d'inclusion et d'exclusion des études et des données.....	25
2.1.9 L'évaluation des données probantes ou des preuves	25
2.1.10 Le développement des recommandations.....	27
2.1.11 L'élaboration d'une stratégie d'implantation	28
2.1.12 Le processus de consultation pour évaluer la version préliminaire du guide de pratique.....	28
2.1.13 La rédaction de la version abrégée	29
2.1.14 La planification de l'évaluation, révision et mise à jour	29
2.2 L'application des recommandations des GsG.....	30
2.2.1 Les données colligées	30
2.2.2 La mise en place des étapes recommandées	31
2.2.3 Le respect des critères proposés par les GsG.....	31
2.3 Les outils d'évaluation des guides	33
2.3.1 Les outils d'évaluation destinés aux chercheurs et aux organisations.....	34
Portrait des instruments.....	34
Évaluation des instruments	37
2.3.2 Les outils d'évaluation destinés au praticien pour décider s'il peut retenir ou non un guide	38

2.4 Tendances et questions de l'heure pour les guides de pratique.....	38
2.4.1 Informatique et technologies de l'information (TI).....	38
2.4.2 Mises à jour et coûts	39
2.4.3 Équilibre simplicité/complexité, standardisation/flexibilité	39
2.5 Les guides en SST	40
3 POSITIONNEMENT SUR LES GUIDES ET FACTEURS QUI ONT UN IMPACT SUR LEUR QUALITE	43
3.1 Positionnements sur les guides.....	43
3.1.1 Critiques et demandes	43
3.1.2 Craintes et inquiétudes des usagers.....	46
3.1.3 Désirs et souhaits	47
3.2 Facteurs qui ont un impact sur la qualité des guides	47
3.2.1 Date	48
3.2.2 Organisation impliquée.....	48
3.2.3 Objet du guide.....	49
3.2.4 Longueur du guide	49
3.3 Opinions sur les guides en SST.....	50
4 UTILISATION ET IMPACT DES GUIDES SUR LA PRATIQUE.....	51
4.1 Sources d'information privilégiées par les usagers.....	51
4.2 Portrait de l'adhésion et de l'adoption des recommandations	52
4.3 Barrières à l'utilisation et incitatifs.....	52
4.3.1 Connaissances nécessaires à l'utilisation des guides	53
4.3.2 Clarté et spécificité	53
4.3.3 Simplicité et convivialité	54
4.3.4 Observabilité	54
4.3.5 Accessibilité et environnement.....	54
4.3.6 Crédibilité des développeurs ou des utilisateurs.....	54
4.3.7 Facteur culturel	55
4.3.8 Portrait global.....	55
4.4 Impact des attributs des recommandations.....	56
4.5 Utilisation des guides en SST	56
5 IMPLANTATION DES GUIDES	57
5.1 Définitions et concepts	57
5.2 Interventions possibles.....	58
5.3 Pratiques évaluatives	58
5.4 Efficacité et coût des interventions.....	59
5.5 Évaluation du potentiel d'implantation.....	60
5.6 L'implantation des guides en SST	Erreur ! Signet non défini.
6 BILAN COMPARATIF SANTÉ VS SST ET PISTES À EXPLORER.....	61
7 PERSPECTIVES ET RÉFLEXIONS.....	65
CONCLUSION	69
BIBLIOGRAPHIE.....	71
ANNEXES	89

Annexe 1.1 – Survol du rôle des États dans la production des guides à l’aube des années 2000 et de quelques acteurs clés.	90
Annexe 2.1 - Les attributs des guides proposés par l’Institute Of Medicine et les ajouts proposés par divers groupes de travail	92
Annexe 2.2 – Les instruments d’évaluation des guides	94
Annexe 2.3 - Tableau compilatif des dimensions, critères ou items proposés pour évaluer les guides développés.....	101
Annexe 2.4 - Attributs des guides de pratique et d'évaluation des traitements de l’American Psychologists Association (2002)	109
Annexe 4.1 - Principaux facteurs ou déterminants du transfert (Faye et al. 2007)	111
Annexe 5.1 - Termes de transfert appliqués aux guides (Faye et al. 2007).....	112
Annexe 5.2 - Résumé des approches destinées à changer les pratiques (Grol 1997)	113
Annexe 5.3 - Guide Line Implementation Appraisal (Shiffman et al. 2005)	114

LISTE DES TABLEAUX

Tableau 2.1 - Les attributs et critères que les guides de pratique devraient posséder	19
Tableau 4.1 - Sources d'information utilisées par les cliniciens (Hayward et al. 1997).....	51

LISTE DES ENCADRÉS

Encadré 1.1 - La croissances de la littérature sur les guides et de la littérature médicale	6
Encadré 2.1 - La formulation des recommandations par NICE.....	28
Encadré 2.2 - Pourcentage des 279 G confirmes aux standards recommandés (Shaneyfelt <i>et al.</i> 1999)	32
Encadré 2.3 - Évaluation de 51 guides sur le cancer (Harpole et al. 2003)	32
Encadré 2.4 - Évaluation des chapitres du Guide de l'American College for Occupational and Environmental Medicine (Cates <i>et al.</i> 2006)	33
Encadré 2.5 - Variantes proposées pour les items portant sur le format et la formulation des énoncés.....	36
Encadré 2.6 - Items de l'instrument de Cluzeau repris par au moins cinq des dix instruments analysés	36
Encadré 2.7 - Propositions d'énoncés pour évaluer directives et recommandations.....	37
Encadré 2.8 - Liste de contrôle proposée pour permettre aux praticiens d'évaluer un guide	38
Encadré 4.1 - Les obstacles ou les facilitateurs de l'adhésion-adoption	53
Encadré 4.2 - Obstacles et incitatifs à l'utilisation des guides	55
Encadré 4.3 - Attributs des recommandations	56
Encadré 5.1 - Étude de Christie (2003) sur les pratiques d'évaluation : points saillants	58

INTRODUCTION

La prévention des risques professionnels est un enjeu majeur pour la santé des individus et la performance des organisations. Elle passe, entre autres, par une meilleure application des connaissances développées dans le domaine. Ces connaissances nécessitent, pour que les milieux se les approprient, une mise en forme opérationnelle, dont les guides sont une manifestation importante. Intégrant les savoirs d'expertise et d'expérience, Mercier (2010) considère d'ailleurs les guides comme représentant la forme la plus achevée de valorisation des connaissances.

Comme les recherches documentaires ont rapidement mis en évidence l'abondance des travaux menés sur les guides dans le domaine de la santé, où non seulement le nombre de guides publiés a explosé au cours des deux dernières décennies, mais où aussi la littérature publiée sur le sujet, a elle-même crû de façon exponentielle, nous avons décidé de nous y centrer pour en restituer la logique et la cohérence. Par conséquent, bien que nous sachions que d'autres domaines sont de grands utilisateurs de guides, comme le domaine social, nous avons délimité nos efforts à celui de la santé (médical pourrait-on ajouter).

L'objectif principal de cette étude est de faire le point sur les écrits publiés sur le développement, l'utilisation, l'évaluation, l'implantation et les impacts des guides en santé afin d'en dégager les enseignements clés et de déterminer ce qui pourrait s'appliquer à la SST. Plus spécifiquement, nous avons cherché à répondre aux questions suivantes :

Quels sont les grands enjeux et les principales perspectives de développement adoptées par les guides?

Que connaît-on sur le développement des guides et leur évaluation?

Comment évalue-t-on les guides?

Que sait-on sur leur utilisation et leur impact?

Comment s'organise leur implantation? Qu'est-ce qui fonctionne bien ?

Dans ce rapport, la littérature recensée porte en fait surtout sur les guides de pratique clinique (*guidelines for clinical practice*) ou *guidelines*. Ces guides ont un caractère relativement directif qui correspond au sens étymologique premier du terme guide et que conserve le terme *guideline*, « une règle ou un principe qui fournit des conseils pour mener à un comportement adéquat »¹. On peut traduire le terme en français par ceux de recommandations ou de directives. Les guides de pratique clinique sont constitués d'un ensemble de *guidelines*. Ils suggèrent, par le biais d'énoncés ce qui devrait être décidé ou fait dans la pratique compte tenu des connaissances scientifiques établies. La notion de «bonne pratique» y est implicite. Sans doute ne représentent-ils qu'une fraction des 10 000 guides de pratique de toutes sortes auxquels réfèrent Beaulieu *et al.* (2001), mais ils sont au cœur d'une abondante littérature et de propositions concrètes. Les différents termes seront revus au chapitre 1, mais déjà le lecteur familier avec les guides développés en SST sait que si certains guides ont des objectifs très prescriptifs, d'autres sont de nature plus suggestive ou informative que directive.

¹ Word reference dictionary. Hhttp://www.wordreference.com/definition. Cité in Schünemann 2006H.

En fait, les guides en SST visent de multiples objectifs : outiller les utilisateurs pour reconnaître les risques susceptibles d'altérer leur santé ou de porter atteinte à leur sécurité, cibler les éléments à transformer, choisir les préconisations adéquates, etc. La conception d'un guide sur les machines ou les convoyeurs ou sur la meilleure méthode de dosage d'un toxique donné côtoie celle sur les guides d'aménagement des horaires ou des espaces, les guides d'évaluation des risques ou de prévention des troubles musculo-squelettiques, comme de santé mentale.

Ce bilan s'adresse avant tout aux chercheurs en SST, pour lesquels il est usuel de s'impliquer dans le développement d'un guide. Par exemple, dans l'enquête de Laroche (2009), le tiers des chercheurs canadiens en SST interrogés rapportaient avoir déjà publié au moins un guide au cours des cinq dernières années. Comme ces derniers ont de multiples intérêts et perspectives, nous avons conçu ce rapport comme un outil de travail où chacun pourrait y puiser des informations, des idées, selon ses intérêts. Ainsi, nous avons laissé autant que possible des traces du travail de compilation des données, car elles pourraient servir de base à des travaux futurs.

Outre l'abondance de la littérature disponible en santé, une autre raison de s'attarder tout spécialement au domaine de la santé est le partage entre les deux domaines d'un objet commun, la santé, et de concepts qui présentent des analogies avec le domaine de la SST : praticiens *vs* préventeurs, patients *vs* travailleurs, transformation des pratiques *vs* prévention/intervention, diagnostic des maladies *vs* des risques. Cette exploration dans la littérature médicale visait à mieux cerner les parentés, mais aussi les différences dans la façon de réfléchir et d'approcher les questions, en l'occurrence, celle des guides. *A priori*, une différence importante est sans doute que la SST tient fortement compte de l'environnement et des cadres matériel et immatériel dans lesquels évolue le travailleur. La santé coexiste avec la sécurité qui intègre des approches fort différentes. Les références à l'épidémiologie en sécurité sont par exemple rarement pertinentes. Un guide sur les machines ou les convoyeurs ou sur la meilleure méthode de dosage d'un toxique donné côtoie des guides d'aménagement des horaires ou des espaces, des guides d'évaluation des risques ou de prévention des troubles musculo-squelettiques ou de la santé mentale, car la SST couvre presque tous les domaines disciplinaires de recherche. Il est probable que certains adhéreront fortement aux points de vue exposés et se sentiront très proches de l'univers santé et d'autres assez loin. Ainsi, bien que le titre parle de transférabilité et que nous chercherons en cours de route à situer autant les différences que les similitudes, sachant la diversité de nos interlocuteurs, c'est avec humilité que nous entreprenons l'exercice.

Étant donné l'abondance de littérature sur les guides, comme nous le verrons au chapitre 1, il était insensé de prétendre à une revue exhaustive. Nous avons plutôt tenté de dégager, à travers les lectures, des points de repère qui peuvent être utiles en SST. Nous avons procédé de façon itérative et organisé la recherche à partir de deux grands axes, l'un centré sur l'objet, soit les guides - Comment les développer? Comment les évaluer? Comment les implanter? - et l'autre, sur les utilisateurs : Quel usage font-ils des guides? Qu'en pensent-ils? Cela change-t-il leur pratique?

Nous avons recherché des articles de synthèse et privilégié les documents d'application élargie par opposition aux documents centrés sur l'application de directives (*guidelines*) spécifiques. Le point de départ (mots clés : *assessment tool/instrument*, évaluation, *guidelines*, guides, etc.) a été l'identification des principaux outils d'évaluation. Ceux-ci nous ont orientés très fortement vers les guides de pratiques cliniques. Le travail effectué a consisté à colliger ce qui est commun à

tous ces outils ainsi que leurs variantes (ajouts, suppressions, nuances). Cette première étape nous a aussi rapidement amenés à constater combien les écrits s’imbriquaient les uns dans les autres (ex. références communes, mêmes documents sources, auteurs communs) ainsi que l’importance de la participation des organisations publiques, voire de l’État lui-même, dans le processus. Il nous a alors semblé important de reconstituer l’historique et le rôle de certaines organisations. Ces historiques ont été retracés à travers des écrits synthèses et les livres publiés sur le développement des guides.

La seconde étape a été de faire le point sur le développement des guides, les guides sur les guides (ou *guidelines* pour développer des *guidelines*, *GsG*), ces derniers s’adressant surtout, sans que l’on puisse dire exclusivement, aux guides de pratique clinique (*guidelines for clinical practice*). Ce travail de développement a été mené essentiellement par de larges groupes de travail - nationaux ou internationaux - sous l’égide d’organisations publiques. Comme pour les instruments d’évaluation, le travail de synthèse a consisté à colliger les principaux *GsG* diffusés afin de dégager ce qu’ils ont en commun, ce qui les différencie et la logique ou la philosophie qui a sous-tendu leur développement. C’est à travers ces lectures qu’il nous est apparu important de faire aussi le point sur les concepts et les approches clés utilisées (ex. les méta-analyses) et d’en restituer la logique et les fondements pour que le lecteur en SST en comprenne les orientations fondamentales. En raison de leur importance en SST, certains éléments ont été retenus, comme les approches de consensus. Ces questions ont été explorées surtout à partir d’articles de synthèse ou de séries d’articles consacrés à ces sujets.

Les questions qui découlent naturellement de cette dernière étape ont été : les *GsG* sont-ils utilisés? Qu’en pensent les développeurs (individuel ou institutionnel) et les utilisateurs? Les guides développés suivent-ils les propositions des *GsG*?

Par la suite, nos efforts ont été centrés sur les écrits portant sur les guides eux-mêmes. Les guides produits sont-ils utilisés? Qu’en pense-t-on? Qu’est-ce qui favorise ou au contraire gêne leur utilisation? Quels sont leurs impacts? Nous avons recherché *a priori* des articles de synthèse qui font le point sur ces questions et privilégié les articles ciblant les guides de pratique d’application élargie par opposition à ceux centrés sur le suivi de directives spécifiques.

La base de données ainsi constituée a permis de recenser aussi un bon nombre d’études portant sur les utilisateurs et les développeurs. Elle a été complétée en fonction des thèmes qui présentaient un intérêt potentiel pour la SST. Par exemple : le rôle des contextes, les écrits qui interpellent sur le savoir inclus dans les guides ou quant aux sources d’information utilisées, les perspectives futures comme l’informatisation des outils de développement. Les pistes ouvertes par ces lectures ont été complétées à partir des références indiquées ou par des recherches plus pointues sur ces questions. Il ne s’agissait pas de faire le point sur chacune de ces questions (nous n’en avons pas les ressources), mais d’inclure des résultats qui permettent d’ouvrir des réflexions. Il s’agit donc le plus souvent de pistes ouvertes à la lecture d’un article et des références afférentes.

La structure du rapport ne suit donc qu’en partie l’organisation de la recherche. Le premier chapitre expose les concepts clés, les approches qui sont au cœur de tout le processus de développement des guides en santé, mais aussi des connaissances. Le second chapitre réunit d’une certaine façon le début et la fin, à savoir les procédures proposées pour développer un

guide de pratique de qualité et en évaluer la qualité. On parle donc des GsG et des instruments d'évaluation des guides de pratique, qui de fait visent à vérifier si les propositions du GsG sont appliquées. Les deux sont donc étroitement imbriqués. Le troisième chapitre est axé sur le guide en tant qu'objet. Il trace le portrait : de la réalité des guides (ex. les procédures sont-elles suivies ?); des grandes critiques qu'on leur adresse ; des caractéristiques des guides qui ont ou pas un impact sur leur qualité ; des opinions des développeurs sur les processus de développement ; des craintes et des appréhensions des utilisateurs.

Les quatrième et cinquième chapitres sont quant à eux centrés sur l'utilisateur des guides, bien qu'il ne soit pas toujours possible de le dissocier de l'objet lui-même. Ils regroupent des études portant sur l'utilisation des guides et sur leur implantation. Dans ce dernier cas, le bilan est surtout méthodologique - approches proposées, découpages de variables, etc. - et les références utilisées sont essentiellement de vastes études de synthèse qui permettent de tracer de grandes conclusions.

À la fin de chaque chapitre, les résultats sont situés par rapport au domaine de la SST, au mieux de notre connaissance. Pour le second chapitre, qui porte sur le développement des guides, les parallèles sont tracés au fur et à mesure des points abordés, vu la longueur du chapitre. Enfin, plutôt que de proposer ou de recommander un outil d'évaluation (ce qui était en partie l'intention initiale) ou une méthode de développement en particulier, nous avons synthétisé au chapitre 6 ce que nous croyons important de retenir pour la SST. Ce bilan synthétique inclut des suggestions. Nous avons opté pour cette approche prudente en raison de la variété des thèmes couverts par les guides en SST et de la nature même des résultats exposés à travers les divers chapitres. Le dernier chapitre présente une réflexion plus globale sur ce qui caractérise fondamentalement le domaine de la SST.

Enfin, comme il nous est apparu sage de laisser au lecteur la liberté d'y trouver les éléments susceptibles de l'intéresser, nous avons inclus de nombreuses notes de bas de page qui apportent des précisions pouvant intéresser des lecteurs spécifiques. Les diverses compilations effectuées sont annexées au présent rapport ou disponibles sur le site web du Réseau de recherche en santé et sécurité du travail du Québec (RRSSTQ).

1 HISTORIQUE, CONCEPTS ET APPROCHES RELIÉS À LA PRODUCTION DES GUIDES

Ce premier chapitre aborde l'historique récent des guides de pratique et précise la terminologie utilisée dans ce rapport. Les approches reliées à la production des guides sont ensuite décrites.

1.1 Survol historique et mise en contexte

Les guides sont utilisés en médecine depuis les années 1930. Il s'agit d'une tradition surtout issue des pays anglo-saxons, en particulier du Royaume-Uni. Cependant, leur production a connu un essor important à partir des années 80 (Woolf 1995). L'élaboration des guides, originellement centrée sur un consensus d'experts, a alors évolué vers une élaboration centrée sur les données de la littérature scientifique, examinée selon des critères précis et le rôle de l'expert, redéfini ou plutôt délimité (ex. combler les lacunes, statuer en cas de controverses). Soulignons ici que le Canada a été un précurseur important dans ce virage, à travers les travaux menés par un groupe d'étude mis en place en 1976 pour se pencher sur la question de l'efficacité des examens périodiques (Canadian Task Force on the Periodic Health Examination 1979). Ce groupe a développé ce qui est considéré comme le ou un des premiers guides fondés spécifiquement sur des preuves et pour lequel il a alors développé un outil de graduation de la qualité de ces preuves (Beaulieu et al. 2001). Ce développement a été étroitement associé à celui de ce qu'on a appelé *evidence-based medicine*, une médecine fondée sur des faits scientifiques démontrés, sur des données dites probantes, issues d'études scientifiques rigoureuses (Beaulieu *et al.* 2001). La définition des procédures et des standards méthodologiques visant à reconnaître ou déterminer les données probantes est donc maintenant au coeur du processus de production des guides, qui sont en corollaire vus par les organismes de santé comme une vaste opération visant au transfert des connaissances issues des recherches médicales. Les procédures pour sélectionner les études et les données, pour en évaluer la qualité méthodologique et pour en faire la synthèse (revues systématiques et méta-analyses) sont ainsi définies de façon rigoureuse.

Les guides sont vus aussi comme des outils d'organisation des connaissances ainsi qu'un moyen privilégié pour standardiser et rationaliser les pratiques, et ce, dans un contexte de croissance des coûts dans le domaine de la santé, dont une partie est attribuée à des pratiques cliniques jugées inadéquates (Woolf 1995). Les guides visent donc aussi à améliorer les pratiques professionnelles, un thème central dans cette littérature. Le développement des guides a aussi pour but de diminuer la variabilité des pratiques - phénomène observé inter- mais aussi intra-organisation - ce qui suscite aussi des inquiétudes en santé (Armstrong *et al.* 2004).

Tout ceci explique sans doute l'importance de la participation des organisations gouvernementales (cf. annexe 1.1) et le mouvement de collaboration internationale mis en place (Ollenschläger *et al.* 2004), lequel n'a pas son équivalent en SST. Plusieurs pays ont même fait des guides de pratique un élément important de leur politique de santé, incluant la mise en place de politiques actives et systématiques d'implantation². Les organisations professionnelles ont aussi fortement participé, en particulier aux États-Unis³. L'énorme production qui en est résultée (cf. encadré 1.1) a par la suite engendré le développement d'outils d'évaluation, dans lequel les organisations publiques ont aussi joué un rôle de catalyseur important.

² C'est en particulier le cas de la Hollande. Plusieurs auteurs majeurs y sont associés.

³ Aux É.-U., les organisations privées et non gouvernementales participent aussi très activement à la production des guides.

À titre d'exemple, en 1989, le Congrès américain amendait le *Public Health Service Act* pour créer l'*Agency for Health Care Policy and Research* (AHCPR) et, sous sa responsabilité, le *Forum for Quality and Effectiveness in Health*. Le mandat de l'AHCPR était de concevoir des activités dans le but d'améliorer la qualité et l'efficacité des soins de santé, entre autres par la création, la diffusion et l'évaluation de guides. Pour accomplir cette mission, l'AHCPR s'est adressé à l'IOM (Institute of Medicine), qui à son tour a mis en place un comité d'étude avec mandat de fournir une assistance technique sur différentes questions : définitions, spécifications des attributs des guides, implantation et évaluation. L'IOM va être le fer de lance dans l'élaboration d'un premier guide sur les guides, le « *Guidelines for clinical practice: from development to use* » dont la première édition est parue en 1990 (auteurs : Lohr et Field). L'IOM a aussi mis au point, en 1992, un instrument destiné provisoirement à l'évaluation des guides de pratique clinique structuré autour de 8 attributs⁴ regroupant 142 items. Quatre attributs portent sur le contenu (validité, fiabilité, applicabilité et flexibilité clinique) et quatre sur le processus de développement et la présentation (clarté, multidisciplinarité, mise à jour et documentation). Cet instrument, qui s'est avéré trop lourd et trop difficile d'application demeure cependant à la base des développements ultérieurs.

Encadré 1.1 - La croissance de la littérature sur les guides et de la littérature médicale

Littérature sur les guides

Croissance du nombre d'articles (Medline) portant sur les guides de pratiques : 444 en 1993 à 4 975 en 2006 (Turner *et al.* 2008).

Croissance du nombre de citations (Medline) sur *practice guideline* de près de zéro en 1990 à près de 2 000 (Marshall 2000).

Guides chez les praticiens : 855 guides collectés à la suite de visites effectuées auprès de 65 praticiens (Hibble *et al.* 1998).

Guides et protocoles destinés aux programmes d'audits régionaux : 2 000 comptabilisés par Stern et Brennan (1994) (cité par Jackson et Feder 1998).

Aux É.-U. : 1 500 guides de pratique avaient déjà été développés en 1993 (Woolf 1995) ; 40 sociétés médicales ont établi des comités pour élaborer des guides.

Littérature médicale

Nombre d'articles annuels : deux millions publiés dans plus de 20 000 revues, et ce, il y a déjà 15 ans (Mulrow 1994).

Croissance du nombre d'articles annuellement indexés (Medline) : en 13 ans, de 413 375 à 727 546 (Turner *et al.* 2008) ; en 1995, Jobson et Potter en recensaient 30 000 par mois (360 000/an).

1.2 Terminologie de base

Étymologiquement⁵, le mot **guide**, apparu en 1370, est un emprunt de l'ancien provençal *guida* (XII^e siècle) « celui qui conduit » et du déverbal de *guidar* « conduite » (XI^e) et de l'italien *guida* (vers 1300), déverbal de *guidare* « accompagner quelqu'un ». Les deux termes, qui ont sensiblement le même sens, sont dérivés du gotique *widam*. Le terme guide a conservé depuis son sens étymologique de personne qui conduit au sens propre et figuré (XV^e). Par analogie, guide se dit (XVI^e) d'« un principe directeur qui inspire quelqu'un » (1534). Par extension, il s'emploie pour un ouvrage qui renferme des conseils d'ordre moral, spirituel (1552). À la même époque, il apparaît aussi dans son sens moderne d'« ouvrage destiné à aider » (ex. en donnant des informations sur un pays). On retrouvera par la suite le terme appliqué à des objets (guide dans le sens de rênes, pour diriger un mouvement en menuiserie, etc.).

Aujourd'hui, le mot guide revêt toujours plusieurs sens, notamment celui d'un « ouvrage destiné à aider par des informations générales ou pratiques »⁶ ou d'un « ouvrage contenant des

⁴ Adoptés par la suite par l'AHCPR.

⁵ Dictionnaire historique de la langue française (2000) sous la direction de A. Rey.

⁶ Grand Robert.

informations pratiques ». Toutes les définitions montrent l'orientation utile, pratique, voire pragmatique prise par les guides, mais elles peuvent être reliées à des finalités différentes : expliquer l'usage d'un équipement, indiquer les conduites à tenir advenant certaines circonstances précises, présenter la démarche à suivre pour réaliser une activité déterminée, etc. Le terme guide a donc un double sens de direction et de regroupement d'informations pratiques.

Dans la langue anglaise, il est fait état beaucoup plus du mot **guidelines** qui est défini par le dictionnaire anglais-français des sciences médicales et paramédicales⁷ comme des « principes directeurs » ou des « directives » ou des « recommandations ». Le mot **guidelines** est considéré comme « [...] *outlines or directions on how to conduct or control some program* »⁸ ou comme « *a rule or principle that provides guidance to appropriate behaviour* »⁹. Vu sous cet angle, le mot **guidelines** est associé à la notion de **bonnes pratiques**. Ce qui englobe et signifie beaucoup plus de choses que le mot guide utilisé en français.

En français, le mot guide peut être suivi de l'adjectif pratique, compris dans le sens « qui s'applique aux situations concrètes », ou du complément « de pratiques » qui recouvre partiellement le sens du mot **guidelines**. Ainsi, la notion de **guide de pratique** est aussi liée au concept de bonnes pratiques ou de meilleures pratiques, lesquelles sont définies comme étant « des savoirs ou une manière de faire qui, dans une organisation, conduisent au résultat souhaité et qui sont portées en exemple auprès des pairs afin de leur faire partager l'expérience qui permettra une amélioration collective »¹⁰. Le concept revêt, ainsi, un caractère polysémique et peut se définir de plusieurs manières selon son objet (aide à la décision, transfert des connaissances, réponse à une question d'éthique) ou ses utilisateurs potentiels (INESSS 2011).

Les **guides de pratiques cliniques**¹¹ quant à eux sont définis comme étant des **énoncés systématiquement** conçus pour aider les praticiens et les patients à prendre les **décisions** quant au choix de soins appropriés dans un **contexte** donné¹². Bien entendu tous les guides de pratiques ne sont pas cliniques et leurs définitions dans les divers domaines de la santé sont un peu plus larges, mais elles reprennent toutes sensiblement les mêmes éléments. Par exemple, selon l'Organisation mondiale de la santé (2003), les GdP : « *are systematically developed evidence-based statements which assist providers, recipients and other stakeholders to make informed decisions about appropriate health interventions. Health interventions are defined broadly to include not only clinical procedures but also public health actions. Guidelines are formal advisory statements which should be robust enough to meet the unique circumstances and constraints of the specific situation to which they are being applied* ». La Fédération québécoise des centres de réadaptation en déficience intellectuelle et en trouble envahissant du développement les définit comme « un ouvrage décrivant la manière d'exercer une activité dans un contexte spécifique s'appuyant sur les **valeurs** et les **principes** reconnus. Il guide les choix

⁷ Dictionnaire anglais-français des sciences médicales et paramédicales. 4^e édition, 1996.

⁸ Dictionary and thesaurus of environment, health and safety US Department of Energy. Office of Environment, Safety and Health. CRC Press inc. (1992).

⁹ Word reference dictionary. <http://www.wordreference.com/definition>. Cité par Schünemann *et al.* (2006).

¹⁰ Dictionnaire illustré des activités de l'entreprise français-anglais. Presses internationales Polytechnique. Montréal, 2008.

¹¹ Définition proposée par l'Institute of Medicine : « *Clinical guidelines are systematically developed statements designed to help practitioners and patients make decisions about appropriate health care for specific circumstances* » (Field et Lohr 1992).

¹² Au Québec, les GdP sont définis comme des « recommandations formulées de façon systématique pour soutenir les praticiens et les patients dans leurs décisions concernant les soins à prodiguer selon les symptômes cliniques » (INSPQ 2011).

d'intervention, détermine les actions pertinentes et **acceptables** et inclut des recommandations visant l'amélioration des services. Il est généralement produit par des comités **d'experts** et devrait contenir des standards de pratiques basés sur des **données probantes**. Il doit être régulièrement **révisé** sur la base des nouvelles données disponibles » (Mercier 2010). Cette dernière définition recouvre des notions (valeurs, principes reconnus, acceptabilité, données probantes, mise à jour) traitées dans les GsG que nous verrons au chapitre suivant.

Dans le domaine de la santé et de la sécurité du travail, l'Organisation internationale du travail a adopté un point de vue proche de celui de la santé. Les GdP appelés tantôt « recueils de directives pratiques », tantôt « principes directeurs » constituent un moyen de plus en plus recouru par l'Organisation pour promouvoir des bonnes pratiques¹³. Ils peuvent être présentés dans des formats différents : protocoles, algorithmes, déclarations de consensus, recommandations de comités d'experts, etc. (OMS, 2003).

Une autre expression souvent rencontrée est celle de **standard de pratique**¹⁴ : « Un ensemble de normes, de pratiques et de règles qui définit une approche **systematique** commune. Le standard de pratique s'inscrit dans un plan d'amélioration de la qualité, il fixe les critères de **performance** et s'appuie sur les données probantes disponibles. Il doit nécessairement comprendre des indicateurs permettant une **évaluation** ».

Le terme **evidence** est un concept qui fait référence aux savoirs produits par la recherche, plus particulièrement à partir d'études dont la rigueur permet d'en assurer la validité. Les termes utilisés en français sont ceux de preuves ou de données probantes (ex. utilisé par la Fondation canadienne de la recherche sur les services de santé, le FCRSS). Différentes expressions en sont dérivées, telles *evidence-based medicine (EBM)*, *evidence-based individual decision making*, *evidence-based guideline*¹⁵. Le concept d'*EBM* a été introduit par G. Guyatt en 1991 (Sackett 1996), mais on en crédite l'origine idéologique à A. Cochrane (1908-1988) dont on a donné le nom aux centres de recherche qui publie et diffuse sur son site des revues systématiques sur les données probantes (Ulvenes *et al.* 2009).

Les **données probantes** sont définies par la FCRSS comme étant « les renseignements qui se rapprochent le plus des faits d'un sujet. La forme qu'elles prennent dépend du contexte. Les résultats de recherches de haute qualité, qui reposent sur une méthodologie appropriée, sont les données probantes les plus précises. Comme les recherches sont souvent incomplètes et parfois contradictoires ou non disponibles, d'autres catégories de renseignements sont nécessaires pour les compléter et les remplacer. »¹⁶ Une définition plus courte proposée par Lomas *et al.* (2005) est « tout élément qui permet d'établir un fait ou donne raison de croire quelque chose. » Conceptuellement, il s'agit pour ces derniers d'une **connaissance qui est explicite** (c'est-à-dire codifiée et propositionnelle), **systemique** (qui utilise des méthodes transparentes et explicites de codification) et **répétable** (c'est-à-dire qu'en appliquant les mêmes méthodes aux mêmes

¹³ Le GdP le plus connu à cet égard est sans doute : « Principes directeurs concernant les systèmes de gestion de la sécurité et de la santé au travail ILO-OSH, 2001 ».

¹⁴ Autres termes utilisés pour référer aux pratiques : *practice policy*, *practice parameters*, *clinical pathway*, *standard* (voir Berg *et al.* 1997).

¹⁵ Le site du Département de médecine familiale de l'Université Laval est cité par Graham *et al.* (2002) comme étant une bonne source d'information à consulter sur la qualité des sites web portant sur la *EBM*. (www.medicine.quebec.qc.ca/default.htm). Il existe une trentaine de sites web sur les guides.

¹⁶ http://www.chsrf.ca/other_documents/evidence_f.php (consulté en juillet 2010).

échantillons, on obtiendra les mêmes résultats). La valeur des données probantes est donc intrinsèquement liée à la qualité des méthodes qui les produisent. Elles sont par conséquent évaluées sur la base de critères méthodologiques. Probant qualifie aussi son caractère d'applicabilité et d'efficacité. Les deux termes - données probantes et preuves - seront utilisés dans les chapitres suivants, selon le contexte.

1.3 Approches reliées au développement des guides

Grol (1997) résume la multiplicité des points de vue à prendre en compte à travers un dialogue fictif sur un problème de taux de césarienne jugé trop élevé. Ce dialogue résume divers éléments de discussion que nous retrouverons associés aux guides.

*« The clinician either denies there is a problem or proposes setting up a well **designed course** to increase clinicians' knowledge and skills.*

*"OK," says the clinical epidemiologist, "but we first need to know what the evidence is on the indications for a caesarean section. We should perform a **meta-analysis** and come up with evidence based guidelines to disseminate among the obstetricians."*

*"No," says the educational expert: "that is a top down approach and such strategies will usually fail. **Form small groups of doctors** and let them **discuss** the problem, using cases and experiences from their own practices as the basis for local arrangements on new routines."*

*"We should take a look at the facts first," says the health services researcher. "Let us set up a **multicentre audit** first and **collect data on actual variation between hospitals** and include data on casemix. Feeding this information back to the hospitals will probably stimulate improvement."*

*"You are all focusing too much on the individual doctor," says the management expert. "The problem is not the doctor, but the **system**. We should analyse the **process of decision making** and performing the caesarean sections and see what structures determine the process. Next we need a quality improvement team."*

*"This is all too much talking," says the representative of the health authorities. "Doctors are sensitive only to what happens to their budgets. We need to put a pressure on them to limit the number of caesarean sections per hospital, give hospitals a reasonable budget, and provide the obstetricians with an **incentive** when they reduce the rate."*

Woolf (1990) distingue quatre approches de développement des guides, chacune visant à répondre aux critiques adressées à la précédente : 1) par consensus informel ; 2) par consensus formel ; 3) centrée sur les données probantes ; 4) l'approche explicite.

L'approche la plus ancienne est le consensus informel. La seconde formalise le processus de consensus, qui se construit alors selon des règles précises, où le consensus est quantifié ou coté. Le but est de limiter les biais potentiels issus d'un consensus informel. La troisième approche, développée à partir des années 80, vise à limiter/délimiter la référence à des opinions et à améliorer la méthodologie de construction des guides. Le groupe témoin doit suivre une méthode systématique pour colliger les études pertinentes, revoir les données et évaluer les liens avec les recommandations. La dernière approche, la plus récente, ajoute une phase dont l'objet est d'explicitier les impacts attendus (bénéfice potentiel, dommages possibles et coût des options) ; elle inclut une étape qui précise les préférences des patients pour chaque résultat attendu. Ce processus d'explicitation a été proposé pour répondre aux diverses critiques émises sur la construction des guides à savoir : ne prend en compte qu'un résultat (Berg *et al.* 1997) ; faiblesse de l'articulation entre les données probantes et les recommandations; manque de lien avec le

processus de décision. Il marque en fait le passage du « quoi » au « comment » (Edwards 2010¹⁷) et une plus grande attention accordée aux contextes pratiques (Browman 2001).

Le choix des connaissances et des données probantes étant au cœur du processus d'élaboration des guides en santé, une abondante littérature traite aussi des approches et des méthodes qui permettent de collecter les connaissances, de les évaluer et de les colliger. Elle correspond à la première grande phase du développement d'un guide, soit celle du « quoi » incluse. Cependant, cette étape ne suffit pas et il faut aussi à un moment ou l'autre s'entendre pour évaluer, interpréter, combler les lacunes et établir des liens (Giacomini et Cook 2000). Les méthodes de consensus ont donc acquis un statut méthodologique important. Enfin, l'intérêt grandissant pour le « comment » a amené les auteurs à s'intéresser aux approches qualitatives. Nous présentons donc maintenant un survol de ces questions où nous y situons les enjeux, les principes clés, les limites et les critiques des modèles et méthodes proposées pour collecter, interpréter et traduire les données probantes en recommandations.

La première section résume les grands modèles utilisés, la deuxième section les deux grandes méthodes de sélection des données probantes que sont les revues systématiques et les méta-analyses. La section suivante traite de l'intégration de ces approches dans les guides. La quatrième section présente une classification des méthodes qualitatives (méthodes phénoménologiques, théorisation ancrée et études ethnographiques) et la dernière section fait le survol des principales méthodes de consensus utilisées (Delphi, technique de groupe nominal, conférence par consensus).

1.3.1 Modèles pour colliger les données

Audet *et al.* (1990) distinguent trois modèles dont les différences résident dans la façon de choisir les données qui seront à la base du guide. Dans le premier modèle, les données sont issues d'une recherche systématique où les faits proviennent essentiellement de la littérature, selon une procédure déterminée¹⁸. Le résultat fait l'objet d'un « *position paper* » qui est soumis à l'évaluation d'experts externes et que l'on fait circuler pour y incorporer des opinions des praticiens et autres spécialistes concernés. La position « publiée » est ensuite ouverte à une période de commentaires. Le second modèle repose avant tout sur le consensus d'un groupe d'experts internes¹⁹ ; les ressources d'experts externes sont utilisées pour compiler et synthétiser la littérature scientifique. Le troisième modèle combine revue systématique et consensus d'experts, ce dernier étant obtenu à travers des procédures formalisées. Les auteurs concluaient qu'aucune étude ne permettait de démontrer quel était le meilleur modèle de développement.

¹⁷ L'auteur fait référence aux guides dans le secteur du travail social.

¹⁸ Le premier modèle est à la base du *Clinical Efficacy Assessment Project* développé par l'American College of Physicians.

¹⁹ Audet *et al.* (1990) citent en exemple le *Clinical Guidelines Program* développé par le Harvard Community Health Plan. Le degré avec lequel on utilise des techniques formelles de consensus pour établir la synthèse varie. Cette synthèse implique des praticiens, des experts locaux, des gestionnaires de clinique qui sont alimentés par des documents et des résumés. Le résultat prend la forme d'un algorithme.

1.3.2 Méthodes d'identification des données probantes : revue systématique et méta-analyse

a) Principes clés

Une revue systématique ou formelle de la littérature scientifique en santé procède par étapes précises (Greenhalgh 1997²⁰, Sackett 1995). Woolf (1992) la définit ainsi : « *An effective scientific technique to identify and summarize evidence on the effectiveness of intervention and to allow the generalisability and consistency of research findings to be assessed and data inconsistencies to be explored* ». On attend de cette revue qu'on y précise les objectifs, la stratégie de recherche des données, les critères utilisés pour inclure ou rejeter les études, une description de la méthode utilisée, une procédure pour graduer la qualité (ou la force [*strength*] de la preuve) des données. Les critères d'évaluation suggérés sont de deux ordres. Le premier groupe porte sur le contenu informatif de l'article et sur sa qualité descriptive. Comme ce premier travail de tri s'effectue à partir des résumés, des propositions ont aussi été développées pour en standardiser le format (ex. Hayward *et al.* 1993). Le second groupe définit les critères pour évaluer la qualité de l'article²¹. Ceux-ci portent généralement sur la méthode d'étude utilisée. Comme les revues combinent des résultats de plusieurs études, la méthode inclut aussi une procédure pour évaluer la qualité des données ou la force des « preuves ». Armstrong (2003) propose trois niveaux selon que les preuves s'appuient sur : 1) plusieurs essais cliniques randomisés ; 2) un essai randomisé et plusieurs études non randomisées ; 3) le consensus d'experts. Les *textbooks* ou ouvrages de référence, en général, et les articles de synthèse publiés ne se qualifient donc pas en tant que revue scientifique systématique ou « *scientific overview* »²².

La revue systématique prend donc valeur de technique scientifique. Certains, par exemple, Grimshaw *et al.* (1995) considèrent que seul le caractère systématique d'une revue permet un haut niveau de qualité scientifique. Une revue non systématique vaut ainsi une cote moyenne, et le consensus d'experts, une cote faible. Selon Mulrow (1994)²³, la revue systématique présente plusieurs avantages. D'abord, elle peut être combinée à un processus de mise à jour continu qui pourrait permettre de diminuer le délai d'intégration des nouvelles connaissances. Comme le délai moyen d'introduction des données sur les traitements les plus efficaces serait de l'ordre de dix ans (Cook *et al.* 2004), l'enjeu est important. En second lieu, la revue systématique permet de statuer sur le caractère généralisable des observations (*findings*). Et finalement, le recensement de multiples études permet de prendre en compte le rôle du contexte et d'évaluer la cohérence des relations, afin d'en expliquer éventuellement les inconsistances. Dans les approches où elle

²⁰ L'auteur a publié une série d'articles en 1997 dans le *BJM* sur le thème de « *How to read a paper* » dont *Statistics for the non-statistician* (315 : 422); *Papers that summarise other papers/systematic reviews and meta-analysis* (315 : 672); *Papers that goes beyond numbers/qualitative research* (315 : 740).

²¹ La littérature sur le sujet est abondante. Guyatt *et al.* 1993, 1994, 1995; Jaeschke *et al.* 1994; Oxman *et al.* 1993. Beaucoup d'articles s'adressent aux praticiens et aux professionnels. Les praticiens en SST n'ont généralement pas ce rapport à la littérature scientifique.

²² Il n'y a pas d'équivalent en français. L'*overview* n'équivaut pas à une revue systématique, même si elle s'en approche. Hayward *et al.* (1993) la définissent ainsi : « *A type of review in which primary research relevant to a question is examined and summarized, and an effort is made to identify all available literature that pertains to this question.* » ... *A systematic overview* est défini dans la base de données sur l'*EBM* de l'Association médicale canadienne comme « *a formal review of a focused clinical question based on a comprehensive search strategy and structured critical appraisal.* » (<http://www.ebem.org/definitions.html>)

²³ L'auteur avait noté, dans une étude précédente portant sur l'analyse de 50 articles, que seulement un décrivait clairement la méthodologie, que peu (3) proposaient une synthèse quantitative. L'auteur se montrait particulièrement critique quant aux méthodes de synthèse utilisées (Mulrow 1987).

est au cœur de la démarche, les propositions sur le contenu de ces revues peuvent être encore plus élaborées et aller jusqu'à en définir le format de présentation. Le but est d'arriver à des procédures standardisées²⁴. Plusieurs articles décrivent ces méthodes (voir Greenhalgh 1997).

Une **méta-analyse** est une revue qui incorpore des stratégies quantitatives²⁵ pour obtenir une estimation globale. Une méta-analyse combine des données de diverses études. Elle accroît donc la puissance des études puisque l'on peut combiner les effectifs, ainsi que la précision dans l'estimation des risques (Mulrow 1994). Que l'on procède par méta-analyse ou par revue systématique, les auteurs doivent préciser les critères utilisés pour trier les études qui seront retenues puis en estimer la valeur.

Globalement, le point de vue est surtout épidémiologique²⁶, et fait peu référence aux concepts développés en recherche évaluative (cf. chapitre 5.3).

b) Limites et critiques

Ces approches qui consistent à tout recueillir, puis à classer, à trier et à évaluer sont coûteuses en temps et en ressources (Browman 2001). On estime que le développement d'un guide prend de un à trois ans et qu'il en coûte de 100 000 \$ à 800 000 \$ (Woolf 1995), dont une grande partie est reliée à l'étape de revue systématique. Ces coûts élevés ont par ailleurs amené diverses organisations et auteurs à promouvoir la mise en commun des ressources et le développement d'adaptations nationales ou locales. Les GsG y consacrent maintenant une étape.

La stratégie principale utilisée pour réduire ces coûts consiste à prendre des raccourcis en utilisant des critères restrictifs pour autant qu'ils soient systématiques. Ces critères sont généralement la langue (ex. ne retenir que les articles publiés en anglais), l'année de publication (ex. moins de cinq ou dix ans), l'appartenance à une base d'articles ou à un type de publication (ex. revue avec comité de pairs). Ces critères introduisent un biais (Browman 2000), auquel s'ajoutent divers autres biais de publication comme la tendance à publier surtout les résultats positifs. À cet égard, il est intéressant de signaler que la publication par *l'American Psychologist Association* d'un modèle de développement²⁷ a entraîné une vague de contestations sur cette façon d'établir quels traitements étaient efficaces (Reed *et al.* 2002). Les approches systématiques finissent par éliminer aussi un ensemble de données (Marshall 2000). Enfin, les études montrent qu'il y a des discordances de conclusions selon que l'on agrège les études/essais ou les données sur les patients eux-mêmes (cette critique fait référence aux méta-analyses).

La critique sans doute la plus importante est que ces données sont généralement spécifiques à une situation précise, et qu'elles ne prennent pas en compte d'autres éléments comme la comorbidité ou la situation du patient. Les « preuves » sont en quelque sorte trop spécifiques à un type de patient, une situation. Des étapes ont d'ailleurs été ajoutées au processus de

²⁴ Ex. l'American College of Physicians Clinical Efficacy Assessment Subcommittee a développé une position sur le contenu de ces synthèses. La proposition inclut une phase de revue par des experts externes et des "content experts" : [...] *Both papers then undergo an extensive outside review incorporating opinions of practitioners and reviews by all relevant specialty and subspecialty societies. Finally, if the papers are published in the Annals of Internal Medicine, then they are opened to the Letters section, a process that solicits alternative views and disagreements on the subject.*

²⁵ Il existe des méta-analyses qualitatives dont le principe est assez différent. Il s'agit ici uniquement de méta-analyses quantitatives.

²⁶ Par exemple, le GsG neo-zélandais suggère d'utiliser le General Appraisal Tool For Epidemiology pour évaluer les preuves.

²⁷ Portait sur les interventions pour les problèmes de santé mentale et les aspects psychosociaux de problèmes de santé physique.

développement des guides pour pallier ces limites. Une critique corollaire est que les analyses systématiques basées sur des données quantitatives mènent à des dérives liées à la notion de « moyenne » (patient moyen, réponse moyenne...) (New Zeland Guidelines Guide 2001). Comme le résumant Cook *et al.* (1997) : « *Clinicians reason about individual patients on the basis of analogy, experience, heuristics, and theory as well as research evidences. Awareness of treatment's effectiveness does not confer knowledge about how to use that treatment in caring for individual patients* ». On pourrait remplacer les termes de patient et traitement par ceux de poste, travailleur ou intervention et nous nous y reconnaitrions très bien.

La critique souligne aussi le fait que s'appuyer essentiellement sur l'expertise externe et nationale (ou internationale, selon le contexte) revient à questionner le jugement du praticien confronté à une situation réelle qui, elle, est locale : « *What appears to be an implicit faith in the "better" judgment of academic experts and the utility of randomized clinical trials needs to be questioned and tested.* » (Audet *et al.* 1990). Cette critique a amené les développeurs à accorder plus de place aux praticiens dans l'élaboration des guides. La nécessité de prendre en compte les contextes²⁸ est maintenant bien reconnue et les écrits sur l'adaptation des guides se sont multipliés. Les GsG prévoient une étape d'adaptation.

Enfin, plusieurs critiques s'adressent à l'*Evidence based-medicine*. Les principales sont le ton évangélique adopté par ses tenants, l'absence de démonstration claire que les données issues des études aléatoires présentent moins de biais que les autres designs ou améliorent la santé des patients, la *scientification* de l'art de la médecine (Ulvenes *et al.* 2009). Cependant, l'enquête menée auprès de 966 praticiens norvégiens²⁹ par ces auteurs montre au contraire que les praticiens ont une opinion globalement très positive sur le caractère aléatoire des études, et systématique des revues. Toutefois, les sources d'information privilégiées quand il s'agit de prendre une décision ne sont pas les revues scientifiques systématiques (Hayward *et al.* 1997).

1.3.3 Classification des méthodes qualitatives de recherche

La place de la recherche qualitative est rarement abordée de front. Cependant, le New Zeland Guidelines Group (2001) s'est attardé à la question. Il les regroupe en trois classes³⁰ : méthodes phénoménologiques (les données - connaissances proviennent d'expériences rapportées par la personne), les théories ancrées (*grounded theory* ; la méthode vise à développer une théorie qui s'articule à partir de la réalité rapportée par les participants), les études ethnographiques (elles sont décrites par leur objet : ex. comprendre comment les gens interprètent leurs expériences et modulent leur comportement en conséquence). Les données sont issues de verbalisations, d'observations ou du recueil de traces (ex. écrits). Les différences entre les méthodes ne sont pas exposées en matière de validité, mais de type d'information recueillie. L'extrait suivant résume bien les différences entre les données quantitatives et qualitatives :

There are major methodological differences ... in terms of the type of question asked, sample selection, data collection and analysis and reporting the results. Qualitative study questions are designed to enquire about how

²⁸ Aux chapitres 2 et 3, le lecteur trouvera plusieurs exemples traitant du rôle du contexte : ressources disponibles, ressources des patients, compétences, support, etc.

²⁹ On demandait aux médecins leur accord (échelle Likert à cinq échelons) par rapport à un énoncé. Ont répondu oui (valeur 4,5) : pour l'*EBM*, les études randomisées sont importantes (87 %), l'*EBM* est fondée sur des revues systématiques indépendantes (77 %), utilise l'expertise des médecins (53 %) et prend en compte les valeurs des patients (53 %).

³⁰ Source utilisée : McKibbin *et al.* (1999) *Evidence-based principles and practice*.

people feel or experience certain situations. The number of people studied in qualitative studies is generally much lower as the focus is on themes rather than effective treatments and adverse events. Participants are often chosen on the basis of their knowledge or experience of the content area (known as an iterative approach) whereas in quantitative research samples are often chosen to be representative of the population, generally by using a random sampling method. The main methods to collect qualitative data are case studies, participant observation, in depth interviews and focus groups. Whereas the unit of analysis in quantitative studies is primarily numbers, the unit of analysis in qualitative studies is a thought, a concept or a theme. Study findings are usually presented in the form of detailed narratives that describe these themes and attempt to analyse how the themes are interrelated. Discourse analysis is another way to analyse the data. (P. 39)

1.3.4 Les principales méthodes de consensus utilisées

Il existe un ensemble de méthodes de consensus. Le rapport de Murphy *et al.* (1998) en présente une excellente synthèse rédigée sur une perspective d'application au développement de guides³¹. Comme ces derniers le soulignent, la littérature sur le sujet en santé est très limitée. Ce sont plutôt les sciences sociales qui abordent ces questions. En santé, le consensus est surtout utilisé pour les questions controversées. C'est un des rôles que l'on réserve à l'expert comme on l'a vu précédemment où le consensus est utilisé principalement pour la prise de décision. Il permet de faire un meilleur usage des données disponibles. La méfiance à utiliser le consensus pour capturer les connaissances collectives³² vient de ce que le consensus peut aussi capturer l'ignorance. Cette crainte est sous-jacente à bon nombre d'écrits sur le développement des guides et les revues systématiques. La crainte de l'erreur n'est pas aussi omniprésente en SST.

Ces méthodes comportent cinq composantes. La façon de les définir et de les structurer varie cependant d'une méthode à l'autre : 1) formulation/utilisation des questions ; 2) participants : choix, rôles... ; 3) informations transmises ; 4) méthode pour structurer les interactions ; 5) méthodes pour synthétiser les jugements indépendants. Nous reprenons ci-dessous les principales méthodes. Elles comportent trois grands types d'activités qui se distinguent dans la façon de les organiser : la planification, le recueil du jugement individuel et l'organisation des interactions.

La méthode Delphi

La méthode a été développée dans les années 50 par la corporation Rand³³ dans une perspective de prévision - d'où la référence à l'oracle grec - par la synthèse des opinions d'experts. La méthode a évolué et est depuis appliquée à de multiples domaines. Delphi est en partie quantitative puisqu'elle cote les estimations des participants. Elle est basée sur l'envoi d'un questionnaire. Il n'y a pas de face-à-face. Un premier questionnaire sert à déterminer les indicateurs, les indices pertinents et un second, les opinions. Les réponses sont colligées et retournées aux participants. Il s'agit d'un processus itératif intégrant les rétroactions, mais il est

³¹ Le GsG du National Institute for Health and Clinical Excellence (NICE) consacre un chapitre à cette question.

³² La notion de connaissances, d'intelligence collective fait l'objet de nombreux écrits en sciences sociales. Cette notion est considérée importante aussi en biologie animale. Wikipedia en fournit la définition suivante : « L'intelligence collective désigne les capacités cognitives d'une communauté résultant des interactions multiples entre des membres (ou agents). Les éléments portés à la connaissance des membres de la communauté font qu'ils ne possèdent qu'une perception partielle de l'environnement et n'ont pas conscience de la totalité des éléments qui influencent le groupe. Sous certaines conditions particulières, la synergie créée par la collaboration fait émerger des facultés de représentation, de création et d'apprentissage supérieures à celles des individus isolés. L'étude de l'intelligence collective implique aussi l'étude des limites des interactions entre membres d'un groupe, limites qui conduisent à des erreurs collectives parfois catastrophiques. » Pierre Lévy (1994) développe cette question sur l'Internet : *L'intelligence collective. Pour une anthropologie du cyberspace.*

³³ Le Value Health Sciences de la RAND Corporation (organisation privée basée en Californie qui publie souvent des documents en transfert de connaissances : <http://www.rand.org/> . RAND a aussi publié un handbook qui fait le point sur les audits dont Delphi (Ling et Villalba van Dijk 2009).

conçu pour éviter les interactions, vues comme une source de biais potentiel. Cela est considéré comme le point faible de la méthode puisqu'elle se prive des impacts positifs des interactions.

La technique du groupe nominal

Cette méthode a été développée à la fin des années 60 dans un contexte de comité décisionnel. Les idées sont collectées individuellement, puis listées pour être soumises à la table ronde, et complétées. Elles sont ensuite discutées et votées individuellement, ce qui permet de faire des statistiques. Il en existe des variantes, dont celle adaptée par la corporation RAND pour les guides (appelée aussi Delphi modifié). Le rôle du facilitateur y est important. La formule est relativement similaire à celle des décisions par consensus éclairé, utilisée en travail communautaire. Berg *et al.* (1997) suggèrent entre 10 et 20 participants. Globalement, on considère que la fiabilité diminue avec des groupes de moins de 6 ; avec plus de 12 membres, le groupe perd en efficacité, les gains en fiabilité se faisant au détriment de l'ensemble des informations (Murphey *et al.* 1998). La taille suggérée dépend de la diversité à intégrer (ex. spécialistes, praticiens, généralistes). Il existe aussi une version dérivée, utilisée pour les questions simples, dite du *staticised group*, appelée aussi parfois groupe nominal, mais dans laquelle il n'y a pas d'interactions.

La conférence de consensus

Un groupe d'experts sélectionnés est réuni (± 10) pour élaborer un consensus sur une question donnée. Ce groupe reçoit des documents et des données. Le format de discussion demeure informel. Il peut y avoir un public et le rôle du président de session est important³⁴. La formule est relativement similaire à celle utilisée pour les décisions par consensus éclairé.

1.4 Mise en parallèle avec la SST

1.4.1 Les données alimentant les contenus des guides

Quantité de données produites : L'explosion du nombre d'articles publiés en santé et de la littérature sur les guides a imposé la nécessité d'élaborer des procédures de « filtration-condensation ». Ce problème n'a pas son équivalent en SST où nous faisons plus souvent face au problème inverse : il nous faut souvent combler les vides pour lesquels aucune recherche n'a été produite. Adopter une solution issue de la santé pour un problème qui n'a pas de commune mesure en SST ne serait pas judicieux.

Source des données. Dans certains domaines de la SST, la source principale des données est la littérature scientifique et le processus de production de connaissances est proche de celui de la santé. C'est le cas par exemple de l'hygiène industrielle. Dans les domaines où la recherche-action ou la recherche terrain prédominent, les données issues des partenaires de la recherche ont un statut important. Le point de vue en est résolument un de co-construction. Des exemples de cette intégration sont donnés dans l'ouvrage publié au RRSSTQ sur la dynamique du transfert de connaissances (Desmarais et Lortie 2011).

³⁴ Ce format est très utilisé dans le domaine qui s'adresse aux décideurs (*policy-making*). Voir la base de données de la communauté en transfert du Réseau de recherche en santé et sécurité du travail du Québec sous le thème prise de décision/policy making. <http://www.rrsstq.qc.ca/fra/activites.asp>

La notion de ce qui est probant peut différer aussi en SST. Par exemple, dans le domaine de la sécurité, le risque est une combinaison de deux éléments : le potentiel de gravité et l'exposition. On reconnaît que des éléments en mouvement (engrenage, lames) présentent des risques potentiels importants ; les données n'ont pas à être issues de la littérature pour reconnaître la nécessité de développer des procédures de cadenassage. L'exposition est évaluée *in situ*.

Validité et synthèse des données. La façon d'évaluer la qualité des données varie selon le domaine. La qualité du design est un enjeu quel que soit le domaine, mais les règles varient. Tester un équipement a peu à voir avec l'aléatoire. Autant le contrôle des variables est central dans un contexte expérimental, autant il devient un biais sur le terrain qui peut nuire à toute la validité de l'étude. En SST, le contexte tient aussi une place parfois centrale. Autant il apparaît imprudent de généraliser à partir de données contextualisées, autant il apparaît aussi imprudent de généraliser à tout contexte des données obtenues dans un milieu contrôlé. Autrement dit, la question fondamentale parfois soulevée est celle de la validité externe. En santé, cette question est sans doute moins importante. Si les données probantes montrent qu'un traitement donné est efficace, le résultat pourra être modulé par des facteurs externes (ex. ressources disponibles) mais sans que cela remette en question la démonstration d'efficacité elle-même. En SST, ce n'est pas toujours le cas. Par exemple, il est démontré expérimentalement que fléchir les genoux sollicite moins le dos qu'une flexion du dos. Cependant, fléchir les genoux nuit à l'équilibre et gêne la mobilité. Dans un contexte où les surfaces d'appui sont étroites et où les imprévus sont fréquents, cela deviendrait à risque (Denis *et al.* 2011). Des données probantes obtenues dans un contexte expérimental ne le sont plus dans des contextes où d'autres facteurs interviennent.

Les procédures pour **colliger les données** apparaissent aussi trop coûteuses et trop rigides pour la SST. Le processus proposé ne permet pas les allers-retours, les échappées pour poursuivre une piste. Il éloigne de la lecture des articles et de la compréhension de comment un problème est réfléchi. La documentation systématique de tout ce qui est fait devient rapidement lourde et nuit au processus créatif de recherche.

Aussi, dans plusieurs domaines de la SST, les situations et les contextes évoluent très rapidement. C'est le cas en particulier des technologies, constamment en avance sur les sciences de l'ingénierie (ex. nanotechnologies). Les données obtenues sur la base de protocoles lourds et complexes peuvent être rapidement périmées. Par exemple, les problématiques émergentes comme les transformations de la main-d'œuvre (nouvelle génération, travailleurs immigrants) nécessitent des solutions avant que les problèmes ne s'installent. Les données du milieu sont sans doute un reflet plus précis de la situation que ce qui est publié dans la littérature scientifique (Gravel *et al.* 2010).

Contexte : La notion de contexte est plus large en SST. En santé, le contexte réfère surtout au patient et aux ressources. En SST, le contexte réfère à l'ensemble des conditions matérielles et non matérielles. Les données, dans un contexte autre que celui où elles ont été produites peuvent ne plus être probantes. Par exemple, les recommandations sur l'autonomie au travail ne peuvent être opérationnelles qu'en relation avec un contexte spécifique.

Consensus : La pratique du consensus a une longue tradition en SST et il intervient à de multiples moments. Pour notre part, nous croyons donc que c'est en « admettant l'existence de consensus qu'à un moment donné, dans un lieu précis, basé sur un lot de connaissances défini,

ainsi qu'en gardant en mémoire l'évolution rapide des savoirs, [...] [qu'] il est possible de parvenir à une idée et à une pratique de la justice ou de la science, à une légitimité elle-même située spatiotemporellement, limitée » (Laroche 2009, p. 2-3).

1.4.2 Les guides produits en SST

Les guides produits en SST ont de multiples teneurs. Le guide à caractère prescriptif ou directif, proche de l'esprit du guide de pratique clinique, côtoie le guide qui collige des informations de façon plus large. Une liste de près d'une centaine de guides francophones en SST publiés au Québec/Canada, France, Belgique et qui peuvent être consultés en ligne ou téléchargeables gratuitement est disponible sur le site web du RRSSTQ. Ces guides traitent : de prévention, d'évaluation, de plan d'action, d'intervention, d'information, d'utilisation ou de gestion. Ils peuvent ne comporter que quelques pages, comme être longs de plus de 100 pages. Le format énoncé ne peut convenir qu'à une petite fraction de ces guides. Cela est d'autant plus vrai que certains guides misent sur le visuel plutôt que sur le texte. C'est par exemple le choix éditorial du groupe qui a développé le guide sur les convoyeurs (Giraud *et al.* 2004).

Les approches utilisées pour développer les guides sont multiples. Les approches qualitatives sont souvent accessoires en ingénierie ou en toxicologie, mais centrales pour les problématiques en santé mentale ou de transfert intergénérationnel. Les procédures de consensus sont aussi au cœur de nombreuses démarches, mais elles varient beaucoup selon les projets de développement concernés. Les règles peuvent varier aussi selon le domaine considéré. Par exemple, en design, on reconnaît que regrouper quatre ou cinq personnes suffit pour identifier 80 % des problèmes potentiels dans des tests de suivis (Wilson et Corlett 2005).

2 LES GUIDES SUR LES GUIDES, LEUR UTILISATION ET LES INSTRUMENTS D'ÉVALUATION

Les guides, à l'exception de ceux États-Unis, sont développés essentiellement par des organisations publiques (professionnelles, gouvernementales/paragouvernementales) qui ont mis en place des structures à cet effet et en ont même professionnalisé la production. Par exemple, dans l'étude Kryworuchko *et al.* (2009) portant sur 1 360 guides, 80 % avaient été développés par une organisation. Ils avaient été produits par 96 développeurs dont 41 étaient responsables de 94 % de la production, à raison de plus de 30 guides chacun.

À la suite du travail effectué par l'IOM pour élaborer un premier GsG et définir les attributs et les critères que les guides de pratique devraient posséder, plusieurs organisations nationales ont aussi produit leur GsG (cf. annexe 1.1). Les attributs spécifient les qualités que devrait avoir un guide (ex. clarté, validité) ainsi que son contenu en matière de rubriques (ex. avoir un plan de révision, d'implantation). Les huit premiers attributs au tableau 2.1 proviennent de l'IOM.

Tableau 2.1 - Les attributs et critères que les guides de pratique devraient posséder

Validité	Si mène aux résultats anticipés en matière de santé et de coûts.
Fiabilité/ reproductibilité	Si d'autres groupes d'experts arrivent aux mêmes conclusions en utilisant les mêmes preuves et méthodes de développement du guide. Si, sous les mêmes conditions cliniques, le guide est interprété et appliqué de façon consistante par les praticiens ou les autres parties.
Applicabilité clinique	A-t-on défini de façon appropriée la population des patients? A-t-on précisé la population <i>cible</i> ((ex. groupes ethniques, économiques)? A-t-on discuté des problèmes cliniques complexes (ex. comorbidité)? A-t-on précisé la logique (<i>rationale</i>) pour exclure des patients?
Flexibilité clinique	A-t-on déterminé les exceptions cliniques (connues ou attendues) aux recommandations? A-t-on répertorié les exceptions non cliniques? A-t-on discuté des préférences des patients?
Clarté	Langage non ambigu ; éviter les termes directifs (<i>aspirational language</i>) tel <i>should, must</i> . Définition précise des termes. Présentation logique, facile à suivre. En-têtes, index clairs, résumés précis. Recommandations consistantes-cohérentes ; recommandations majeures distinguées des autres?
Processus multi- disciplinaire	S'est-on assuré de la participation des représentants des groupes affectés (ex. table ronde : fournissent des données probantes ou des points de vue, renvoient les recommandations)? A-t-on décrit les parties concernées, leurs <i>credentials</i> et intérêts? A-t-on tenu compte des biais et des conflits d'intérêts potentiels? A-t-on élaboré des méthodes pour tenir compte du point de vue de ceux qui ne siègent pas sur la table ronde? A-t-on questionné les sources de financement et les biais possibles?
Plan de révision	Quand sera-t-il révisé? Sera-t-il révisé advenant de nouvelles preuves cliniques ou changement de consensus?
Documentation	Description méticuleuse de la procédure utilisée, des participants impliqués, des preuves utilisées, des hypothèses et de l'argumentaire, des méthodes analytiques employées.
Diffusion	A-t-on prévu comment les guides seront distribués aux utilisateurs? A-t-on adopté des stratégies pour que patients, médecins et autres collaborent à sa diffusion et adoption?
Implantation	Stratégies pour planter les guides. Faisabilité (politiques et implications administratives de l'utilisation).
Évaluation	Comment le GP sera-t-il évalué une fois implanté (ex. santé économique, satisfaction des patients et des <i>providers</i> , résultats [<i>outcomes</i>])? Adhérence au GP par les utilisateurs attendus.

Pour rencontrer ces attributs, les GsG ont défini des éléments à vérifier (*key items*) et des procédures à respecter dont nous résumons les grandes lignes ci-dessous (les variantes et ajouts introduits par les groupes sont compilés à l'annexe 2.1). Nous analyserons par la suite l'utilisation du processus proposé et les principaux outils d'évaluation des guides.

2.1 Les étapes d'élaboration des guides

Globalement, tous les GsG partent des mêmes attributs et proposent sensiblement les mêmes grandes étapes d'élaboration. Ce qui les distingue les uns des autres est l'importance ou l'attention accordée à certains attributs et à la façon d'y arriver. Les derniers GsG (ou versions) produits intègrent les critiques plus récentes. D'après l'analyse comparative de Turner *et al.* (2008), les GsG les plus complets et les plus clairs sont ceux développés par le National Institute for Health and Clinical Excellence (NICE, britannique) et le New Zealand Guidelines Group (NZGG). Mais selon Shünemann *et al.* (2006), les meilleurs GsG seraient ceux développés par NICE et le Scottish Intercollegiate Guidelines Network (SIGN).

Le but des sections suivantes n'est pas de comparer les GsG, mais de situer leurs différences et leurs complémentarités. Comme certains attributs, critères et items définis font l'objet de nombreux écrits, nous ne citerons les études que lorsqu'elles éclairent sur un point qui n'est traité que dans ce chapitre. D'ordre général, les GsG suggèrent les 14 étapes suivantes pour l'élaboration des guides :

1. Sélection du sujet.
2. Délimitation du but et du public cible anticipé (*scope*).
3. Adaptation du guide.
4. Formation du groupe de développement.
5. Participation des patients.
6. Élaboration des questions.
7. Revue systématique.
8. Critères d'inclusion et d'exclusion des études et des données.
9. Évaluation des données probantes ou des preuves.
10. Développement des recommandations.
11. Élaboration d'une stratégie d'implantation.
12. Consultation pour évaluer la version préliminaire du guide de pratique.
13. Rédaction de la version abrégée.
14. Planification de l'évaluation, révision et mise à jour.

2.1.1 La sélection du sujet

D'abord, les GsG déterminent les critères pour choisir un sujet. Un premier groupe est lié directement aux demandes des mandataires qui appuient l'organisation et qui en définissent les grandes orientations stratégiques. Nous le laisserons donc de côté. Le second groupe définit de réels critères de choix. Les trois éléments clés qui reviennent comme critères sont l'importance du problème (gravité, nombre de gens affectés), la variabilité des pratiques et le potentiel d'amélioration ou d'impact (ex. réduire la mortalité). Le GsG néo-zélandais (NZGG) résume ces critères ainsi : là où un guide peut faire la différence.

Ces critères de sélection du sujet se rapprochent des exercices de priorisation effectués en SST, sauf en ce qui a trait à la variabilité des pratiques, plutôt considérée - dépendamment des sujets -

comme un mécanisme d'adaptation. Pour l'éco-sais SIGN, la variabilité des pratiques reflète une incertitude clinique ; le Conseil européen la situant comme un symptôme d'une qualité des soins variables. En SST, le choix peut-être relié directement aux contextes organisationnels d'application.

Le GsG le plus exhaustif en la matière est celui développé par le NZGG. Plus que tout autre, il pose le choix du sujet en matière de potentiel et de conditions d'implantation. Selon ce GsG, pour retenir un sujet, il faut que :

- suffisamment de données probantes soient disponibles sur le sujet (Conseil de l'Europe aussi),
- le sujet soit suffisamment complexe pour justifier des débats sur les recommandations,
- les recommandations soient acceptables pour les utilisateurs (inclut la communauté),
- l'implantation soit faisable et que les obstacles ne soient pas trop majeurs,
- il soit clair que l'implantation résoudra le problème (NHMRC),
- l'implantation ne mobilise (ou n'épuise) pas trop les ressources de la communauté.

À ces critères, aussi applicables en SST, les GsG ajoutent : les domaines qui nécessitent une mise à jour ou qui nécessitent des informations de pointe ; les sujets qui impliquent une sous-utilisation des technologies existantes (OMS) ; les interventions répandues, mais dont l'efficacité est remise en question (perspective de coûts, OMS) ; les maladies émergentes et les interventions qui peuvent poser un conflit de choix entre l'individu et la société (préoccupation politique, OMS).

2.1.2 La délimitation du but et du public cible anticipé

Le NZGG, le NICE et le National Health and Medical Research Council d'Australie (NHMRC) sont très explicites quant aux critères de délimitation du but et du public cible anticipé. Ils définissent des conditions beaucoup plus précises qu'il est usuel de le faire en SST : type de soins (primaires, secondaires, tertiaires), groupe de patients et d'utilisateurs, critères d'exclusion ou de non-inclusion, type d'intervention, de traitement ou de test, d'environnement, etc. NICE se distingue en insistant sur la nécessité de délimiter l'ampleur de ce qui sera couvert, qui doit être raisonnable, s'étaler sur une période de temps délimitée et clairement formuler les questions cliniques qui orienteront la revue de littérature.

2.1.3 L'adaptation du guide

Les guides sont développés par trois types d'organisation : gouvernementales, professionnelles et privées (ex. compagnies d'assurances). Comme on le verra au chapitre suivant, l'opinion dominante est que les guides gouvernementaux et nationaux sont de meilleure qualité. Cependant, les guides locaux sont souvent mieux implantés et mieux acceptés. On reconnaît que l'application dépend de nombreux facteurs locaux (système de santé, moyens technologiques, etc. ; Armstrong 2003) et que les contextes d'utilisation jouent un rôle important sur le potentiel d'implantation. Les GsG prévoient donc une étape d'adaptation qui relève des experts locaux et des praticiens (Berg *et al.* 1997). Une seconde raison invoquée est le coût élevé du développement des guides : il importe alors de répertorier d'abord les guides existants puis d'adapter.

Il existe différentes propositions sur la façon d'adapter un guide, mais d'après la revue de la question effectuée par Fervers *et al.* (2006), il n'existe pas de processus d'adaptation validé. Sur les 10 444 publications sur les guides recensées, 18 rapportaient un modèle ou une expérience d'adaptation, neuf discutaient de la question et trois proposaient une approche pour recueillir les données aux fins d'adaptation (*tailoring*) culturelle. En somme, les auteurs préconisent de répertorier d'abord les guides existants, puis de procéder à leur adaptation. Les étapes proposées sont sensiblement les mêmes, bien qu'elles puissent varier en ce qui a trait à leur exécution (recensement des guides existants, évaluation de leur qualité, choix, adaptation au contexte local) et à leur perspective. Par exemple, Graham *et al.* (2003, 2005) insistent sur le fait que l'adaptation locale ne doit jamais modifier les recommandations. Les connaissances locales portent exclusivement sur les conditions d'application. Cela est basé sur la prémisse que les connaissances valides sont celles issues de consensus d'experts internationaux, voire nationaux, qui suivent des règles précises. Ce processus d'adaptation s'inscrit aussi dans la mouvance de la littérature sur la *translation* qui porte sur la nécessaire adaptation des connaissances pour qu'elles deviennent utilisables ou applicables (Faye *et al.* 2007). Fervers *et al.* (2006) ajoutent une phase de révision, comme s'il s'agissait d'un nouveau guide (le modèle proposé se nomme ADAPTE³⁵).

Les procédures proposées incluent une étape d'évaluation (ex. avec un instrument comme AGREE), puis une étape d'adaptation aux conditions locales ou nationales. SIGN propose d'utiliser ADAPTE et il participe lui-même au développement de versions locales dérivées de ses guides nationaux. NHMRC et NICE insistent sur l'importance de la qualité des données probantes (mises à jour, révisions) et sur la nécessité de ne pas affaiblir l'intégrité scientifique du guide source en modifiant les recommandations. Si un groupe local altère les lignes directrices, il doit en informer les développeurs du guide initial (NHMRC). NICE considère cependant normal de reformuler les énoncés ; il spécifie une procédure de mise à jour (mais il ne s'agit pas d'adaptation). Le Conseil de l'Europe (CE) insiste sur la nécessité de revoir et de tester le guide en matière d'applicabilité.

De l'ensemble, c'est le NZGG qui formule la question de la façon la plus élargie et qui intègre le mieux les conclusions des études ci-haut mentionnées, à travers les questionnements suivants: les patients et les utilisateurs néo-zélandais sont-ils comparables? Les recommandations sont-elles « faisables » dans le contexte néo-zélandais? Doivent-elles être réécrites? Qu'est-ce que cela implique en matière de ressources? L'implantation est-elle « faisable »? L'adaptation est vue comme une étape qui permet aux ressources locales de s'approprier le guide, d'y injecter leur point de vue, voire des innovations locales. L'adaptation d'un guide pourrait aller jusqu'à en diminuer la validité scientifique.

En SST, la tradition de développement local ou régional est forte. Une des raisons principales est que les structures de fonctionnement de la SST, de même que les utilisateurs, varient beaucoup d'un pays à l'autre. Les cultures et les formations diffèrent. Par exemple, la France dispose d'un vaste bassin de médecins du travail. Les ergothérapeutes ont une fonction plus importante au Québec. Les approches de terrain étant très développées, les chercheurs en SST ont une grande sensibilité à l'impact des différences : secteur de production, taille de l'entreprise, moyens disponibles, compétences, barrières de langue, alphabétisation, situation économique, etc.

³⁵ C'est aussi l'acronyme utilisé pour nommer le groupe élargi qui a développé le modèle. Il incluait un groupe québécois.

2.1.4 La formation du groupe de développement

Deux des grandes critiques adressées aux stratégies de développement basées sur l'*evidence-based medicine* sont d'avoir un point de vue limité sur les connaissances et de ne pas suffisamment tenir compte des contextes et de la réalité des problèmes. Comme le souligne Dowie (1996), les données probantes et les recommandations peuvent ainsi être valides, mais peu utiles ou inapplicables. Une partie de la réponse à ces critiques a pris la forme de l'identification des sources porteuses de ces connaissances et de l'élargissement des groupes impliqués dans le développement. Par exemple, pour Edwards (2010), toute base de données devrait comporter des connaissances issues de cinq sources : organisation, praticiens, décideurs, chercheurs et usagers. Pour Grimshaw *et al.* (1995), la formation des groupes, soit le processus de composition et la représentation multidisciplinaire, constitue deux des trois clés à la base de la production de guides de qualité (troisième clé : identification et synthèse des données probantes).

Tous les GsG abordent cette question, de façon souvent très détaillée, mais en mettant l'accent sur des aspects différents. Tous insistent sur la nécessité de groupes multidisciplinaires capables de représenter des groupes d'intérêts variés, dont les patients, et sur l'importance de ressources en méthodologie pour l'évaluation des données³⁶. Plusieurs proposent des modes d'organisation ou de participation. Le nombre jugé adéquat est de l'ordre de 15 (SIGN : 15 à 25 ; NICE : 13 à 15). En SST, les groupes qui développent les guides sont généralement plus petits, mais les processus de consultation et d'allers-retours sont plus élaborés et plus étalés dans le temps. Mais ils peuvent être aussi élargis. Par exemple, le comité « convoyeur » chargé de rédiger le guide cité précédemment était composé de 15 personnes, dont trois issues d'un organisme de recherche (IRSST) et 12 de l'organisme chargé de gérer la SST au Québec (CSST). Les personnes issues de la CSST avaient des profils variés : inspecteur, chef d'équipe ou spécialiste provincial. Un sous-comité de recherche de cinq personnes était chargé de trouver des solutions scientifiques ou fiables (issues de normes ou autres documents nationaux) afin de les présenter à l'ensemble du comité pour discussion et adoption.

SIGN insiste sur le problème des rapports hiérarchiques liés aux différences de statut professionnel, problématique moins présente en SST. NHMRC et NZGG insistent sur la question des conflits d'intérêts (question moins présente en SST³⁷). NZGG propose toute une procédure de discussion – négociation à mener au sein du groupe de développement avant d'amorcer les travaux³⁸. NICE insiste sur l'équilibre des représentations ethniques ou régionales (Angleterre vs Écosse) et la nécessaire reconnaissance des variantes régionales. Les recommandations sommaires du CE portent plutôt sur l'importance d'essais auprès des utilisateurs. L'OMS se démarque par l'accent mis sur la structure, les définitions des fonctions et la composition de chaque comité : comité de pilotage (*steering group*), groupe de développement technique, groupes de travail complémentaire, secrétariat.

³⁶ Résultats de l'enquête menée par Burgers *et al.* (2003) auprès de 18 organisations sur les groupes : taille du groupe : 10 à 20; 3 à 5 disciplines; 15 avaient invité un épidémiologiste, 12 un spécialiste en bibliothéconomie, 4 un statisticien.

³⁷ La SST n'est certainement pas à l'abri de conflits d'intérêts potentiels (ex . lien d'emploi médecin du travail/entreprise), mais cela n'est présentement pas une problématique importante et ne risque pas d'avoir beaucoup d'impact sur la conception des guides.

³⁸ Ex. de points couverts : acceptation et tolérance aux points de vue divers, accord sur la préséance des faits sur les opinions, entente quant à la répartition de la charge de travail, les aires de confidentialité, la contribution, l'assistance aux réunions.

2.1.5 La participation des patients

Dans tous les GsG, la place accordée aux patients ou aux consommateurs est discutée. Comme les patients sont souvent organisés en regroupements, c'est généralement à travers ces structures et par le biais de leurs représentants que les processus de consultation sont mis en place. Les propositions des GsG varient surtout quant aux détails et aux précisions fournis sur le statut ou le rôle des patients et la façon de les intégrer³⁹. Dans l'ensemble, il s'agit d'intégrer leur point de vue pour s'assurer que le guide répond à leurs besoins, à leurs préoccupations, et intègre leurs perspectives. Deux formules sont privilégiées : des consultations organisées par voie épistolaire et la tenue de groupes témoins.

En SST, l'analogie la plus proche est celle du travailleur et le type de représentation, le syndicat. Cependant, la diversité des acteurs potentiellement impliqués et des modes de représentation y est notablement plus importante : gestionnaires, inspecteurs, hygiénistes, préventeurs, etc. ; associations paritaires, équipes de santé au travail, etc. L'approche paritaire mise de l'avant dans les politiques de SST fait qu'il existe déjà une forte tradition d'inclusion. En matière d'interlocuteurs, le travailleur, peut-être aussi le gestionnaire, le préventeur, l'inspecteur. Il faudrait parler ici de participation des acteurs.

Suite à cette étape, les GsG abordent le travail de synthèse qui est à faire pour l'élaboration des guides. Ce travail de synthèse est découpé en quatre thèmes : l'élaboration des questions, qui orientera la recherche, les procédures de recherche des données, la définition des critères d'inclusion ou d'exclusion des documents, l'évaluation des données.

2.1.6 L'élaboration des questions

Dans la mesure où les guides de pratique sont définis comme des suites d'énoncés, la formulation des questions constitue une étape naturelle. Leur rôle est de mieux guider la recherche de la littérature et des données probantes. Ce sujet est abordé de façon relativement détaillée par NICE, NZGG et SIGN.

NICE et NZGG proposent d'utiliser le format du guide PICO (*Patients, Intervention, Comparison, Outcomes*) : 1) population de patients à cibler (ex. âge) lors de la recherche des données probantes sur les interventions ; 2) intervention : ce qui doit être fait (ex. traitement, maladie, facteur de risque) ; 3) comparaisons à établir (ex. vs placebo, pas de traitement, comparaison de thérapies) ; 4) effets attendus. Le nombre de questions varie selon les guides, mais NICE propose le repère suivant : pour un guide qui exige 18 mois de travail, 30 questions.

2.1.7 La revue systématique

Comme la revue systématique relève généralement du mandat du groupe de développement, il est assumé qu'elle ne pourra généralement pas avoir l'ampleur prescrite.

NHMRC insiste sur le caractère formel de la revue, où l'étalon-or est la méta-analyse combinant des essais contrôlés randomisés : « *The more formal the review, the more credible and valid the*

³⁹ Par exemple, SIGN précise l'agenda en matière d'échéancier; NZGG énumère une longue liste de contributions potentielles et insiste sur leur rôle dans l'implantation et la diffusion. SIGN signale leur apport en tant que complément potentiel aux données probantes. Intégration : ex. SIGN précise le type d'aide à prévoir pour les représentants des patients.

resulting guidelines ». NICE, dont l'approche est similaire, insiste sur l'importance de choisir les meilleures données ; la recherche, centrée sur les bases de données électroniques⁴⁰, est menée par du personnel spécialisé. Les étapes de la recherche systématique sont décrites précisément et sont hiérarchisées, y compris quant aux bases de données à utiliser et à quel moment. Il fait cependant une plus grande place aux partenaires qui peuvent être invités à soumettre leurs données. Ce qui peut être soumis est précisé.

SIGN signale l'importance d'intégrer les perspectives des patients dès le départ et suggère d'inclure la recherche d'écrits généralement non couverts par les revues systématiques (par exemple les écrits qui relèvent du nursing et de la psychologie) et d'inclure celle des guides existants et leur évaluation. NZGG insiste aussi sur la nécessité d'élargir les champs d'investigation vu l'étendue de la littérature quant à ses formats de diffusion : journaux, publications électroniques, présentations affichées, livres, etc. Il propose d'effectuer cette recherche avec des chercheurs et des praticiens proches du domaine. Il se distingue, avec NICE, par la place qu'il accorde aux données qualitatives auxquelles il consacre un chapitre. On y décrit les grandes méthodes qualitatives utilisées et on situe leur rôle en tant que méthodes qui permettent d'éclairer la signification des événements : « *The focus on quantitative research is the frequency or the occurrence of events whereas qualitative research aims to provide insights into the meaning of the events* ». Elles sont vues comme nécessaires pour préciser les résultats attendus, comprendre les comportements dans un contexte quotidien réel, mieux prendre en compte la réalité, planifier les stratégies d'implantation et mieux déceler les barrières potentielles. Les deux organisations internationales que sont l'OMS⁴¹ et le CE abordent plus sommairement cette question par le biais de références à consulter.

2.1.8 Les critères d'inclusion et d'exclusion des études et des données

La question des critères d'inclusion et d'exclusion des études et des données est traitée essentiellement par NICE et SIGN. La première étape vise à exclure ce qui n'est pas approprié pour le sujet exploré ou les doublons, la seconde, à exclure les documents dont les résumés ne répondent pas aux standards et la troisième, à retirer les articles qui ne répondent pas aux standards méthodologiques établis.

2.1.9 L'évaluation des données probantes ou des preuves

Le sujet de l'évaluation des données probantes ou des preuves est couvert de façon précise par SIGN, NICE et NZGG. SIGN (NICE s'y réfère) utilise essentiellement la liste de contrôle mise au point par un groupe australien, MERGE (*Method for Evaluating Research and Guideline Evidence*)⁴², qu'il a adaptée pour mieux équilibrer les exigences de rigueur méthodologique en regard des besoins pratiques. Les principaux éléments couverts sont : 1) la revue systématique et la méta-analyse ; 2) les essais contrôlés aléatoires ; 3) les études de cohortes ; 4) les études de

⁴⁰ Le lecteur intéressé pourra consulter sur le site du RRSSTQ (<http://www.rrsstq.qc.ca/>) en allant sur le portail en transfert de connaissances, section publications, onglet outil, une compilation des portails d'accès et sites de référencement des bases de données dans le domaine de la santé.

⁴¹ L'OMS a commandé depuis une série de revues sur la question du développement des guides qui implique des ajustements qui ne sont pas pris en compte ici (Oxman *et al.* 2006).

⁴² Liddle *et al.* (1996). Groupe du New South Wales Department of Health. Cf. l'annexe C du GsG de SIGN.

cas-témoins ; 5) les études de diagnostic ; 6) les évaluations économiques. Nous résumons ci-dessous les trois étapes principales du processus :

- Étape 1. Identification et sélection du matériel : 1) Question clairement formulée? 2) Méthodologie de recherche des documents décrite? Systématique? etc. 3) Recherche assez rigoureuse pour répertorier les études pertinentes (ex. : base de données utilisée ; ajout d'une recherche « manuelle » dans des revues clés)? 4) Évaluation de la qualité des études (aux fins d'inclusion, d'exclusion)? 5) Vérification de la similarité des études avant de les combiner?
- Étape 2. Évaluation du matériel : 1) Attribution d'une cote sur la qualité méthodologique des études? 2) Si la recension des écrits est pauvre ou limitée, quel en est l'impact?
- Étape 3. Organisation du matériel : 1) Types d'études couvertes par la revue et résumé des conclusions avec les forces et faiblesses de l'étude. 2) Permettent-elles de répondre aux questions?

NHMRC et WHO se réfèrent essentiellement aux directives du Cochrane *textbook*⁴³. Les preuves sont évaluées à travers quatre paramètres : 1) niveau : le design de l'étude permet-il de minimiser les biais, l'étude aléatoire étant l'étalon-or⁴⁴ ? 2) Qualité des méthodes utilisées pour minimiser les biais : on considère que les critères pour évaluer la qualité des études randomisées sont bien établis, et que ce n'est pas le cas pour les autres types de design. On reconnaît qu'il y a des zones de recouvrements et qu'une étude d'observation de haute qualité vaut éventuellement plus qu'une étude aléatoire de faible qualité. 3) Pertinence : il peut s'agir de la proximité des données probantes par rapport à la question formulée, de la pertinence par rapport aux effets à évaluer, de l'applicabilité à d'autres patients ou à d'autres contextes. 4) Force de la preuve : grandeur, précision et reproductibilité des résultats obtenus. NZGG combine plusieurs approches⁴⁵, mais le point de vue est similaire ; il ne réfère plus aux études qualitatives.

Les GsG adoptent tous sensiblement la même approche pour évaluer la qualité des données, soit la méthode utilisée. Le système de cotation des preuves⁴⁶ est le suivant :

- Une revue systématique de toutes les études d'essais aléatoires pertinentes.
- Au moins une étude randomisée de qualité ; 1) les études sont pseudo-aléatoires, mais elles sont de bonne qualité ; 2) les études sont comparatives avec groupe contrôle ; 3) les études sont comparatives avec groupe de contrôle historique.
- Études de cas, pré-tests, post-tests.

⁴³ Pour une version plus récente, voir : Cochrane Handbook of Systematic Reviews dans sa dernière version accessible à www.cochrane-handbook.org. Le lecteur intéressé à un versant plus social pourra consulter le site Campbell <http://www.campbellcollaboration.org/library.php>

⁴⁴ En SST, cela peut introduire un biais majeur dans la mesure où cela amène à sélectionner des contextes qui se prêtent à ce type de design. SIGN considère par ailleurs la validité externe et si le lien est direct avec l'application.

⁴⁵ Harbour and Miller (2001). A new system for grading recommendations in evidence-based guidelines (Harbour est le directeur du SIGN Grading Review Group); Greer *et al.* (2000). A practical approach to evidence grading. Harris *et al.* (2001). Current methods of the US Preventive Services Task Force. Shekelle *et al.* (1999)

⁴⁶ Source citée par les auteurs : US Preventive Task Force (1989) Guide to Clinical Preventive Services : an assessment of the effectiveness of 169 interventions. Il existe des variantes. Ex. : Marshall (2000) introduit le facteur hétérogénéité et la valeur de l'intervalle de confiance : 1) randomisée homogène et intervalle de confiance sous le seuil; 2) randomisée mais hétérogène; 3) observationnelle (ex. un groupe delphi). Voir aussi : Voir aussi le travail historique de la Canadian Task Force on the Periodic Health Examination (1979) ainsi que Guyatt GH *et al.* (1995) *Users' guide to the medical literature IX: a method for grading health care recommendations*.

En SST, la recherche basée sur des devis expérimentaux est peu présente. Est-ce en lien avec les paradigmes de la recherche, des raisons éthiques ou pragmatiques de résolutions de problèmes? Quoi qu'il en soit, en SST il n'y pas de tradition de la recherche centrée sur ces approches expérimentales avec essais randomisés. Comme on l'a vu précédemment, la production des données procède d'approches et de points de vue disciplinaires multiples. Cela peut expliquer l'absence de références communes. Néanmoins, les critères de rigueur scientifique des diverses disciplines contributives comme la sociologie, la toxicologie, la psychologie, l'ergonomie ou les relations industrielles sont autant de références pour construire adéquatement une méthode systémique d'élaboration des guides en SST. Un travail consensuel est à faire.

2.1.10 Le développement des recommandations

Pour effectuer le développement des recommandations, la première étape proposée est de juger et évaluer les preuves. Les concepts utilisés sont les mêmes que ci-dessus, mais le jugement à porter concerne aussi l'applicabilité et l'impact potentiel. Ce dernier est fondé sur le jugement des membres du groupe de travail. Le statut des recommandations ne s'établit donc pas uniquement en fonction de la force des preuves, mais aussi de leur mise en pratique. Des efforts sont consacrés au développement d'outils pour graduer aussi cette dimension⁴⁷.

Pour SIGN, la proposition intègre les études qualitatives essentiellement à cette étape. Le commentaire sur les études qualitatives en tant que sources de données probantes est fort intéressant : « *Qualitative methods are increasingly being used to inform practice in some aspects of medical care. At present, there is no mechanism for incorporating such studies in the evidence base. Some progress has been made on methods of identifying qualitative studies, and in evaluating their methodological quality. The use of qualitative evidence to identify issues of concern to patients, and to help identify key questions to be addressed in the guideline is becoming an established part of SIGN methodology. A pilot exercise looking at the formal inclusion of qualitative evidence in developing a SIGN guideline has been carried out and will form the basis of future developments in this area.* »

NHMRC insiste de son côté sur les résultats (*outcomes*) attendus, le compromis bénéfice-risque, les avantages comparatifs (ex. autres traitements), les coûts. Il s'agit d'éléments dont l'absence ou l'insuffisance étaient fortement critiquées jusqu'alors. L'OMS insiste sur le compromis à établir entre coûts et bénéfices attendus. Pour sa part, NZGG résume ainsi la question : « *It is surprisingly rare for the evidence to show clearly and unambiguously what course of action should be recommended for any given question. The quality of all the evidence needs to be linked with other aspects that require "considered judgment"* » Le groupe insiste sur l'importance d'analyser l'applicabilité et d'être capable de prendre en compte les différences culturelles et organisationnelles propres à la Nouvelle-Zélande lorsqu'il s'agit d'extrapoler les résultats issus d'études étrangères. L'évaluation de l'impact clinique⁴⁸ est considérée comme un aspect central du processus d'élaboration des recommandations. Il souligne le risque de réintroduire des biais, d'où l'importance accordée à la méthodologie de décision par consensus.

NICE aborde quatre aspects : 1) la traduction des preuves en recommandations (idem aux autres GsG) ; 2) la formulation des recommandations ; 3) les approches possibles pour établir les

⁴⁷ Ex. The GRADE Working Group (2004). *Grading quality of evidence and strenght of recommendations* (cité par SIGN).

⁴⁸ L'outil utilisé « *Considered Judgment Form* » est une version adaptée du formulaire développé par SIGN. Les éléments considérés sont : 1. quantité de preuves; 2. applicabilité et généraibilité; 3. consistance; 4. impact clinique; 5. autres facteurs; 6. preuves en matière d'énoncés (*evidence statement*); 7. recommandations.

priorités en matière d'implantation ; 4) la formulation de recommandations de recherche (pour préciser les questions pour lesquelles les connaissances sont insuffisantes⁴⁹). Partant du principe que les utilisateurs n'auront pas le temps de lire l'ensemble des éléments à la base des recommandations et que ces dernières doivent donc être rédigées soigneusement, il insiste sur la *translation*⁵⁰ et la mise en mots des recommandations (cf. encadré 2.1). Pour NICE, l'établissement des priorités est vu comme étroitement lié au potentiel d'implantation. Il revient au groupe de développement de statuer sur les priorités (entre cinq et dix recommandations normalement). Plusieurs critères peuvent être utilisés (ex. : impact sur la morbidité, réduction de la variabilité des pratiques), mais ils doivent être clairement indiqués.

Encadré 2.1 - La formulation des recommandations par NICE

Les recommandations doivent

- être complètes en elles-mêmes et ne pas nécessiter de recourir à d'autres matériels ;
- être claires et concises ;
- être orientées vers l'action ;
- (devraient) prendre la forme du *should* et non du *must* ;
- (devraient) être formulées pour souligner le droit du patient à participer aux discussions (ex. les traitements offerts plutôt que prescrits ou donnés) ;
- (peuvent) indiquer si d'autres traitements ou activités peuvent être altérés, voire arrêtés.

Les données de directives d'ordre temporel (temps d'attente) doivent être évitées.

Les tableaux ne doivent être utilisés que s'ils en améliorent substantiellement la clarté.

Les termes *people* ou *patients* devraient être préférés à ceux de *subject* ou *individual*.

En SST, les recommandations doivent souvent être situées dans un contexte systémique. La recherche de compromis est usuelle. Mais, surtout, il existe une étape supplémentaire qui lui est propre, soit la traduction de la recommandation en solution. Plusieurs solutions peuvent être envisagées pour une même recommandation.

2.1.11 L'élaboration d'une stratégie d'implantation

Tous les GsG consultés soulignent l'importance de traiter de l'implantation. Essentiellement, ils proposent une synthèse de la littérature à partir de laquelle ils font le point sur l'efficacité des différentes stratégies. Ils insistent sur l'importance de déceler les obstacles potentiels et d'élaborer les stratégies d'implantation en fonction du contexte. NICE propose des outils sur son site web pour aider à l'implantation : par exemple, des outils pour estimer les coûts d'implantation/économies, des audits pour contrôler l'implantation locale.

2.1.12 Le processus de consultation pour évaluer la version préliminaire du guide de pratique

Pour évaluer la version préliminaire du guide, le processus comporte des propositions de consultations auprès des experts du domaine, des professionnels et des groupes de représentants concernés. Le processus comporte la mise en place de larges colloques ou d'ateliers de travail plus restreints, d'audits, comme des consultations individuelles et des mises à l'essai.

⁴⁹ Utilise aussi la liste de contrôle basée sur le modèle PICO (Patient/population ou problème, Intervention/facteurs de risque considérés, à qui cela sera comparé [ou ce qui sera contrôlé], résultats attendus.)

⁵⁰ Nous conservons le terme de *translation*, utilisé dans la littérature en transfert de connaissances dans un sens d'adaptation.

SIGN insiste sur l'organisation de rencontres à un niveau national⁵¹ pour recueillir les commentaires et les suggestions (y compris sur les données additionnelles à prendre en compte), mais aussi susciter une adhésion en vue de l'implantation. Le guide est revu par des évaluateurs externes - invités en tant qu'individus et non en tant que représentants⁵² - dont le rôle est de se pencher plus particulièrement sur la clarté, la précision des recommandations et l'utilité pratique du guide. L'intégration des commentaires est d'abord réalisée par le comité de développement, puis vérifiée par le comité éditorial. Le comité éditorial révisé tous les commentaires et leur donne suite. Pour les essais pilotes, SIGN privilégie les essais locaux en partant du point de vue que ce qui fonctionne dans un contexte donné peut échouer dans un autre. Les propositions de NICE sont relativement semblables. Le but premier des rencontres est de vérifier la pertinence, la couverture globale et le réalisme des attentes. NICE aborde plus longuement la question de l'intégration des commentaires. Ses propositions se rapprochent d'un protocole avec échancier intégré. Il distingue les composantes à évaluer : les recommandations, l'ensemble du guide, le manuel d'accompagnement, le guide abrégé. Le NHMRC insiste sur le fait d'inclure les groupes d'intérêts qui n'auraient pas été inclus au cours du processus de développement du guide (ex. groupes industriels, autorités locales). Le NZGG centre le processus sur les partenaires et la nécessité de réaliser des pré-tests dans des milieux variés.

2.1.13 La rédaction de la version abrégée

Les abrégés-résumés visent à faciliter l'utilisation et la diffusion du guide. NZGG centre ses recommandations sur l'abrégé destiné à l'utilisateur, réputé ne pas avoir le temps de lire la version complète du guide et/ou de la conserver avec lui. Il propose un format de présentation qui inclut l'introduction d'un algorithme⁵³. L'abrégé devrait tenir sur le recto-verso d'un feuillet standard. HMRC insiste surtout sur la diffusion, plus particulièrement les voies de diffusion du guide. Le résumé – en fait les résumés – vise à atteindre un public plus vaste. Les formats suggérés sont multiples : brochure, vidéo, encart, article... SIGN insiste sur l'importance de mettre en évidence les messages clés et de prévoir une version pour les patients. Les GsG suggèrent généralement la mise en place d'un format électronique accessible sur site web. Comme nous le verrons au chapitre 5, les diffusions larges sont considérées peu efficaces.

2.1.14 La planification de l'évaluation, révision et mise à jour

« Evaluation of CPGs is essential. How will we know if guidelines are worth developing, disseminating and implementing if we do not know whether they make a difference to clinical practice and health outcomes? The amount of practical detail available to enable guideline developers to carry out these processes varies between handbooks, and often discussion is quite limited. » (Turner et al. 2008)

En fait, les GsG consultés abordent surtout la question de la mise à jour des guides. Les références au processus d'évaluation des impacts sont généralement assez sommaires. NICE propose, pour l'évaluation, un intervalle de deux ans⁵⁴ et NHMRC, de trois à cinq ans. SIGN s'engage à réviser ses guides tous les trois ans et envisage qu'il soit possible d'instaurer un système en continu. Les six composantes de l'évaluation résumées par NHMRC sont :

⁵¹ Suggère un auditoire de 150 à 300 professionnels de la santé, d'accréditer ces rencontres comme activités de développement professionnel et de mettre le GdP en disponibilité sur le site web un mois à l'avance.

⁵² On assume que les organisations seront représentées plutôt au colloque national.

⁵³ Plusieurs auteurs considèrent les algorithmes indispensables à la pratique (cf. chapitre 3). La proposition vise à y pallier.

⁵⁴ Voir la section 3.2.1 sur la durée de vie des guides. NICE est celui qui s'étend le plus sur la question (chapitre 15, p. 86-89).

- la dissémination ;
- l'évolution de la pratique. Va-t-elle dans le sens des recommandations? Le NZGG formule la question en matière de réduction du fossé pratique/recommandations ;
- les effets sur la santé ;
- les changements de pratique ;
- l'impact sur les connaissances des patients et leur compréhension (indicateurs - clarté, accessibilité, acceptabilité, satisfaction, pertinence) ;
- le coût (ex. pré vs post). Le NZGG pose la question de l'utilisation des ressources.

Maintenant que nous avons vu les 14 étapes d'élaboration des guides suggérées, la prochaine section sera consacrée à l'étude de l'application des recommandations des GsG.

2.2 L'application des recommandations des GsG

Audet *et al.* (1990) écrivaient il y a 20 ans qu'il n'y avait pas de données probantes claires sur la meilleure méthode de développement des guides et, 15 ans plus tard, Burgers *et al.* (2003) ajoutaient qu'il y avait peu d'études portant sur ce qui rend un guide efficace. Globalement, les études menées auprès des développeurs montrent que les recommandations ne sont pas très suivies (Schünemann *et al.* 2006), entre autres à cause du manque de ressources.

2.2.1 Les données colligées

Les développeurs utilisent, voire privilégient encore, des approches non systématiques et s'appuient sur le consensus d'experts (Calder *et al.* 1997⁵⁵ ; Harpole *et al.* 2003). Par exemple, en Finlande, sur les 719 guides analysés, 26 % avaient procédé avec une analyse systématique (Grol *et al.* 1998) ; deux incluaient une méta-analyse. Dans celle de Pencharz *et al.* (2002) sur 18 guides portant sur l'ostéoarthrite des membres inférieurs, environ la moitié avait coté les données probantes et 2 avaient effectué une méta-analyse.

L'enquête de Burgers *et al.* (2003) menée auprès de 18 organisations⁵⁶ montre qu'il y a eu une certaine évolution et que les revues systématiques et l'application des méthodes de l'*EBM* sont plus utilisées que les stratégies de consensus. Cependant, les recherches électroniques continuent d'être complétées par des recherches manuelles. Une étude plus récente portant sur 1 664 guides canadiens développés au cours de deux périodes (1994-99 vs 2000-05) montre au contraire que la situation a peu évolué⁵⁷ (Kryworuchko *et al.* 2009).

Certains chercheurs se sont penchés sur l'origine des informations ou des sources de références utilisées. Deux constats s'en dégagent. Le premier est que les auteurs des guides privilégient deux sources : les références nationales et les références issues de publications américaines (Burgers *et al.* 2002). Le second constat est l'étalement des références citées, la majorité n'étant citée qu'une ou deux fois⁵⁸. Dans l'étude de Christiaens *et al.* (2004), sur les 205 références colligées dans 4 guides portant sur un même sujet spécifique, 7 étaient partagées par 3 des 4

⁵⁵ Revue de 59 références concernant des guides portant sur la transfusion sanguine : 17 études cliniques, 42 revues.

⁵⁶ Neuf sociétés professionnelles, trois agences gouvernementales, deux nationales et quatre académiques.

⁵⁷ La majorité utilisent une stratégie informatique pour répertorier les données (94 %), mais la moitié (52 %) seulement utilisent une procédure pour graduer la qualité des données et moins de la moitié (42 %) publient leur stratégie de recherche.

⁵⁸ Les auteurs ont analysé 15 guides de 13 pays. Sur les 1033 citations, 18 étaient partagées par au moins un autre guide. Seulement 1 % était mentionné dans 6 guides et plus.

guides et une seule, par les 4 guides. L'analyse des références utilisées par les intervenants à la Commission des lésions professionnelles au Québec sur le syndrome du canal carpien a montré la même tendance (Tcaciuc *et al.* 2011) : les références locales étaient privilégiées et partagées alors que les autres références étaient peu partagées. Une hypothèse avancée est que les développeurs accordent plus d'importance et de crédibilité aux sources « nationales » car leurs chercheurs prendraient mieux en compte les éléments contextuels ou les pratiques développées. Les résultats seraient en quelque sorte plus proches de la réalité.

2.2.2 La mise en place des étapes recommandées

Audet *et al.* (1990) ont mené leur enquête auprès de huit organisations américaines sur quatre aspects : objectifs, méthode utilisée, approche pour implanter et méthode d'évaluation. Leurs principales conclusions sont que : 1) les objectifs sont le plus souvent définis de façon générale pour éviter la controverse, ce qui nuit à la planification des évaluations ; 2) peu de ressources sont consacrées à l'implantation par rapport au développement ; 3) l'évaluation des impacts demeure rudimentaire à cause de ses coûts et de l'impact négatif sur la réception des guides.

Dans l'enquête de Burgers *et al.* (2003), la majorité des 18 organisations questionnées avait mis en place une unité composée de personnel permanent dédié au développement de guides. Les groupes de développement, composés de 10 à 20 membres, s'adjoignent assez systématiquement des experts en méthodologie et les membres reçoivent en général une formation sur le développement de guides. La plupart utilisent des mécanismes de révision externe (ex. par l'organisation professionnelle ciblée), la majorité produit des outils d'application (ex. graphique de cheminement, algorithme) et des versions résumées. Onze avaient produit des versions pour les patients. Ce qui distingue le plus les organisations est en fait l'importance accordée à la dissémination et à l'implantation, les stratégies utilisées étant très variées. Les étapes d'évaluation sont aussi plus présentes : la moitié effectuent un suivi ou évaluent au moins certains aspects des guides. Les organisations utilisent des approches par consensus (formelle pour sept) et des mécanismes de vérification externe. Globalement, c'est le passage des données probantes aux recommandations qui est vu comme étant une des étapes les plus délicates et sujettes à discussion. Une des difficultés serait d'arriver à expliciter les valeurs sous-jacentes à la formulation des recommandations⁵⁹ (Cook et Giacomini, 1999).

Dans l'étude précitée de Kryworuchko *et al.* (2009) menée sur les guides eux-mêmes, 42 % avaient publié leur stratégie de recherche, 12 % évaluaient l'efficacité de la stratégie de dissémination et 24 %, les résultats (*outcomes*) pour la santé.

2.2.3 Le respect des critères proposés par les GsG⁶⁰

La première vaste enquête a été menée par Shaneyfelt *et al.* (1999) sur 279 guides. Sur les 25 critères considérés, le nombre moyen d'items jugés satisfaisants a été de 12 par guides, soit 43 % (cf. encadré 2.2, conformité pour 10 critères). Aucun n'a rencontré les 25 critères. Ne pas décrire comment on combinait les données avec l'opinion des experts a été l'un des critères le moins souvent satisfaits (8 %).

⁵⁹ Les GsG et les organisations en parlent peu, mais la question est régulièrement soulevée par les utilisateurs (cf. chapitre 3).

⁶⁰ La majorité a utilisé comme instrument d'évaluation AGREE, que nous verrons à la section 2.3.

Encadré 2.2 - Pourcentage des 279 GdP conformes aux standards recommandés (Shaneyfelt *et al.* 1999)

But	75 %
Raisons et importance du GdP	88 %
Participants et leur champ d'expertise	26 %
Problèmes de santé ciblés ou les technologies	61 %
Population de patients ciblée	46 %
Utilisateurs ou public ciblé	51 %
Options disponibles spécifiées aux cliniciens et patients	82 %
Résultats attendus	40 %
Méthode utilisée pour revoir le GdP (external review)	32 %
Date où le GdP doit être revu	11 %
Adhésion moyenne (écart-type)	51 % (25 %)

Grilli *et al.* (2000) a pour sa part vérifié l'adéquation de 431 directives pour 3 grands critères. L'étude a montré que seulement 5 % répondaient aux trois critères : 1) informations sur les partenaires et types de professionnels impliqués : absent de 67 % des guides ; 2) stratégie pour répertorier les données probantes : absent à 88 % ; 3) règles de graduation des recommandations : absent de 82 % des guides. Dans l'étude de Burgers *et al.* (2004), les items cotés en commun les plus bas ont été la participation des patients, la description des critères de sélection, la description d'essais pilotes, les conflits d'intérêts et l'indépendance éditoriale.

Deux autres études ont été conduites sur des guides spécifiques. Celle de Harpole *et al.* (2003) a porté sur 51 guides sur le cancer du poumon comportant un total de 880 recommandations (8 à 118 par sujet) : environ la moitié était basée sur des approches par consensus (*vs* 53 % *EBM*). Les résultats, résumés à l'encadré 2.3, montrent que l'applicabilité et l'indépendance éditoriale obtiennent des résultats particulièrement bas, et que la qualité des informations fournies varie considérablement. Cranney *et al.* (2002) ont jugé pour leur part que la qualité moyenne des 21 guides examinés sur l'ostéoporose était faible, avec des cotes particulièrement basses pour l'applicabilité ; 10 ont été jugés inacceptables et 10 acceptables avec modifications.

Encadré 2.3 - Évaluation de 51 guides sur le cancer (Harpole *et al.* 2003)

Items	% (variations) qui répond aux standards d'AGREE
But et objectifs	70 % (29 à 97 %)
Implication des partenaires	35 % (3 à 70 %)
Rigueur de développement	52 % (2 à 95 %)
Clarté	57 % (11 à 90 %)
Applicabilité	20 % (0 à 90 %)
Indépendance éditoriale	24 % (0 à 83 %)

Dans les études de Brouwers et Charette (2001) et de Watine *et al.*⁶¹ (2006), l'applicabilité est aussi faiblement cotée. Ces derniers observent que les guides présentant les meilleures qualités méthodologiques ne conduisent pas nécessairement aux meilleures recommandations et vice-versa⁶². Cette conclusion est partagée par Pencharz *et al.* (2002), qui constatent toutefois que si la qualité des guides varie, ils sont cependant plutôt consensuels quant aux recommandations, les divergences portant sur des questions controversées. Parmi les points faibles soulignés, on peut noter le manque de participation et d'intégration du point de vue des partenaires et des patients et la faisabilité d'implanter les recommandations.

⁶¹ 18 chapitres de 2 guides de chiropractie pour les premiers et 11 guides pour les seconds. Le bilan de Cates *et al.* (2001) sur les deux guides de chiropractie est aussi assez négatif.

⁶² Littlejohns *et al.* (1999) dans leur revue de 45 guides sur la dépression ont conclu que malgré des qualités très variables, les performances résultantes étaient semblables.

Enfin, une dernière étude est d'intérêt pour nous puisqu'elle portait sur les chapitres d'un guide de médecine du travail⁶³. Les principaux résultats, semblables d'un chapitre à l'autre, sont résumés à l'encadré 2.4. Le faible résultat obtenu pour l'item rigueur vient du manque d'informations sur la méthode utilisée pour sélectionner les données probantes, sur les critères utilisés pour les trier, sur la méthode utilisée pour formuler les recommandations et établir des liens entre les données et les recommandations. Globalement, les experts ont cependant jugé le contenu supérieur à sa documentation sur son processus d'élaboration.

Encadré 2.4 - Évaluation des chapitres du Guide de l'American College for Occupational and Environmental Medicine (Cates *et al.* 2006)

Items	Moyenne
But et objectifs	80 %
Implication des partenaires	46 %
Rigueur de développement	27 %
Clarté	87 %
Applicabilité	31 %
Indépendance éditoriale	19 %

2.3 Les outils d'évaluation des guides

Nous avons abordé dans les sections précédentes le processus d'élaboration des guides, ainsi que l'utilisation réelle du processus recommandé. Les évaluations que nous venons de mentionner à la section précédente ont été effectuées avec des outils. Ces derniers constituent l'objet de cette section.

The purpose of evaluation is to determine what outcomes—both desired and undesired, anticipated and unanticipated—have occurred as a result of a policy or program. For practice guidelines, the primary outcome variables identified in the legislation are the quality, appropriateness, effectiveness, and cost of care provided to Medicare beneficiaries. However, evaluation that concentrates solely on ultimate outcomes and ignores intervening events may be incapable of distinguishing why a policy succeeded or failed... (Field et Lohr 1990, p. 91).

L'IOM a ainsi proposé dès le départ de distinguer entre l'évaluation des pratiques (*outcomes*) et l'évaluation des guides et des directives elles-mêmes. Beaucoup d'efforts ont alors été consentis au développement de procédures et d'instruments d'évaluation, instruments qui diffèrent selon les destinataires ciblés : des chercheurs/organisation/développeurs ou des praticiens. Les instruments proposés se distinguent aussi selon qu'il s'agisse d'évaluer le guide lui-même ou de recueillir les opinions des utilisateurs. Dans cette section, nous nous concentrons sur les instruments d'évaluation des guides destinés aux chercheurs, organisations et développeurs⁶⁴. Nous résumons à la section 2.3.2 les outils destinés aux praticiens en vue de les aider à discriminer entre les bons ou les moins bons guides.

⁶³ Plusieurs systèmes de compensation ont adopté des lignes directrices comme standard pour l'évaluation et la gestion des blessures de travail, en particulier pour ce qui est des troubles musculo-squelettiques.

⁶⁴ Nous avons exclu les manuels d'évaluation qui ont une portée plus large comme celui d'Eddy (1992) : *Assessing health practices & designing practices policies*. Ils témoignent de l'importance accordée au processus évaluatif.

2.3.1 Les outils d'évaluation destinés aux chercheurs et aux organisations

a) Portrait des instruments

Quatre groupes ont produit des instruments qui sont très référencés. Ces outils sont structurés autour des attributs et de leurs items d'évaluation. Tout comme pour le GsG, le document et l'instrument source (10 attributs, 142 items) à la base de tous les autres est issu du travail effectué dans le cadre de l'IOM (Field et Lohr 1990). Le second (37 items), de Cluzeau *et al.* (1999), a fait l'objet de nombreux travaux et a engendré plusieurs dérivés dont l'instrument AGREE⁶⁵ (The Agree collaboration 2003), une version simplifiée (23 items), développée par un groupe international et dont les échelles ont été validées. On recommande donc en général d'utiliser plutôt AGREE (Vlayen *et al.* 2005), diffusé sur Internet et périodiquement révisé (<http://www.agreetrust.org>). Le quatrième (25 items), de Shaneyfelt *et al.* (1999) est considéré comme un des mieux validés.

Nous en avons retenu quatre autres, parce qu'ils étaient soit souvent mentionnés, soit présentaient des différences intéressantes. Celui de Ward et Grieco (1996), australien, comporte sensiblement les mêmes éléments, mais la formulation des items tient mieux compte de certaines critiques (ex. flexibilité) et de certains aspects pratiques (ex. clarté). Le SIGN (2008), un des plus récents, est relié au groupe de développement des GsG. L'outil proposé par Grilli *et al.* (2000) est très court. Nous avons trouvé intéressant de présenter une version très simplifiée et synthétique. Enfin, le dernier instrument (33 items) a été développé à la demande de l'Association américaine de psychologie⁶⁶ (Reed *et al.* 2002). Il prend la forme d'une liste de contrôle. Les items sont sensiblement les mêmes que pour la santé, mais la façon de définir le contenu présente différentes nuances et variantes. Par exemple, la question sur la rigueur méthodologique fait référence au poids relatif accordé aux différentes sources (opinions cliniques, experts, littératures...). La question sur les résultats insiste sur la variété des éléments pris en compte (qualité de vie, conséquences à long terme, satisfaction, etc.). L'instrument accorde plus d'importance à la variable patient. À bien des égards, les items identifiés sont proches de la SST. Il offre un bon exemple d'un processus d'adaptation⁶⁷.

Le vocabulaire utilisé pour décrire ces outils présente des variantes, mais la logique demeure sensiblement la même. On parle d'attribut ou de sujet, et de critères ou d'explications qui s'y rattachent. Les attributs peuvent être regroupés en dimension. Chaque critère se déploie lui-

⁶⁵ Le lecteur intéressé pourra consulter l'article de Hannes *et al.* (2005) qui résume les essais de AGREE effectués auprès de 33 utilisateurs. Les auteurs ont organisé dans le cadre de cette étude divers ateliers d'échanges sur AGREE.

⁶⁶ L'APA distingue les guides de traitement des guides de pratique. Les premiers visent à fournir des recommandations spécifiques sur les interventions cliniques (en général basées sur des désordres spécifiques). Les seconds comportent des recommandations portant sur la conduite et les questions à prendre en compte dans un domaine donné de la psychologie. Les recommandations de l'American Psychologist Association sont résumées à l'annexe 2.4.

⁶⁷ Fondamentalement, la proposition de l'APA (2002) couvre les mêmes attributs, mais elle y explicite de façon plus claire des éléments soulignés dans les textes critiques en santé : faisabilité, délimitation des buts, documentation des points de vue alternatifs, exposition de la logique et des raisonnements, flexibilité et reconnaissance de la priorité du jugement du professionnel. Les notions d'efficacité, d'utilité, d'applicabilité, de faisabilité, de convivialité sont plus abordées. L'APA consacre ainsi une longue section à l'évaluation de l'efficacité et de l'utilité des traitements; elle convient qu'il n'est pas toujours possible de les évaluer et propose une stratégie centrée sur le niveau de confiance dans le jugement qu'on peut porter sur l'efficacité. Sa position sur la standardisation des pratiques est aussi très différente. Les variations sont vues comme des adaptations à des contextes spécifiques et non comme un problème à résoudre. Elle met en garde contre tout système rigide qui détruit l'habileté du praticien à répondre aux besoins individuels.

même à travers un ou plusieurs items. Un bref historique du développement des principaux instruments et la structure des instruments cités ci-dessus est présenté à l'annexe 2.2.

L'objet de la présente section n'est pas de comparer les instruments proposés, ce que plusieurs articles de synthèse⁶⁸ font très bien (ex. Graham *et al.* 2000, Vlayen *et al.* 2005). Il nous a semblé cependant pertinent de compiler les variantes. Le tableau de compilation (cf. annexe 2.3) a été structuré à partir du Cluzeau, car c'est le plus exhaustif et tous suivent sensiblement sa structure. Nous y avons ajouté deux autres références (Shiffman *et al.* 2003, Graham *et al.* 2000). Pour chacun des 37 items du Cluzeau, nous avons noté pour les autres instruments : 1) s'il était présent ; 2) si on proposait des variantes ou des éléments complémentaires ; 3) les explications fournies. L'instrument de l'IOM n'est pas inclus puisque tous se sont assez rapidement entendus sur sa trop grande complexité. Comme il s'agit du matériel source à partir duquel les critères et items ont été développés, il a été cependant utilisé pour compléter des explications. Ce tableau de compilation est proposé comme outil de travail aux chercheurs que la question d'évaluation des guides intéresse de près. Chacun pourra y trouver des points de repère ou de réflexion propres à son domaine d'intérêt. Pour notre part, nous en proposons ci-dessous une lecture globale.

Une lecture globale montre que les items évaluatifs s'articulent majoritairement autour de deux grands volets : 1) le processus : par exemple, comment les données probantes sont-elles identifiées? Comment le groupe est-il constitué? Quelles sont les méthodes pour proposer les recommandations? 2) Le contenu, à savoir les éléments qui doivent être présents dans le guide : description des patients ciblés, description des obstacles, évaluation des coûts, des bénéfices, etc. Peu d'items s'adressent en tant que tels au format et à l'organisation du contenu. Les quelques items qui traitent de ces aspects sont décrits de façon peu opérationnelle, c'est-à-dire sans critère d'évaluation. Par exemple : le format est-il convivial? Les recommandations sont-elles clairement présentées?

Ce qui distingue en fait le plus les instruments est la philosophie sous-jacente : certains privilégient l'approche « décrit-on? », d'autres l'approche « il doit y avoir... »

Globalement, les auteurs de ces instruments proposent des variantes pour près de la moitié des items, soit 16 sur 37. Il s'agit le plus souvent d'ajouts ou de précisions. Par exemple, pour l'item financement « des fonds ont-ils été reçus? », un instrument demande de préciser aussi le rôle ; à l'item « l'organisation responsable est-elle clairement identifiée? », un autre ajoute « l'organisation endosse-t-elle officiellement le guide? », « décrit-on son rôle? ». À l'item portant sur la « description des sources d'information utilisées pour choisir les données probantes à la base des recommandations », l'un demande si on a décrit la stratégie de littérature utilisée, un autre en fait un critère, à savoir la méthode utilisée doit être systématique. Dans la majorité des cas (10 critères), les divers auteurs proposent une ou deux variantes. Ce qui retient l'attention, c'est le niveau de détail scruté, qui est assez loin de la culture dominante en SST où par exemple les partenariats sont complexes. Les deux items pour lesquels les variantes sont plus nombreuses (au moins quatre variantes) sont la clarté des recommandations (item 28, relève du critère clarté) et la formulation des énoncés (item 33 sur la dissémination et l'implantation). Ces

⁶⁸ Graham *et al.* (2000) ont comparé 13 instruments (11 dimensions/attributs, 44 items) et Vlayen *et al.* (2005) 24, soit les 13 de Graham *et al.*, plus 11 autres sur 10 dimensions (50 items).

variantes sont présentées à l'encadré 2.5. En fait, au début de notre démarche, notre questionnaire fondateur avait été : quand on ouvre un guide, y a-t-il des critères qui permettent de statuer sur sa qualité? Nous pensions alors en particulier à la qualité de la présentation.

Encadré 2.5 - Variantes proposées pour les items portant sur le format et la formulation des énoncés

Item de Cluzeau <i>et al.</i> (1999)	Variantes*
Les recommandations sont-elles clairement présentées?	Le format est convivial. Les termes utilisés sont non ambigus, les définitions sont précises. Le résumé est structuré. Énonce clairement les actions recommandées, spécifie les circonstances.
Le document suggère-t-il des méthodes de dissémination et d'implantation?	Les barrières potentielles organisationnelles sont-elles discutées? Les barrières potentielles sont-elles décrites? Fournit-on des documents auxiliaires pour aider à l'implantation? Fournit-on un graphique (ou algorithme) des étapes et du processus décisionnel?

* La formulation peut prendre autant la forme d'une question que d'une affirmation. Nous avons uniformisé le format pour un item donné pour faciliter la lecture.

Quatrième constat, seulement le tiers (12) des 37 items sont repris par au moins cinq des instruments consultés (cf. encadré 2.6). Aucun item portant sur la dimension « application » (trois critères, cinq items), n'est repris par au moins la moitié des instruments. Le consensus sur la façon d'évaluer cette dimension est donc limité. Or, elle est centrale en SST.

Encadré 2.6 - Items de l'instrument de Cluzeau repris par au moins cinq des dix instruments analysés

Dimension/Critère	Item
I. Rigueur du développement	
1.2 Groupe qui développe	4. Décrit-on les individus?
1.3 Identification et interprétation des données probantes (DP)	6. Y a-t-il une description des sources d'information utilisées pour choisir les données probantes à la base des recommandations?
1.4 Formulation des recommandations	13. Le lien entre les principales R et le niveau des DP est-il explicite?
1.5 Révision pré-diffusion	14. Les directives ont-elles été revues par un processus indépendant avant leur publication/diffusion?
1.6 Mise à jour du guide	18. Fait-on mention d'une date de révision/mise à jour?
II. Contexte et contenu	
2.1 Objectifs	22. Les objectifs sont-ils clairement définis?
2.2 Contexte	23. La description des patients à qui s'adressent les directives est-elle satisfaisante? 25. Énoncé explicitant comment les préférences du patient seront prises en compte ? 26. Le G décrit-il sans ambiguïté les conditions à détecter, à traiter ou à prévenir?
2.3 Clarté	27. Les différentes options pour gérer la condition sont-elles clairement énoncées?
2.4 Coûts et bénéfices	29. Décrit-on adéquatement les bénéfices pour la santé attendus? 30. Y a-t-il une description adéquate des dommages et des risques potentiels?
III. Application	Aucun.

Dernière observation : les instruments proposés utilisent différents systèmes de réponses (cf. encadré 2.7). Sur les 24 instruments analysés⁶⁹ par Vlayen *et al.* (2005), dix utilisaient un format de réponse oui/non, cinq, une forme d'échelle et neuf, un système de cotation non spécifique.

⁶⁹ 6 États-Unis, 5 Canada, 4 Royaume-Uni, 2 Australie, 2 Italie, 1 France, 1 Allemagne, 1 Espagne et 2 internationaux. Tous étaient américains avant 1995.

Encadré 2.7 - Propositions d'énoncés pour évaluer directives et recommandations

Amstrong (2003)	Accord quant à la procédure ou le traitement
	<ul style="list-style-type: none"> • Accord général : utilisé et efficace. • Divergences ou conflits d'opinions sur le poids des évidences: (a) en faveur (b) pas bien établi. • Accord et évidences sur : généralement pas efficace/utile, peut occasionner des dommages.
Woolf (1995)	Gradation des évidences/preuves
	<ul style="list-style-type: none"> • Excellentes (<i>good</i>) • Suffisantes (<i>fair</i>) ... • Insuffisantes pour statuer sur le pour comme le contre. • Suffisantes ... pour rejeter/exclure . • Excellentes ...pour rejeter.

b) Évaluation des instruments

Une des revues les plus extensives menées sur l'évaluation des instruments (elle couvre plus large que les outils développés pour évaluer les guides) est celle de Katrak *et al.* (2004). Elle porte sur 121 outils d'évaluation (108 articles). Leurs constats sont que : 1) la construction des outils est très variable, que ce soit en matière d'intention, de composantes ou des propriétés psychométriques⁷⁰ ; 2) peu documentent la validité et la fiabilité de leurs items ; 3) la construction est généralement empirique ; 4) la grande majorité (n=104) est élaborée pour évaluer un design spécifique et peu sont génériques (n=16).

Dans le cas des instruments d'évaluation des guides, on s'entend sur le fait qu'il n'y a pas d'étalon-or et qu'il existe peu de données permettant de trancher entre eux (Graham *et al.* 2000). Cluzeau *et al.* (1999) font la mise en garde qu'un processus de développement apparemment rigoureux peut cacher des recommandations cliniques aberrantes.

Deux grands types de critiques sont destinés à ces instruments : le premier concerne la validité et la fiabilité des instruments proposés. Peu ont fait l'objet de procédures de validation spécifiques (exceptions : Cluzeau, AGREE – a validé les échelles numériques – et le Shaneyfield). Dans l'étude de Vlayen *et al.* (2005) menée sur 24 outils, quatre avaient fait l'objet d'études de validation. Les attributs (cf. tableau 2.1) de ces différents outils ont été testés par deux évaluateurs⁷¹. Ils ont été d'accord pour l'ensemble des items pour quatre des 24 outils. L'attribut où il y a eu le plus de désaccords est la flexibilité clinique (10 des 24 instruments). Pour chaque instrument, les deux évaluateurs ont été en désaccord sur 22 dimensions.

La seconde grande critique concerne le contenu évaluatif en lui-même : 1) peu d'items évaluent l'ensemble du contenu clinique ou la valeur clinique elle-même du guide (Graham *et al.* 2000, Vlayen *et al.* 2005) ; 2) la faisabilité de l'implantation est peu considérée (Graham *et al.* 2000) ; 3) le contenu peut être valable, mais ne pas être aussi utile qu'attendu et l'utilité ne peut être déterminée que par des tests sur la dissémination et l'utilisation en pratique (Graham *et al.* 2000).

⁷⁰ 12 % des 121 outils (n=14) ont été élaborés sur une base empirique (ex. : revue des outils existants, enquête Delphi, questionnaires, consensus... recherche de l'intrant de partenaires). 78 % (n=94) ont procédé d'un design d'étude primaire plutôt que secondaire (ex. méta-analyse). Un total de 1 475 items, regroupés en 173 types, ont été identifiés : les deux plus fréquents étaient l'évaluation de la validité externe (n=35) et la méthode d'analyse (n=28).

⁷¹ On a extrait des outils 469 énoncés ou questions, regroupés en 50 items/10 dimensions.

2.3.2 Les outils d'évaluation destinés au praticien pour décider s'il peut retenir ou non un guide

Le praticien doit aussi pouvoir évaluer les propositions d'un guide, mais il dispose de peu de temps. Il existe plusieurs propositions : nous présentons ci-dessous quatre orientations. Les deux premières sont résumées à l'encadré 2.8.

Woolf (1995) suggère une liste écourtée de questions où il s'agit essentiellement de vérifier si certaines informations sont fournies. La seconde proposition (Hayward *et al.* 1995) va plus loin – trop loin pour des praticiens? - en suggérant tout un ensemble d'éléments que ce dernier devrait jauger. L'outil statue sur ce qui doit être présent. Par exemple, le guide devrait préciser les critères d'inclusion et d'exclusion des articles.

La proposition de Brosek *et al.* (2009) reprend les mêmes éléments, mais de façon plus explicative et moins prescriptive. Quant au questionnaire proposé par Mottur-Pilson (1995), il amène en fait le futur usager à réfléchir sur l'applicabilité du guide et sa clarté. Le guide est alors considéré comme un outil de communication (Mottur-Pilson 1993), une dimension importante pour les guides en SST.

Encadré 2.8 - Liste de contrôle proposée pour permettre aux praticiens d'évaluer un guide

Woolf (1995)	Hayward <i>et al.</i> (1995)
<ul style="list-style-type: none"> • Qui a développé (organisation, table ronde /composition) ? • Âge du guide. • Méthodologie décrite? • Comment les données probantes ont-elles été évaluées : revue de littérature, classification des études, prises en compte des sources de biais, méthodes pour synthétiser les données? • Comment l'opinion des experts/cliniciens a-t-elle été évaluée? • Comment les décisions ont-elles été prises : documentation de l'argumentaire, règles de décision, liens entre la force des recommandations et la qualité des données probantes, caractère approprié du langage utilisé pour formuler les recommandations? • Les questions de politique publique (en santé) sont-elles prises en compte? • Faisabilité considérée? Capacité du Md de l'implanter (temps, simplicité, remboursements). • Évalué par un groupe d'experts? • Comment ce GdP est-il relié aux autres G? Conflits? Impact sur d'autres interventions? 	<ul style="list-style-type: none"> • (Le groupe...) a utilisé des méthodes appropriées... pour appuyer les recommandations. • Les autres options ont été considérées ainsi que tous les résultats potentiels (et conséquences ; parle de question des coûts en tant que critère de décision). • Méthode, processus de recherche utilisés pour identifier, sélectionner et combiner preuves explicitées? (ex. critères inclusion et exclusion) • La nature et la pertinence d'expertise dans le développement des guides sont chaudement débattues. RAND Corporation a développé un protocole <i>for recording and quantifying expert assessments</i>. Il est peu probable de trouver des informations sur comment les experts ont été choisis, mais on devrait savoir comment ils ont été utilisés pour palier à ce qui manque. • Résultats (<i>outcomes</i>) : la valeur qu'on y attribue dépend des gens et est relative. Mais les développeurs devraient rapporter la source de leur jugement de valeur et la méthode de consensus. Propose des éléments qui portent sur la composition de la table ronde (suppose que les praticiens connaissent les membres !!). Les méthodes informelles de discussion peuvent induire des biais et ouvrir la porte à l'influence individuelle. • Comment les préférences du patient ont-elles été prises en compte? • Les dates sont importantes pour intégrer les dernières données. • Le guide a-t-il été soumis à une revue de pairs et testé?

2.4 Tendances et questions de l'heure pour les guides de pratique

2.4.1 Informatique et technologies de l'information (TI)

La littérature récente aborde de plus en plus la question de l'intégration des guides aux TI et vice versa. La question du développement de logiciels et de règles de syntaxe, le choix du langage, est

abordée depuis une quinzaine d'années⁷². Les chercheurs tentent de développer des logiciels (ex. ALCHIMIST de Sanders *et al.* 2000) centrés sur les modèles décisionnels afin de répondre à une grande critique adressée aux guides, soit celle de ne pas correspondre suffisamment au processus clinique de décision⁷³. Les discussions en sont au stade du développement de systèmes d'adaptation semi-automatiques⁷⁴. Il faut signaler ici que le format énoncé préconisé par les guides favorise l'adaptation informatique. Son efficacité dépendrait cependant des domaines d'application (Bero⁷⁵ *et al.* 1998).

Ce qu'il faut en retenir, c'est que ces systèmes fonctionnent sur la base de règles⁷⁶. Les connaissances sont manipulées de façon déclarative. Les nœuds des algorithmes sont les conditions des règles et les embranchements représentent les décisions et les conséquences potentielles. Ces algorithmes sont souvent fondés sur les graphiques de cheminement⁷⁷. Un des avantages du passage à l'informatique serait que cela oblige à préciser ces éléments, celle-ci tolérant mal les omissions, les trous et le manque d'explicitations (Shiffman *et al.* 2004)⁷⁸. Cela présente toutefois un désavantage : la rigidité.

2.4.2 Mises à jour et coûts

La maintenance et la mise à jour des guides sont très onéreuses et nécessitent des infrastructures et des ressources importantes (Lugtenberg *et al.* 2009). La mise à jour individuelle, avec les changements de pratique que cela peut impliquer, présente aussi un défi majeur (Woolf 1995). Les solutions préconisées sont l'établissement de partenariats, l'organisation d'échanges et de coopérations pour éviter la duplication d'efforts et le partage des coûts⁷⁹. L'approche préconisée est d'adapter les guides aux besoins locaux (Browman 2000, Carlsen *et al.* 2010). Adapter suppose cependant souvent un rapport aux connaissances centralisateur et hiérarchique.

2.4.3 Équilibre simplicité/complexité, standardisation/flexibilité

Les praticiens veulent des guides simples, rapides et pratiques à consulter (Carlsen et Kjellberg 2010) avec pour défi d'éviter de transformer les guides en « livres de recettes » (Woolf 1995). Cela correspond au mode de fonctionnement dit heuristique⁸⁰ qui permet de simplifier le processus décisionnel (Baunilauri Ferreira *et al.* 2010). Les guides trop simples sont perçus en contrepartie comme mal adaptés aux situations complexes. Les praticiens apprécient les

⁷² Le lecteur intéressé peut consulter Ohno-Machado *et al.* (1998).

⁷³ D'après leurs auteurs, les GdP qui en découlent sont de meilleure qualité, car les algorithmes et les graphiques de cheminement obligent à préciser les options, les stratégies recommandées, les résultats attendus, les risques, les coûts, les bénéfices.

⁷⁴ Le lecteur intéressé peut consulter par exemple les travaux Ongenae *et al.* (2010).

⁷⁵ Ex. : améliore la performance dans des cas comme le dosage des médicaments, dans la gestion des patients, mais pas dans le cas du diagnostic.

⁷⁶ Même principe que pour les systèmes experts, dont le développement dans le domaine médical remonte à l'implantation des ordinateurs personnels. Le but était d'implanter ces systèmes dans les pays où les ressources médicales étaient limitées.

⁷⁷ Cook *et al.* (1997) distinguent les notions de cheminement et d'algorithme en matière de complexité : « *Clinical Pathway : organize, sequence and time the care given to a typical uncomplicated patient. Clinical algorithm : a set of more complex instructions for addressing a particular issue...* » Les graphiques de cheminement sont très utilisés en génie industriel.

⁷⁸ Les auteurs proposent aussi une méthode pour traduire les connaissances en graphique de cheminement (*flow chart*). Les buts sont d'améliorer la spécificité, de « désabstraire » les données en les rendant opérationnelles, de diminuer l'ambiguïté, de vérifier si elles sont complètes (y compris en matière d'explications) et de pouvoir élaborer les énoncés d'exécution. Il s'agit en quelque sorte d'une méthode de *translation* qui pourrait servir de cadre d'analyse aussi.

⁷⁹ D'autres approches sont aussi proposées. Ex. : Gartlehner *et al.* (2004) combinent expertise et recherche de nouvelles données.

⁸⁰ Un peu comme une règle du pouce.

procédures standardisées, mais ils insistent sur l'individualité des patients. En fait, Burgers *et al.* (2003) estiment que les recommandations ne devraient s'appliquer qu'à seulement 60 à 90 % et qu'il peut être tout à fait approprié de ne pas être d'accord.

Carlsen et Kjellberg (2010), qui font un retour historique sur l'élaboration des GdP en Angleterre, résumant bien ces dynamiques de tension et comment cela se reflète dans les décisions⁸¹. Les auteurs expliquent que le modèle biomédical basé sur les pratiques en milieu hospitaliers s'est imposé dès la fin du 19^e siècle et que les praticiens ont cherché à s'y associer progressivement pour maintenir leur statut. Une vague « anti modèle biomédical », en partie associée à l'organisation et à la prise de parole des patients (*patient empowerment*), a conduit au développement d'approches holistiques plus centrées sur le patient. Par la suite, les autorités de santé sont revenues à des approches plus médicales axées sur la notion d'*EBM*, autant pour standardiser que réguler la pratique médicale. D'après les études récentes, les praticiens adhèrent aux efforts de standardisation et de rationalisation menés à travers les GdP, même s'ils se défendent de maintenir une approche holistique.

2.5 Les guides en SST

Comme nous l'avons vu au chapitre précédent, les guides de pratique en SST prennent de multiples formes et le format énoncé ne peut convenir qu'à une fraction d'entre eux. L'ampleur des ressources financières consacrées au domaine de la santé n'a pas non plus de commune mesure avec ce qui est disponible en SST.

Le développement des guides en SST implique aussi régulièrement des partenariats complexes, à géométrie variable et le développement peut être étalé dans le temps. On en trouvera plusieurs exemples dans Desmarais et Lortie (2011) : le développement d'un guide sous forme de vidéo pour les jockeys (Brunet, p. 99), la grille d'auto-évaluation des risques destinée aux centres de tri de matières recyclables (Lavoie *et al.*, p. 79), le programme de formation à l'aiguisage de couteaux pour les abattoirs (Desmarais *et al.*, p. 169), le développement d'un guide de sécurité sur les convoyeurs à courroie (Trépanier *et al.*, p. 45). Plusieurs de ces projets se sont échelonnés dans le temps et résultent de multiples allers-retours. Les chercheurs y sont très impliqués, car ces guides sont le plus souvent un résultat de leur propre travail. Ces guides ont un caractère très contextuel et la connaissance développée est elle-même reliée à ces contextes. C'est particulièrement le cas des guides de prévention qui doivent tenir compte à la fois du secteur de production, des moyens de prévention dont dispose l'entreprise, et de la capacité du travailleur à contribuer à l'implantation des guides et à leur utilisation. Des barrières de langues, d'alphabétisation, de conception de la SST ou de son utilité sont autant de facteurs de succès ou d'échec à l'implantation de guides dont les concepteurs cherchent à tenir compte d'emblée. La problématique dominante est en fin de compte moins celle de la généralisation que du transfert d'un contexte à l'autre et du potentiel d'ajustement.

Ces guides coexistent cependant avec des guides qui visent d'emblée une utilisation généralisable, décontextualisée. Le *Quick Exposure Check-list* utilisé par la Commission de la santé et de la sécurité du travail (CSST) en est un bon exemple. Ce guide développé en Angleterre a été adapté aux fins d'utilisation par les inspecteurs. La question des contextes y est posée plus en termes de limite d'interprétation que d'adaptation.

⁸¹ Fait référence à des travaux menés sur l'évolution de la gestion des pratiques professionnelles et de l'identité professionnelle.

La question de mise à jour est aussi présente, mais elle se pose souvent plus en termes de mise au rancart, de pérennité que d'actualisation. En fait, c'est souvent l'évolution des technologies des environnements de travail qui rend caduques les guides produits. C'est en bonne partie l'évolution des contextes plus que celle des connaissances qui détermine la durée de vie d'un guide.

Enfin, il faut signaler que les caractéristiques des utilisateurs des guides en SST couvrent un spectre beaucoup plus large qu'en santé. Cela va orienter la façon d'organiser les développements, les groupes consultés et les stratégies adoptées (ex. importance du visuel).

3 POSITIONNEMENT SUR LES GUIDES ET FACTEURS QUI ONT UN IMPACT SUR LEUR QUALITÉ

Ce chapitre est centré sur les guides en tant qu'objet de recherche. Nous traçons d'abord un portrait des opinions émises sur les guides, puis nous présentons les facteurs susceptibles d'avoir un impact sur leur qualité.

3.1 Positionnements sur les guides

Les opinions émises ont été réparties en trois sections. La première regroupe surtout des articles d'opinion, tels les éditoriaux ou les articles de synthèse, dans lesquels les auteurs exposent un point de vue critique sur les guides et/ou sur les points à améliorer. La seconde section regroupe des enquêtes menées auprès des usagers ; elles permettent de saisir les inquiétudes que soulève l'implantation des guides. La dernière section résume les souhaits des deux parties.

3.1.1 Critiques et demandes

On s'entend globalement autant sur les qualités désirées des guides que sur les critiques qui leur sont destinées. Le lecteur reconnaîtra que la plupart des points critiqués sont abordés dans les GsG, dans lesquels plusieurs des étapes proposées visaient à corriger les lacunes soulevées. En fait, plusieurs des critiques portent implicitement sur le fait que les GdP n'atteignent pas leurs objectifs de correction ou de modification des pratiques. Ces critiques sont maintenant détaillées.

a) Être plus centré sur le patient

Cette première critique porte sur le passage difficile des données portant sur un ensemble de sujets à une décision qui concerne un individu spécifique. Comme le rappelaient Jackson et Feder (1998) dans un éditorial du *BMJ* « *one of the cornerstones of EB practice is the requirement that the evidence is relevant to individual patient* ». Il s'agit sans doute de la principale critique adressée aux guides : les GdP ne s'appliquent pas assez à l'individu et n'intègrent pas suffisamment les facteurs de décision prise en compte dans la réalité (Armstrong *et al.* 2004). La critique estime que bien que la majorité des données soient présentées de façon compréhensive et claire, le format ne permet cependant pas de les situer suffisamment dans un contexte de soins individuels, puisqu'elles sont présentées sous l'angle de risques relatifs alors qu'en pratique la décision est basée sur le risque absolu, individuel. Ce qui est mieux pour un ensemble peut être inapproprié pour un patient spécifique (Woolf *et al.* 1999). « *The more guidelines can present evidence which can be applied to individual, the more useful they will be for real life clinical decision making. Explicit statements about the benefits and risks of treatments can then be weighed by patient preferences and available resources.* » (Dowie 1996).

Une seconde critique importante concerne le fait que les guides ne prennent pas suffisamment en compte l'ensemble des facteurs que le praticien doit considérer dans sa pratique, tels le point de vue du patient, ses valeurs, les coûts et surtout la comorbidité (Lugtenberg *et al.* 2009). Les enquêtes menées auprès des usagers montrent que les praticiens considèrent la dimension relationnelle centrale à leur pratique. Or, les recommandations peuvent interférer avec le point de vue du patient (Carlsen et Kjellberg 2010).

Enfin, une des questions soulevées par ces critiques est celle du choix des sujets dans les études. Comme le signalent Audet *et al.* (1990), les études sont centrées sur des catégories de patients

(ex. : sexe, âge, ethnie) qui les définissent de façon tronquée ou trop étroite pour refléter leur profil réel. La critique remet ainsi en question la signification des données probantes à partir desquelles se prennent les décisions.

b) Être plus proche du processus de décision

La question du patient est étroitement liée à celle de la décision où cette dernière nécessite de prendre en compte un ensemble de facteurs. La pratique comporte en fait deux activités conceptuellement distinctes : prendre une décision et passer à l'action (Dowie 1996). Les GdP issus de l'*EBM* ne seraient pas bien adaptés à ces activités, celles-ci étant basées sur des processus de résolution de problèmes et non de décision⁸², les deux relevant de deux paradigmes différents. Il serait nécessaire de mieux comprendre le processus de décision (Browman 2000) pour pouvoir élaborer des GdP efficaces. De plus, la décision peut être aussi un processus collectif ou institutionnel. Il ne s'agit pas alors nécessairement de prendre la meilleure décision pour un patient, mais de prendre une décision optimale par rapport à un ensemble de facteurs, dont les coûts (Woolf *et al.* 1999).

Par ailleurs, on constate que les travaux sur les processus de décision se sont multipliés ces dernières années. Ils sont associés au développement d'algorithmes et de propositions informatiques, comme on l'a vu précédemment. L'élaboration d'algorithmes présente cependant le risque de réduire le processus de décision à des séquences de choix binaires traduisant mal la complexité des décisions à prendre et les processus mis en jeu, souvent itératifs et parallèles (Woolf *et al.* 1999). En fait, pour l'instant, les solutions proposées sont surtout centrées sur la traduction des connaissances en graphique de cheminement pour développer des outils d'aide à la décision, des processus bien connus par les gestionnaires d'entreprises et en génie.

c) Être plus proche des conditions d'application

Une troisième critique formulée à l'égard des guides a trait au fait que les données probantes ne sont pas suffisamment développées pour s'appliquer au contexte où le problème se pose. Les résultats de l'analyse de Dowie *et al.* (1998) sur 193 articles portant sur la gestion des victimes d'arrêt cardiaque et sur 51 traitant des cas de douleur aiguë à la poitrine offrent un bon exemple de cette critique. Les auteurs voulaient vérifier si les résultats s'appliquaient au cas de la gestion ambulancière. Dans le premier cas, il s'est avéré que 23 % des études pouvaient s'y appliquer, dans le second cas, 2 %. Un corolaire de cette critique est que les chercheurs s'intéressent possiblement trop à la question d'applicabilité sous l'angle des obstacles (ex. équipements ou ressources non disponibles) et pas assez à celle de la conformité aux conditions d'application telles que rencontrées dans la réalité (Audet *et al.* 1990). Ces conditions variables nécessitent de prévoir des alternatives, ce que les GdP ne feraient pas suffisamment.

Globalement, les difficultés d'applicabilité sont considérées comme des obstacles importants à l'implantation des guides (Lugtenberg *et al.* 2009). Dans l'étude de Burgers *et al.* (2003) menée auprès de 200 praticiens hollandais, sur un total de 96 recommandations issues de 12 GdP, l'applicabilité ressortait en effet comme le facteur dominant en ce qui concerne l'adhésion. S'il y

⁸² Dowie (1996) définit ainsi la pratique : « *Taking into account and bearing in mind* ». Il cite les questions de coûts et de valeurs comme exemple d'éléments qui s'intègrent mal dans un processus de résolution de problème, mais naturellement dans un processus de décision. Kellett et Clarke (1995) ont ainsi introduit la notion de *Decision Analysis Based Medical Decision Making*.

a un bon consensus sur la nécessité d'accorder plus d'importance aux questions d'applicabilité et aux partenaires (Burgers *et al.* 2004), le « comment » est cependant moins consensuel. Par exemple, certains suggèrent de centrer le contenu des GdP sur la gestion d'un problème plutôt que sur l'application ou non de procédures (Audet *et al.* 1990), et d'inclure plus de considérations sur les barrières potentielles (Harpole *et al.* 2009). D'autres le voient comme un problème d'adaptation à des conditions locales, la majorité des guides nécessitant des adaptations (Cook et Giacomini 2009).

Sankaran (2006) pose la question un peu différemment en insistant sur l'écart entre les connaissances scientifiques et les pratiques organisationnelles⁸³. Le débat sur le fossé « académique vs pratique »⁸⁴ est posé sous l'angle de débat entre rigueur et pertinence. Il soulève les différences ontologiques et épistémologiques entre les objectifs de généralisation des scientifiques et les préférences des praticiens pour des solutions particulières (Aram et Salipante 2003) et qui présentent des avantages économiques. En fait, une partie de ces éléments sont intégrés dans les GsG (ex. liens avec les partenaires, identification des besoins), mais sans appui sur un cadre théorique conséquent. La proposition de Gibbons⁸⁵ *et al.* (1994) de différencier les modes de production des connaissances selon qu'ils soient théoriques et disciplinaires (Mode I) ou centrés sur un problème et transdisciplinaires (Mode II) est aussi intéressante. Selon Laroche (2009), ce modèle est passablement révélateur de l'organisation de la recherche en SST. Dans ce domaine, au Canada, 17,5% des chercheurs interrogés auraient le Mode I comme mode de production de la recherche, alors que 39,6% auraient le Mode II et 40,6% un mode hybride, situé entre les deux. Quant aux usagers, leurs critiques réfèrent surtout à des aspects pratiques tels : temps disponible, compétences nécessaires, accessibilité aux ressources, accessibilité des guides eux-mêmes (Carlsen et Kjellberg 2010). Comme on le verra au chapitre 5, les stratégies d'implantation ciblent particulièrement ces éléments, bien documentés par les études sur l'utilisation.

d) Être plus holistique

Les GdP sont souvent développés par des comités d'experts qui ont des intérêts spécifiques pour un problème clinique. Le focus serait trop étroit et pas suffisamment sensible (*awareness*) aux autres problèmes, y compris ceux reliés à la santé (Woolf 1995). Comme on l'a vu au chapitre 2, les GsG ont cherché à répondre à cette critique en élargissant les processus de consultation⁸⁶.

e) Reconnaître les risques d'erreurs et de dérives

La reconnaissance des risques d'erreurs et de dérives est davantage une limite formulée qu'une critique formelle. Il s'agit en fait d'une demande, soit celle de reconnaître qu'une recommandation puisse être erronée ou inappropriée pour un patient donné, que le passage des

⁸³ Les auteurs cités sur lesquels s'appuie son argumentaire sont très identifiés à la littérature sur le transfert : Rynes, Schön, Barley...

⁸⁴ Pour une discussion plus complète de cette question, le lecteur peut consulter l'article de synthèse d'Avenier et Bartunek (2010). *Bridging a supposedly unbridgeable gap : Elaborating scientific knowledge from and for practice.*

⁸⁵ Voir : http://en.wikipedia.org/wiki/Mode_2. Gibbons *et al.* sont connus pour leurs écrits sur les modes de production des connaissances. Le mode dit « 2 » situe la production des connaissances dans un contexte d'application. La distinction recherche appliquée vs fondamentale est considérée non pertinente. Cette recherche s'inscrit automatiquement dans un cadre de collaboration et doit prendre en compte les valeurs et les besoins des partenaires. Elle est située comme étant transdisciplinaire. Pour un résumé simple, voir Aasen *et al.* (2010) et Gibbons (2001).

⁸⁶ Les réflexions menées sur la multidisciplinarité vs l'interdisciplinarité vs la transdisciplinarité sont pertinentes à cet égard. Les propositions des GsG s'apparentent à de la multidisciplinarité où les modèles additifs ne sont pas toujours intégratifs.

données aux recommandations puisse être erroné. Woolf *et al.* (1999) rappellent que les recommandations sont influencées par le point de vue des membres du groupe de développement et que ces derniers peuvent mal estimer certains aspects. Des jugements portés sur la pratique professionnelle à partir des GdP pourraient aussi mener à des erreurs malencontreuses.

f) Conflits d'intérêts

Eichacker et Natanson (2008) donnent une série d'exemples illustrant des conflits d'intérêts, en particulier ceux liés au secteur privé⁸⁷. Leurs recommandations placent la question des conflits au cœur des GdP : « *Professional societies and other stakeholders must work together to promote a consistent guideline-development process, a robust rating system for guidelines that is applicable to all subspecialties and a policy that prohibits the pharmaceutical and medical-device industries from directly or indirectly funding or influencing practice standards. The challenges involved in producing first-rate guidelines and performance standards are only exacerbated by the intrusion of marketing strategies masquerading as evidence-based medicine* » (Eichacker *et al.* 2006). Les problèmes rapportés ne sont pas spécifiques au domaine de la santé, mais ils sont particulièrement aigus dans les secteurs où il y a des intérêts économiques. Ainsi, l'analyse de Culleton (2009) sur 44 G a montré que respectivement 87 % et 55 % des auteurs étaient en interaction avec l'industrie pharmaceutique et avaient des relations avec les compagnies dont les médicaments étaient examinés⁸⁸. En SST, les occasions de conflit d'intérêts sont plus rares. Cependant, il peut exister un pôle de tension syndicat – patronat qui peut influencer le contenu d'un guide selon la proximité des auteurs pour l'un ou l'autre.

g) Coûts associés aux recommandations et aux alternatives

Les auteurs demandent de fournir les informations qui permettent d'évaluer les bénéfices attendus, mais aussi les risques et les coûts associés, avec des alternatives permettant la résolution de problèmes (Burgers *et al.* 2003). Cela rejoint la demande de base des utilisateurs d'être avant tout informés pour pouvoir prendre des décisions éclairées⁸⁹.

h) Respect des règles de l'art quant au développement

Le reproche adressé est de ne pas spécifier suffisamment le contenu et comment le guide a été développé (Shaneyfelt *et al.* 1999). C'est en bonne partie pour répondre à cette critique, à laquelle on a accordé beaucoup d'importance, que les GsG ont été élaborés.

3.1.2 Craintes et inquiétudes des usagers

Diverses craintes et inquiétudes ont également été émises à l'égard des guides. Notamment, les éléments suivants ont trouvé écho chez certains auteurs :

⁸⁷ *Medical Guidelines and Performance Measures: The Need to Keep Them Free of Industry Influence*. Pour le lecteur intéressé à cette question, voir : Kassirer (2004), Schunemann *et al.* (2006), Harris et Roberts (2007), Egilman et Druar (2011); site web Institute of medicine as a profession - marketing and medicine, http://www.imapny.org/media_library/marketing-and-medicine.

⁸⁸ Un résultat semblable est rapporté par Choudhry *et al.* (2002) : 55 % des auteurs avaient des liens financiers avec le secteur pharmaceutique.

⁸⁹ Des approches évaluatives sont proposées dont les « *balance sheets* » par Eddy (1990).

Perte d'autonomie et rigidité. La perte d'autonomie et la rigidité associées à l'application des guides sont vues comme diminuant la satisfaction associée au travail de professionnel et à la pratique (Striker *et al.* 1999, Hayward *et al.* 1997).

Peut réduire le rôle du jugement clinique. Le jugement clinique est au cœur de la pratique professionnelle. La crainte manifestée dans la littérature est associée au fait que le GdP ne se substitue au jugement professionnel (Berg *et al.* 1997). C'est la raison pour laquelle les auteurs insistent sur l'adoption d'une formulation qui reflète clairement le caractère non coercitif des recommandations⁹⁰. Par exemple, pour Browman (2000), on devrait parler de « *overall concordance* » et non de « *compliance* ». Le conditionnel « *could* » devrait primer sur le « *should* ». Dans le domaine de la SST, l'OIT a pris soin de souligner le caractère volontaire et non contraignant des GdP dans leur application. L'OIT souligne qu'ils n'ont pas vocation à se substituer aux normes, à la réglementation ou à la législation en vigueur (ILO-OSH 2001).

Impact légal et utilisation coercitive. Les praticiens (surtout américains) craignent que les GdP ne puissent être utilisés dans un cadre légal. Les propositions d'utiliser des incitatifs légaux ne rassurent pas. Par exemple, des États américains requièrent la « *compliance* » avec les GdP pour obtenir la licence. Les praticiens craignent que leur utilisation soit invoquée par la justice et par les compagnies d'assurance (Woolf 1995, Striker *et al.* 1999).

Intérêt et crédibilité. Dans la mesure où les enjeux économiques et de rationalisation sont très présents dans le système de santé, les praticiens craignent que cela teinte le jugement ou les actions posées par des organisations publiques. Les GdP seraient utilisés pour appliquer un agenda gouvernemental. Ces méfiances dépendent cependant des contextes socioculturels. Par exemple, dans l'étude de Carlsen et Kjellberg (2010), les praticiens danois se situaient en tant que courroie de transmission entre l'État et le patient, alors que leurs confrères norvégiens se voyaient plutôt comme un rempart de protection entre le patient et l'État. Les études menées au Canada montrent que les praticiens québécois sont moins réticents que ceux des autres provinces face aux organisations gouvernementales (Kryworuchko *et al.* 2009).

3.1.3 Désirs et souhaits

On s'entend bien sur les qualités recherchées pour la qualification des bons guides, soit d'être clairs, concis, compréhensifs, consensuels, d'un coût acceptable, contemporains et centrés sur le patient (Armstrong 2004). Les critiques adressées ci-haut ciblent ainsi des déficiences qui portent sur les qualités nécessaires pour que les GdP soient acceptés, utilisés et aient de l'impact : clarté, spécificité, flexibilité et validité (Audet *et al.* 1990). Les praticiens désirent aussi des GdP courts, pratiques, informatifs et qui évitent les « certitudes pieuses » plutôt que des guides précisant ce qu'il faut faire (Cook et Giacomini 1999).

3.2 Facteurs qui ont un impact sur la qualité des guides

Plusieurs études ayant montré que la qualité des guides varie (Striker *et al.* 1999, Burgers *et al.* 2003), dans ce rapport, nous nous sommes intéressés à vérifier les facteurs qui influent sur cette qualité.

⁹⁰ Reed *et al.* (2002) en font une recommandation formelle dans le cas des GdP de l'Association américaine de psychologie, à savoir : un guide devrait mentionner qu'il n'est ni obligatoire, ni exhaustif, ni définitif, ni applicable à toute situation clinique et qu'il n'a pas préséance sur le jugement professionnel (cf section 2.3.1).

3.2.1 Date

Parmi les 855 guides recueillis chez les praticiens par Hibble *et al.* (1998), 62 % indiquaient la date de publication⁹¹. D'après l'étude de Burgers *et al.* (2003) qui ont évalué la qualité de 86 guides, l'année de publication (1992 vs 1999) a peu d'impact. Kryworuchko *et al.* (2009) qui ont comparé 1 664 guides canadiens diffusés au cours 1994-99 vs 2000-05 arrivaient à la même conclusion. Cependant Shaneyfelt *et al.* (1999), qui ont évalué 279 G, observent que le pourcentage d'adhésion aux recommandations est passé de 37 % pour les G édités en 1985 à 57 % en 1992. Ce serait donc l'attitude face aux GdP qui aurait changé plus que leur contenu.

On s'entend aussi sur le fait que les guides ont une durée de vie limitée et qu'ils nécessitent d'être mis à jour. D'après l'étude de Shekelle *et al.* (2001), la durée de vie d'un GdP est de moins de six ans pour 50 % d'entre eux⁹². Parmi les 17 guides évalués dans cette étude (1990 à 1996), 7 avaient besoin d'une révision majeure et 6, d'une mineure.

3.2.2 Organisation impliquée

a) Nationale vs locale

Les études tendent à montrer que les organisations gouvernementales produisent des guides de qualité supérieure⁹³ et que les guides nationaux sont de meilleure qualité que les guides locaux⁹⁴ (Berti et Grilli 2003, Burgers *et al.* 2003, Grilli *et al.* 2000, Cluzeau *et al.* 1999). Cependant, Fischer Walker *et al.* (2006), qui ont comparé l'efficacité de sept guides auprès d'une population de 410 enfants égyptiens⁹⁵, ont constaté que les guides développés en Égypte étaient parmi les plus performants⁹⁶. Les auteurs croient que le contexte de développement a permis de prendre en compte des éléments spécifiques aux enfants égyptiens. Dans l'ensemble, les résultats de certaines études permettent de penser que les guides locaux intègrent en effet des éléments de connaissance qui leur sont propres. Par exemple, Christiaens *et al.* (2004) ont observé dans une étude comparative de quatre guides (sur la cystite) qu'il existe de bonnes différences quant au diagnostic (différences de point de vue sur la sécurité, sur les meilleures données pour prendre une décision...). Le facteur local jouerait un rôle dans la production de guides, dans la sélection de la littérature et dans la formulation des recommandations. Les auteurs, qui ont contacté les développeurs pour obtenir leur point de vue, concluent que la littérature locale correspond peut-être mieux aux conditions locales et que les différences observées étaient liées à la prise en compte de facteurs locaux plutôt que résultant d'un processus différent. Une même donnée peut être utilisée différemment selon le cadre d'application. Les guides locaux, qui ne sont généralement pas publiés dans des revues de pairs (Pencharz *et al.* 2002), représentaient 60 % des guides recueillis par Hibble *et al.* (1998) auprès de praticiens.

Au Canada, sur les 1 664 guides étudiés par Kryworuchko *et al.* (2009), 1 232 étaient disponibles en anglais (74 %), 24 en français (1 %) et 304 dans les deux langues. Les auteurs ne précisent

⁹¹ 57 : premier trimestre de 1997; 138 : 1996; 73 : 1995; 8 : 1990.

⁹² Étude menée par questionnaire postal (n=121 répondants, taux de réponse : 71 %). Taux de survie calculé : 90 % valides après 3,6 ans, 80 % après 4,4 ans et 50 % après 5,8 ans.

⁹³ Burgers *et al.* (2003) ont évalué 86 guides provenant de 11 pays et Cluzeau *et al.* (1999), 53. La notion de « local » peut varier selon les contextes.

⁹⁴ Les auteurs font référence à des études menées au Royaume-Uni et en Finlande.

⁹⁵ Les G portaient sur les règles pour prévoir s'il est nécessaire de procéder ou non à un test de culture pour une pharyngite afin de déterminer s'il faut prescrire ou non des antibiotiques.

⁹⁶ Le guide le moins performant dans ce cas a été celui de l'OMS.

pas si l'organisation était nationale ou régionale, le contexte de développement des guides bilingues ou si la langue a un impact sur le contenu des guides.

b) Professionnelle vs gouvernementale

D'après l'étude de Stricker *et al.* (1999) les guides issus d'organisations professionnelles utilisent moins un contexte de développement interdisciplinaire que les agences gouvernementales. Ceci les rendrait plus vulnérables à des biais médicaux (ex. accent sur des traitements médicaux). Par contre, les cliniciens américains consultés par Hayward *et al.* (1997) ont plus confiance dans les guides produits par des organisations professionnelles que gouvernementales.

c) Spécialisation de l'organisation impliquée

Les études de Burgers *et al.* (2003, 2004) amènent un élément intéressant dans l'étude spécifique de l'applicabilité. Une première conclusion est que le type d'organisation n'a pas beaucoup d'impact sur l'applicabilité en matière de qualité recherchée (2003). Cependant, une étude subséquente (2004) menée sur des guides en oncologie – selon qu'ils provenaient d'une organisation exclusivement dédiée aux cancers ou non – a montré que les premiers obtenaient des scores plus élevés pour la rigueur, mais moindres pour l'applicabilité. Les auteurs émettent l'hypothèse que les organismes plus généraux ont davantage l'habitude du travail multidisciplinaire et que cela influencerait sur l'applicabilité.

3.2.3 Objet du guide

Les facteurs que nous venons de voir ne peuvent à eux seuls expliquer la variation observée de la qualité des guides (Striker *et al.* 1999, Burgers *et al.* 2003). On a donc examiné si le champ d'application ou l'objet avait un impact sur la qualité (inter guide), mais aussi sur sa variabilité (intra guide). Il semble que oui. Par exemple, dans l'étude de Burgers *et al.* (2003), le niveau de soin concerné (primaire, secondaire, tertiaire), l'objectif poursuivi (prévention, diagnostic, traitement, mixte), la présence ou non d'un programme structuré sur la qualité des guides affectaient la rigueur du développement. Une étude comparante 11 GdP en oncologie avec des GdP qui ne sont pas en oncologie a aussi donné des profils de scores bien différents (Watine *et al.* 2006). Concernant un même type de guide, Adams *et al.* (1999) ont trouvé qu'il y avait peu de différences de qualité pour les cinq guides d'hypertension étudiés alors que Harpole *et al.* (2003) ont constaté le contraire pour les 51 guides sur le cancer examinés.

3.2.4 Longueur du guide

Dans l'échantillon de 855 guides collectés par Hibble *et al.* (1998), 28 % étaient longs d'une page, 23 % de 2 pages et 19 % en avaient plus de 10. Dans l'étude de Burgers *et al.* (2003) la longueur moyenne était de 15 pages et plus. Les guides plus longs respectent mieux les standards préconisés dans les GsG et les instruments d'évaluation. En revanche, ils seraient moins utilisés (Cook et Giacomini 1999). En SST, la longueur des guides est plus variable et étendue (cf. section 1.4.1).

Pour conclure, cette brève revue des facteurs confirme la conclusion de Burgers *et al.* (2003) : il y a peu de recherches – et de résultats – sur les caractéristiques des guides qui permettent de prédire leur qualité.

3.3 Opinions sur les guides en SST

La littérature sur le sujet est pauvre (tout comme celle en santé, proportionnellement). Peu de recherches ont été menées sur les guides dans leur ensemble. Dans un domaine en particulier, il existe par ailleurs peu de guides couvrant le même sujet. Les souhaits exposés ci-haut correspondent à ce qui est souvent considéré comme important en SST : applicabilité, importance du processus de décision, variabilité et complexité des facteurs à prendre en compte, point de vue holistique, nécessaire latitude. L'introduction de guides ne soulève pas, à notre connaissance, les craintes explicitées ci-dessus. Une explication possible est que les guides ont rarement un ton très directif.

4 UTILISATION ET IMPACT DES GUIDES SUR LA PRATIQUE

Les auteurs ont cherché à répondre à plusieurs questions : Connaît-on les guides? Y adhère-t-on? Sont-ils utilisés? Modifient-ils la pratique? Ont-ils un impact sur la santé du patient⁹⁷? Quels sont les obstacles?

Les études sur l'utilisation, par rapport à celles que nous verrons au chapitre suivant sur l'implantation, explorent surtout le point de vue de l'utilisateur plutôt que son comportement. Dans ce chapitre, nous faisons d'abord le point sur les sources d'information privilégiées par les usagers, puis nous présentons un portrait sur l'adhésion et l'adoption des GdP, pour terminer avec ce qui fait obstacle ou au contraire facilite l'utilisation ou l'adhésion.

4.1 Sources d'information privilégiées par les usagers

Les cliniciens ne disposeraient que d'une heure par semaine pour lire la documentation scientifique (Sackett -1996-, cité par Culleton 2009). Une des questions explorées par les études a donc été celle des sources d'information utilisées, entre autres pour situer la place des guides et le crédit qu'on leur accorde. Ainsi, Hayward *et al.* (1997) ont mené une enquête par questionnaire sur la fréquence (rare, mensuel, quotidien/hebdomadaire) avec laquelle diverses sources d'information sont consultées et l'impact sur la pratique. Les quatre sources les plus utilisées sont les collègues, les articles de synthèse, les *textbooks* et les articles originaux ; 35 % disent ne consulter que rarement les GdP, sans toutefois mettre en cause leur crédibilité. Le problème serait plutôt leur adéquation avec la réalité. Les *textbooks*, ici deux fois et demie plus souvent utilisés que les GdP sont décrits par Cook *et al.* (1997) comme des revues narratives qui offrent une perspective globale (physiologie, pathologie...) sur un ensemble de sujets plutôt que de couvrir en profondeur un sujet pointu. Ils ne fournissent pas de réponse quantitative sur une question clinique spécifique.

Tableau 4.1 - Sources d'information utilisées par les cliniciens (Hayward et al. 1997)

(n=1878)	rare	mensuel	quotidien/ hebdo.	impact important
Collègues, consultants	3	16	81	83
Article de synthèse	9	51	40	59
Livre de référence (textbook)	18	42	40	57
Article de recherche original	31	41	28	30
Guide de pratique clinique	35	48	17	44
Éducation continue	48	38	14	69
« Position paper »	61	33	6	23

Les études menées en Australie et au Danemark montrent aussi que la source première d'information est les collègues et les ouvrages de référence (Ulvenes *et al.* 2009). C'est d'ailleurs une des raisons pour lesquelles les stratégies d'implantation sont souvent centrées sur l'identification des personnes clés crédibles⁹⁸. L'enquête menée auprès de 966 médecins montre que la première source utilisée pour prendre une décision est un collègue (86 %), d'autres spécialistes de confiance (78 %), un ouvrage de référence (76 %). Le site du Medline/Pubmed est consulté par 56 %, un journal médical par 52 % et Cochrane, par 27 %. La moitié (54 %) considère qu'il est difficile de trouver le temps pour chercher des informations.

⁹⁷ Cette question n'est pas abordée dans la présente revue, mais les études consultées montrent que les résultats sont décevants. Ex. étude de Worral *et al.* (1997) sur 13 essais de guides : seulement 5 ont donné ds résultats statistiquement significatifs.

⁹⁸ Cela renvoie en anglais à la notion de 'face validity' (Browman 2001).

4.2 Portrait de l'adhésion et de l'adoption des recommandations

Les études portent en général sur des directives ou des recommandations incluses dans les GdP.

Adhésion. Les utilisateurs ont globalement une opinion positive sur les guides, mais ils expriment des réserves, dans des proportions qui varient selon le domaine et le type de guides (Carlsen *et al.* 2007, Lugtenberg *et al.* 2009). Par exemple, dans l'étude de Burgers *et al.* (2003) menée par audit, sur 96 recommandations, seulement une sur trois a obtenu un taux d'adhésion supérieur à 60 %, alors que dans l'étude menée auprès des praticiens hollandais, le taux moyen variait de 60 à 70 % (Llentjens 2008). Dans la revue effectuée par Cabana *et al.* (1999)⁹⁹, l'adhésion pouvait être quasi parfaite (99 % : test pour le cancer du sein) comme le désaccord quasi général (91 % : utilisation d'un médicament antiviral en pédiatrie). Dans un certain nombre de cas, on ne croit pas à l'efficacité de ce qui est proposé¹⁰⁰. Les raisons de s'opposer sont cependant multiples : simplifie la réalité, pas applicable à la population soignée, réduit l'autonomie, diminue la flexibilité, rend la relation patient-praticien impersonnelle, pas pratique (Carlsen *et al.* 2007)¹⁰¹.

Adoption. Globalement, les praticiens suivent les recommandations à des degrés divers, dépendant des régions, des pays et des domaines de pratique (Carlsen et Kjellberg 2010). Par exemple, un tiers des praticiens hollandais prescrivent les médicaments tel que recommandé par les guides nationaux (Lugtenberg *et al.* 2009) et 55 % des patients américains sont traités tel qu'il est recommandé (Divac *et al.* 2009). La raison principale n'en serait cependant ni la méconnaissance ni la non-adhésion (Rosser et Palmer 1993¹⁰², Dahan *et al.* 2008¹⁰³).

4.3 Barrières à l'utilisation et incitatifs

Nous abordons les barrières à l'utilisation et les incitatifs de façon simultanée, l'un étant en partie le miroir de l'autre. Les auteurs ciblent tantôt spécifiquement les barrières¹⁰⁴, tantôt ce qui facilite l'adhésion et l'adoption ou, de façon globale, ce qui influence l'adhésion-adoption. Les études s'intéressent parfois au GdP dans son ensemble, parfois aux directives spécifiques ou à des barrières spécifiques. Nous présentons dans l'encadré 4.1 trois propositions de découpage. La première est centrée sur les obstacles à l'adoption liés aux attitudes et aux expériences des praticiens ; la seconde couvre des variables qui influencent l'adoption. La troisième proposition, qui porte aussi sur les obstacles, est la plus élaborée. Il s'agit d'une méta-analyse dont nous en présentons les résultats ci-dessous en y intercalant les résultats d'autres études.

⁹⁹ 17 études.

¹⁰⁰ Recommandations dont on croit le moins à l'efficacité : prévention sur les abus d'alcool 90 % vs 10 % pour celles sur l'examen clinique des seins.

¹⁰¹ Méta-analyse de 12 études qualitatives publiées entre 1998 et 2006.

¹⁰² 78 % des praticiens ontariens connaissent les recommandations sur l'abaissement des lipides, 5 % l'appliquent.

¹⁰³ GdP sur les problèmes lombaires : 90 % des 132 répondants connaissaient le guide, 24 % en faisaient usage .

¹⁰⁴ Les barrières sont définies comme tout facteur qui limite ou restreint l'adhésion (Cabana *et al.* 1999).

Encadré 4.1 - Les obstacles ou les facilitateurs de l'adhésion-adoption

Barrières à l'adoption (Carlson <i>et al.</i> 2007)	Variabes qui influencent l'adoption (Grilli et Lomas 1994)	Barrières à l'adhésion (Cabana <i>et al.</i> 1999)
Adhésion	Qualité des guides (recommandations, changement de pratique)	Cognitives
Contenu	Caractéristiques des professionnels ciblés	Être au courant
Format	Caractéristiques de l'environnement	Être familier : Le praticien s'estime-t-il assez familier pour... ? Ses connaissances sont-elles adéquates?
Expérience individuelle des GdP	Incitatifs (ex. légaux, financiers)	Liées aux attitudes
Relations patient-praticien	Facteurs liés aux patients	Désaccord avec les recommandations?
Responsabilité professionnelle	Réglementation (organismes qui peuvent émettre des réglementations)	Coût jugé trop élevé
	Adhésion aux recommandations comme facteur d'accréditation	Ratio temps énergie/bénéfice insatisfaisant
		Liées au comportement
		Recommandations malcommodes, confondantes
		Patient
		Environnement

4.3.1 Connaissances nécessaires à l'utilisation des guides

La grande question formulée est celle de savoir si les recommandations sont ou ne sont pas utilisées parce qu'on ne les connaît pas ou parce que le praticien n'a pas les connaissances suffisantes pour les mettre en pratique (Cabana *et al.* 1999).

Être au courant. Les 46 études révisées montrent que le pourcentage de praticiens qui connaissent l'existence des recommandations dépend fortement du domaine. Il varie de 84 % (US Preventive Services Task Force guidelines) à 1 % (asthme, immunisation contre la rougeole).

Être familier. Deux aspects ont été évalués (31 études) : 1) Le praticien s'estime-t-il assez familier pour l'utiliser? 2) En a-t-il une connaissance adéquate? Le pourcentage de praticiens qui estiment que le manque de familiarité est un obstacle varie de 89 % (*American College of Physicians exercise stress testing*) à 0 % (asthme). La familiarité, explorée en matière d'inertie liée aux pratiques antérieures, y compris celles du patient, ressort aussi comme un facteur majeur¹⁰⁵.

Quand la question est creusée plus finement, les résultats sont parfois surprenants. Par exemple, Divac *et al.* (2009), qui se sont demandé pourquoi les pratiques médicales contemporaines en tant qu'alternatives à la prescription de médicaments avaient autant de difficulté à s'implanter auprès des praticiens serbes, ont constaté que la majorité (65 %) trouvait important de suivre les guides¹⁰⁶. Cependant, ils ne connaissaient pas bien les guides (50 % ne pouvaient en nommer aucun), tout en souscrivant aux mêmes principes, sans toutefois vraiment les intégrer dans la pratique. Ces écarts s'expliqueraient par le fait que le praticien intègre des expériences personnelles dans son processus de décision qui ne sont pas couvertes dans les recommandations.

4.3.2 Clarté et spécificité

Une des études les plus élaborées à propos de la clarté et de la spécificité est celle de Shekelle *et al.* (2000) qui a étudié l'impact du format des recommandations. Ces dernières portaient sur la

¹⁰⁵ Familiarité (14 études) : médiane : 42 %; taux le plus élevé (66 %) : la position de sommeil du bébé; le plus bas : la vaccination (23 %).

¹⁰⁶ Les quatre facteurs dominants ont été : sécurité (78 %), efficacité (73 %), recommandé par G (65 %), remboursement (46 %).

pertinence ou non d'utiliser un test¹⁰⁷. Essentiellement, douze situations/conditions de patients possibles étaient illustrées sur des vignettes (groupe témoin). Elles pouvaient être accompagnées de recommandations générales (groupe A) ou spécifiques (groupe B). Les trois groupes ont ordonné les vignettes de base de la même façon (le test est approprié ou pas) : ils adhéraient donc aux recommandations. Le suivi des décisions prises par la suite a montré que l'ajout de recommandations, générales ou spécifiques¹⁰⁸, avait peu d'impact. D'après les auteurs, les recommandations générales n'apportaient rien et les recommandations spécifiques ne l'étaient pas suffisamment pour permettre au praticien de voir à quel patient s'appliquaient les recommandations. Le problème ne se situait donc pas au niveau de la validité mais de l'adéquation avec un contexte donné.

4.3.3 Simplicité et convivialité

La facilité d'utilisation du guide (*userfriendliness*) est un facteur majeur d'adoption, les praticiens privilégiant les formats courts (Hayward *et al.* 1997). Comme le résumait Carlsen et Kjellberg (2010), les praticiens utilisent ce qui facilite le travail, pas ce qui complique la pratique. La simplicité des recommandations est aussi le premier facteur expliquant son adoption d'après le bilan de l'étude de Grilli et Lomas (1994) de 23 essais sur l'efficacité de la dissémination. Cabana *et al.* (1999) le retiennent aussi comme facteur prépondérant (23 études).

4.3.4 Observabilité

Le fait qu'une recommandation ne comporte pas de bénéfices visibles est vu comme un obstacle important¹⁰⁹ (19 études, médiane : 13 %). Cela fait aussi partie des facteurs explicatifs importants relevés par la revue de Grilli et Lomas (1994), à savoir : pouvoir observer les effets et voir les recommandations utilisées par d'autres praticiens.

4.3.5 Accessibilité et environnement

L'environnement a été le facteur le plus sondé (45 études, dont deux sur l'environnement organisationnel). Dans chaque étude, au moins 10 % des répondants le citent en tant qu'obstacle. Les problèmes d'accessibilité (ex. ordinateur, Hayward *et al.* 1997) ou la présence d'un environnement adverse est souvent notée (Carlsen et Kjellberg 2010).

4.3.6 Crédibilité des développeurs ou des utilisateurs

Un guide endossé et utilisé par un collègue respecté a un impact important, tout comme l'organisation qui l'endosse¹¹⁰. En revanche, ce qui est crédible varie selon les régions. De façon générale, les praticiens ont plus confiance dans les sources locales (Hayward *et al.* 1997), car les auteurs auraient une meilleure compréhension des contextes de pratique. Les guides locaux sont mieux acceptés et ont plus d'impact sur la pratique professionnelle. Dans une étude menée sur les sources utilisées à la Commission des lésions professionnelles par les intervenants (Tcaciuc *et*

¹⁰⁷ Test d'électrodiagnostic ETD (EMG combiné à un test de vélocité) s'appliquant à des patients souffrant de douleurs aiguës ou chroniques au dos. Les 12 vignettes (préparées par un groupe d'experts) décrivaient en mots et en images des patients pour lesquels on pouvait considérer pertinent ou non d'utiliser ce test.

¹⁰⁸ Le groupe avec recommandations spécifiques a commandé + 8 % d'ETD lorsque approprié et -5 % quand inapproprié.

¹⁰⁹ Médiane : 13 %. % le plus élevé, 65 % : guide en éducation en nutrition; le plus bas, 1 % : conseils sur les exercices.

¹¹⁰ Dans l'enquête menée auprès de praticiens canadiens, 78 % considéraient que l'endossement par un collègue estimé avait un impact très important, plus que l'endossement par une organisation (62 %) (Hayward *et al.* 1997).

al. 2011), les guides locaux étaient aussi la source la plus utilisée. L'étude menée par Carlsen et Kjellberg (2010) en Norvège et au Danemark montre que malgré des réserves exprimées sur les preuves à la base de certaines recommandations, le consensus en est un de confiance envers les organisations qui les développent. À cet égard, Fretheim *et al.* (2006) recommandent à l'OMS d'appuyer les développements locaux, option bien différente de l'adaptation souvent préconisée.

4.3.7 Facteur culturel

Il y a consensus quant à l'importance de tenir compte du facteur culturel dans la conception des guides¹¹¹, mais aussi en ce qui a trait à leur adhésion et à leur adoption. Par exemple, dans l'étude de Kryworuchko *et al.* (2009) menée auprès des praticiens canadiens, ceux de langue française valorisaient davantage les *textbooks* et avaient plus confiance dans les guides que leurs collègues anglophones. Les praticiens norvégiens de Carlsen et Kjellberg (2010) sont apparus plus réservés que leurs collègues danois, car ils considéraient que les GdP pouvaient camoufler des enjeux de rationalisation, primant sur les décisions d'ordre clinique ; ils se voyaient en partie comme des alliés de leur patient face aux instances gouvernementales. Les praticiens danois appréciaient au contraire les efforts de rationalisation et les efforts d'implantation du gouvernement¹¹².

4.3.8 Portrait global

Présenter les résultats facteur par facteur présente l'inconvénient de perdre de vue le portrait global. Les résultats globaux de deux études sont donc résumés à l'encadré 4.2. La seconde étude a été retenue, car elle s'est adressée aux organisations en tant qu'utilisatrices, ce qui est rare.

Encadré 4.2 - Obstacles et incitatifs à l'utilisation des guides

Raisons invoquées par 30 praticiens pour ne pas adhérer ou utiliser des recommandations issues de 11 guides nationaux
(Lugtenberg *et al.* 2009)

Contenu lui-même : 68 % des recommandations clés. Surtout à cause de leur applicabilité (57 %), mais aussi de l'insuffisance des évidences ; on ne croit pas que l'effort vaut la peine (23 %) ou que cela va améliorer sensiblement les résultats.

Présence de facteurs **d'environnement** adverses comme les **contraintes organisationnelles** : 52 %.

Connaissance insuffisante des recommandations (46 %) ou un **manque de motivation** ; le facteur patient (40 %), comme le fait qu'il ait d'autres préférences et leurs limites.

Formulation (32 %) comme le manque de clarté ou l'ambiguïté et la complexité.

Conclusion des auteurs : les études ne prennent pas suffisamment en compte les barrières individuelles par rapport aux barrières opérationnelles.

Enquête menée par questionnaire électronique auprès de 28 établissements¹¹³ **sur un GdP biomédical diffusé en 2002**
(Farges 2004)

Utile : 57 %

Exploitable : 36 %

Compréhensible : 14 %

Mis en œuvre : 29 %

Prévoit mettre en œuvre certains items : 25 %

Finalement, comme les études en transfert de connaissances ont exploré ces questions de façon plus large, nous joignons en annexe (4.1) la synthèse qu'en font Faye *et al.* (2007).

¹¹¹ Manna *et al.* (2003) ont comparé comment les guides issus de divers pays abordaient la question de l'ethnie des patients.

¹¹² L'interprétation proposée : les praticiens danois auraient intégré la politique de rationalisation des soins du gouvernement.

¹¹³ Il y a eu 315 envois d'un questionnaire électronique. Cependant, 96 % des répondants connaissaient le guide.

4.4 Impact des attributs des recommandations

Les attributs sont les caractéristiques qualifiant les recommandations (cf. encadré 4.3). Les auteurs ont vérifié leur influence sur l'adhésion à des recommandations ou le désir de s'y conformer. Ils rejoignent en grande partie les facteurs dont on vient de faire le tour, mais ces études portent spécifiquement sur les recommandations elles-mêmes. Une des plus intéressantes sur ce sujet est celle de Grol *et al.* (1998) qui ont analysé l'impact de 12 attributs sur l'utilisation de 47 recommandations issues de 10 GdP chez 61 praticiens qui ont eu à prendre 12 880 décisions. Les recommandations étaient suivies 61 % des fois. Les attributs ayant eu le plus d'impact ont été : le fait d'être controversé, l'incompatibilité avec les valeurs, la clarté, la complexité, les impacts sur la pratique quotidienne (ex. demande de changer de routine, provoque des réactions négatives des patients). Dans l'étude de Burgers *et al.* (2003)¹¹⁴, l'attribut dominant a été le fait qu'une recommandation soit fondée sur des preuves scientifiques ; cependant pour les guides de diagnostic, c'est le fait de pouvoir rapidement en voir les effets qui était ensuite le plus déterminant, alors que pour le guide de traitement, cela a été la discussion sur les bénéfices et le fait d'évoquer les réactions potentiellement négatives des patients. La barrière la plus importante était liée à la nécessité de développer de nouvelles habiletés et à la complexité.

Encadré 4.3 - Attributs des recommandations

<p>Influence leur adoption (Grilli et Lomas 1994)</p> <p>Observabilité : exemple de quelqu'un d'autre les appliquant Avantages : c.-à-d. supérieure aux anciennes pratiques, et ce, de façon démontrable Compatibilité : avec les valeurs ou croyances, par rapport aux expériences antérieures Complexité : en matière de difficulté à comprendre ou à intégrer dans sa pratique Testabilité : <i>ex.</i> peut-on essayer assez facilement le GdP ou au moins une partie?</p>
<p>Facilite ou nuit à l'adhésion (Burgers <i>et al.</i> 2003)</p> <p>Facilite si la recommandation est</p> <ul style="list-style-type: none"> Soutenue par des évidences scientifiques Appuyée par une discussion sur les bénéfices Accompagnée d'une discussion sur les risques et dommages possibles (l'effet négatif) Facile à suivre L'effet peut être vu rapidement Compatible avec les normes et valeurs actuelles <p>Obstacle si la recommandation</p> <ul style="list-style-type: none"> Fait partie d'un arbre complexe de décision et Demande, pour être appliquée, de nouvelles connaissances Des habiletés nouvelles De changer l'organisation De changer les routines et les habitudes Peut entraîner des réactions négatives de la part des patients

4.5 Utilisation des guides en SST

À notre connaissance, il n'existe pas d'études portant sur l'utilisation des guides en général en SST. Cependant, il existe des études ciblant des guides spécifiques. À titre d'exemple, le guide sur les convoyeurs a fait l'objet d'un suivi auprès des utilisateurs (Trépanier *et al.* 2011). La plupart des facteurs mentionnés ci-dessus sont pertinents à la SST.

¹¹⁴ Combine les attributs de Grilli et Lomas (1994) et de Rogers (1995). Des 16 attributs initiaux, 12 ont été conservés : six facilitateurs et six obstacles.

5 IMPLANTATION DES GUIDES

Depuis 20 ans, l'ensemble des études montre que la dissémination du matériel a peu d'effets sur les pratiques professionnelles (Hayward *et al.* 1997, Bero *et al.* 1998, Lomas *et al.* 1989, Cabana *et al.* 1999) ou du moins des effets mitigés (Davis et Taylor-Vaisey 1997, Grimshaw et Russell 1993¹¹⁵). Les conclusions du Nuffield Institute for Health et du Center for Economics Health (1994) sont plus positives (43 des 44 études les plus fiables montrent des impacts significatifs), mais les conclusions sont posées du point de vue de l'amélioration des pratiques plutôt que sur celui de l'adhésion ou de l'effet sur le patient.

Un des reproches adressés aux guides a donc été la faiblesse des efforts et des ressources dédiées à leur implantation (Browman 2000, Engelbrecht et Courte-Wienecke 1999), faiblesse qu'ont voulu corriger les GsG. La littérature qui porte sur le sujet est par ailleurs de plus en plus abondante. Deux grandes questions la traversent : qu'est-ce qui empêche les praticiens d'adhérer (cf. chapitre 4) et quelles interventions sont les plus efficaces pour favoriser l'utilisation des GdP¹¹⁶? Dans ce chapitre, après une mise à plat des termes clés, nous brossons le portrait des interventions possibles, leur fondement, leur logique et leur efficacité potentielle. Nous terminons avec une section sur les outils d'évaluation de l'implantation (en tant qu'objet d'étude), outils en général surtout centrés sur l'évaluation des obstacles. Nous présentons une proposition qui organise l'ensemble des déterminants de l'implantation.

5.1 Définitions et concepts

Diffusion, dissémination et implantation sont des moyens de transfert. L'adoption est un résultat. On retrouve généralement la notion d'adhérence ou adhésion dans le sens de 'être d'accord'. Nous reprenons ici les définitions de Cabana *et al.* (1999)¹¹⁷ :

- **adoption** : engagement du destinataire à utiliser le GdP et décision de modifier sa pratique ;
- **diffusion** : distribution des informations et adoption sans aide des recommandations;
- **dissémination** : processus plus actif que la diffusion. Les informations pour améliorer les connaissances ou les habiletés sont communiquées ;
- **implantation** : dissémination active, impliquant des stratégies pour surmonter des barrières.

Comme ces termes sont très utilisés en transfert de connaissances (TC) – implantation et adhésion le sont moins –, les principaux termes définis dans le guide sur le TC produit pour le RRSSTQ sont repris à l'annexe 5.1 (Faye *et al.* 2007).

¹¹⁵ Les auteurs ont révisé 59 études publiées : 55 rapportent des améliorations statistiquement significatives quant au processus de soins; 9 études sur 11 rapportent un impact positif en ce qui a trait aux effets (*outcomes*). Cependant, les améliorations sont souvent faibles et limitées sauf, par exemple, pour un seul élément.

¹¹⁶ La littérature sur l'implantation rejoint en partie celle sur l'utilisation puisqu'elle se vérifie en partie à travers l'utilisation.

¹¹⁷ Basées sur le travail de Davis et Taylor-Vaisey (1997). Lomas (1993) fait aussi une intéressante revue de la question. Parmi les auteurs aussi souvent cités pour la conception des études d'implantation, on retrouve l'ouvrage de Mäkelä et Thorsen (1999).

5.2 Interventions possibles

Globalement, on distingue trois modes de diffusion : 1) passif (ex. envoi de pamphlets) ; 2) éducation ; 3) actif. Le mode passif est généralement celui qui est le plus utilisé (Kryworuchko *et al.* 2009). Par exemple, pour 84 % des guides de la cohorte 2000-2005 un moyen de diffusion passif avait été utilisé, pour 58 %, un moyen lié à l'éducation, et pour 30 %, une méthode d'implantation¹¹⁸.

Grol (1997) différencie les implantations selon qu'elles mettent en jeu un processus interne, comme l'éducation, ou externe, comme l'organisation. Il résume dans un excellent texte synthèse les différentes approches possibles, leur théorie sous-jacente, sur quel facteur porte l'intervention et la stratégie qui en découle (annexe 5.2)

D'après l'étude menée par Grimshaw *et al.* (2004), les deux interventions uniques les plus fréquentes sont les rappels (13 %) et la diffusion de matériel éducatif (6 %). Les deux plus rares¹¹⁹ sont les interventions sur des éléments organisationnels. Dans le cas des multiinterventions, les moyens utilisés sont : 1) du matériel éducatif : 48 % ; 2) des rencontres éducatives : 41 % ; 3) des rappels : 24 % ; 4) des audits et des rétroactions : 31 % ; 5) des interventions auprès des patients : 18 %.

5.3 Pratiques évaluatives

Les pratiques évaluatives sont l'objet d'un vaste domaine d'investigation¹²⁰. Il ne saurait être question d'en faire un bilan. Nous trouvons cependant intéressant de résumer à l'encadré 5.1 quelques points de la vaste enquête menée par Christie (2003).

Encadré 5.1 - Étude de Christie (2003) sur les pratiques d'évaluation : points saillants

Perspectives

- académiques : visée quantitative
- partenariale (*stakeholder service*) : questions évaluatives basées sur les besoins des partenaires
- décisionnelle : formule les questions en fonction des décisions à venir et de la législation
- résultat : efficacité? Est ce que ça vaut la peine?

De la théorie à la pratique : A questionnaire 8 théoriciens et 138 évaluateurs sur leur pratique

- Très peu (10 %) se réfèrent à un cadre théorique particulier
- 8 sur 138 évaluateurs ont pu citer un livre guidant leur pratique

Rôle des partenaires : Veut-on l'impliquer? Portée de son implication? Sur quoi?

- Modèle décisionnel : il participe au processus mais pas à la sélection des méthodes, ni à l'interprétation
- Modèle de perspective sociale : il est inclus dans la délibération et le dialogue ; il est vu comme un moyen de diffusion, de sensibilisation.

Farges (2007) suggère cependant une approche assez différente qui pourrait être fort appropriée à la culture en SST, soit l'utilisation d'un diagnostic collectif. Cela augmenterait la capacité de consensus collectif sur les actions d'amélioration et les échanges entre départements. En transfert de connaissances, l'utilisation est vue aussi sous l'angle d'une dynamique d'échanges qui permet aussi d'évaluer les problèmes d'applicabilité¹²¹.

¹¹⁸ Les résultats sont un peu meilleurs pour la cohorte des guides publiés entre 1994 et 1999. Il s'agit d'un effet temps (plus de temps pour diffuser).

¹¹⁹ Deux sur les 235 études menées.

¹²⁰ Le lecteur intéressé pourra consulter l'ouvrage de Alkin (2004) *Evaluation roots – Tracing theorists' views & influence* et pour une lecture plus synthétique : *Evaluation Theory and practice : Insights and new directions* (2003).

¹²¹ Un exemple : l'utilisation des rétroactions dans le cadre d'une formation à l'aiguisage des couteaux (Desmarais *et al.* 2011).

5.4 Efficacité et coût des interventions

Globalement, à peu près tous (National Health and Medical Research Council 1995, Oxman *et al.* 1995, Bero *et al.* 1998, Grol et Grimshaw 1999, Grimshaw *et al.* 2004, Farmer *et al.* 2009) s'entendent sur les points suivants : 1) la dissémination passive (ex. envois postaux, électroniques, pamphlets, *newsletters*, revues) est peu efficace pour modifier les pratiques ; 2) les interventions gagnent en efficacité quand elles font participer au processus les personnes ciblées ; 3) l'implantation qui combine des moyens est plus efficace.

Une des premières revues exhaustives sur les études d'implantation a été celle de Bero *et al.* (1998)¹²² sur la consistance des résultats, suivie quelques années plus tard par la revue par Grimshaw *et al.* (2004) de 235¹²³ études rapportant 309 comparaisons. Le but était de situer l'importance de l'impact (faible : moins de 5 % ; modeste : 5 à 10 % ; modéré : 10 à 20 % ; important : plus de 20 %). Nous combinons ci-dessous les deux types de résultats. Les conclusions de l'étude de Grimshaw sont indiquées entre parenthèses.

Les interventions jugées¹²⁴ **effectives** sur une **base consistante** sont les : 1) visites à visée éducative ; 2) rappels écrits ou informatiques (modéré) ; 3) interventions multiples qui combinent des moyens tels audits + rétroactions, rappels... (modeste, le nombre n'améliore pas l'impact, mais certaines combinaisons pourraient être plus efficaces; les écarts sont toutefois faibles d'une combinaison à l'autre). De façon générale, le couplage avec le rappel augmenterait l'impact; 4) rencontres éducatives interactives.

Les interventions suivantes ont des **succès variés** : 1) audits, rétroactions (modeste) ; 2) utilisation des leaders d'opinion ou les praticiens reconnus par leurs collègues comme étant influents. La stratégie est recommandée par le NHMRC ; 3) processus d'élaboration d'un consensus local (les praticiens participants sont inclus dans un processus de discussion pour s'assurer qu'ils adhèrent au G) ; 4) interventions auprès des patients (modéré).

Les interventions suivantes ont **peu ou pas de succès** : 1) matériel éducatif distribué (inclut les guides), surtout si non sollicité (modeste) ; 2) rencontres à caractère éducatif comme des conférences (peu d'effet).

Globalement, les impacts sont donc généralement modestes, au mieux modérés¹²⁵. Aucune méthode n'a d'impact important et on ne sait pas très bien ce qui peut être efficace dans un contexte donné (Grol et Grimshaw 1999, Fretheim *et al.* 2006¹²⁶). Le NHMRC qui a fait aussi une revue assez complète des moyens d'implantation trace un portrait semblable dans ses grandes lignes. Il ajoute cependant une section sur les incitatifs : bonus financiers, accréditations, points de formation continue, invitation à des conférences, protection accrue dans les cas de

¹²² Synthèse de 18 études à partir d'une sélection de 1 139 références; les conclusions sont pratiquement les mêmes que dans la revue de Davis et Taylor-Vaisey (1997).

¹²³ Répertoire de 150 000 références à partir des mots clés choisis; en retiennent 5 000, puis 863 textes, pour en conserver finalement 285 (en partie parce que les informations transmises ne permettaient pas aux auteurs de juger la qualité) décrivant 235 études. Provenaient de 14 pays : 71 % États-Unis; 11 % Royaume-Uni; 6 % Canada; 3 % Australie pour un total de 91 %. Les résultats portent sur ce qui marche ou pas, en pays anglophone.

¹²⁴ On retrouve aussi en fait en parallèle des propositions d'implantation qui s'apparentent plus aux approches en amélioration continue (ex. Gates 1995) et qui ne font pas partie des grandes études évaluatives.

¹²⁵ Grol et Grimshaw (1999) concluent aussi à des impacts modérés (à partir de 51 revues répertoriées sur la question).

¹²⁶ Les auteurs ont aussi effectué une revue de la question à la demande de l'Organisation mondiale de la santé.

litige, réglementation, etc. Comme on l'a vu au chapitre 4, plusieurs de ces moyens touchent des cordes sensibles et suscitent des critiques. Au final, comme le soulignaient Bero *et al.* il y a une quinzaine d'années (1998), on en sait encore peu sur les stratégies plus efficaces et encore moins sur les coûts associés¹²⁷. Cependant, comme ces études sont très coûteuses, la proposition de Marriott *et al.* (2000) de formuler autrement les questions nous apparaît ici pertinent. Ils proposent de répondre aux cinq questions suivantes : 1) qui a besoin de savoir? (audience) ; 2) qu'ont-ils besoin de savoir? (contenu) ; 3) de qui ont-ils besoin de l'entendre? (source) ; 4) quelle est la meilleure façon de les atteindre? (véhicule) ; 5) qu'est-ce qui influence leurs attitudes face au message reçu? (contexte, environnement).

5.5 Évaluation du potentiel d'implantation

On sait qu'au total, c'est une combinaison de facteurs qui font qu'une implantation fonctionnera ou pas, et que parmi les éléments clés on retrouve : on y croit ou non, services/ressources disponibles ou non, incitatifs, engagement envers le guide (*advocacy*) (Reddy 1999). Diverses propositions ont émergé ces dernières années pour évaluer le potentiel d'implantation intrinsèque des recommandations (*implementability*). Elles sont à la frontière des outils d'évaluation abordés au chapitre 2, car elles visent aussi à déceler les barrières reliées à des facteurs intrinsèques (ex. recommandation ambiguë, inconsistante, incomplète) vs extrinsèques. Parmi les propositions les plus récentes, nous avons retenu et annexé (5.3) le *GuideLine Implementability Appraisal*¹²⁸ (GLIA) développé par Shiffman *et al.* (2004). Chaque recommandation ou énoncé est évalué spécifiquement.

En santé et en sécurité du travail, il n'y a pas de grandes études comme en santé. Mais il existe clairement des guides qui sont largement utilisés et dont l'implantation a été un succès. Un exemple est le *Quick Exposure Check-list* introduit par la CSST. Le fait d'être utilisé par les inspecteurs a été en soi un incitatif : il est régulièrement utilisé dans les entreprises. Les succès d'implantation sont jugés en partie au nombre d'exemplaires commandés. Par ailleurs, le processus d'implantation s'imbrique souvent dans celui du développement, non pas en tant qu'étape planifiée spécifiquement, mais de façon entremêlée avec le développement. La conception en soi est réfléchie avec les partenaires en termes de potentiel d'implantation. Bien que nous en sachions très peu sur l'implantation des guides en SST, de façon plus générale, les chercheurs en SST au Canada rapportent utiliser la dissémination active comme principal mécanisme de transfert de leurs résultats de recherche (Laroche 2009). Alors que la diffusion vient au deuxième rang, il semble que l'implantation des connaissances soit le mécanisme le moins souvent rapporté par les chercheurs en SST.

¹²⁷ Revue des études par Graham *et al.* (2004) : 29 % rapportaient des analyses de coûts, mais la qualité des données était pauvre.

¹²⁸ Il avait été au départ développé pour aider à raffiner les recommandations, mais il a été utilisé par la suite pour évaluer les obstacles à l'implantation. D'après les auteurs, cela prend une vingtaine de minutes par énoncé. La version annexée est la proposition électronique de Hill et Lalor (2009).

6 BILAN COMPARATIF SANTÉ VS SST ET PISTES À EXPLORER

La question de départ formulée était : sait-on ce qu'est un bon guide et comment l'évaluer? Ceci nous a menés à une littérature qui accorde une importance centrale aux procédures utilisées pour développer les guides. Ce chapitre fait le point sur ce que la revue nous a appris à ce sujet. Nous avons retenu 27 points qui sont des questions ou des objets traités par les auteurs sur les guides. Pour chacun, nous résumons brièvement le point de vue exprimé en santé, puis le situons par rapport au domaine de la SST. Comme on le verra, si les solutions adoptées en santé apparaissent souvent problématiques en SST, les questionnements et les réflexions menées sont cependant fort pertinentes pour le développement des guides en SST. Nous tentons de dégager, au fil de ce bilan comparatif, des pistes de réflexions et de recherche à explorer. En fait, cette revue nous a amenés à constater la nécessité de mieux articuler ou préciser nos cadres théoriques ou nos approches. Le chapitre suivant proposera d'ailleurs un bilan de cette réflexion.

Données et connaissances

- 1. La gestion des connaissances et le choix des données.** La croissance explosive des données publiées en santé pose un sérieux problème de gestion en termes de triage, d'évaluation et de synthèse. La réponse à ce problème a été le développement de procédures qui se veulent les plus systématiques et objectives possibles. Ces procédures, relativement rigides et coûteuses, sont essentiellement linéaires (choix des questions, des mots clés, des critères d'inclusion, d'exclusion, etc.). Les études menées auprès de développeurs de guides montrent qu'une minorité utilise ces procédures. La SST n'a pas ce problème. La production y apparaît minuscule en comparaison. Documenter systématiquement tout ce qui est fait devient rapidement lourd et nuit au processus créatif de recherche. Les informations présentées dans les discussions ou les introductions, la logique, le raisonnement élaborés deviennent négligés. Le processus ne permet pas les allers-retours, les échappées pour poursuivre une piste. Cependant, nous gagnerions à mieux préciser les procédures ou du moins les approches utilisées en SST.
- 2. Source des données.** En santé, les données considérées valables sont essentiellement celles issues de la recherche. En SST, l'utilisateur est souvent associé au processus de production de connaissances ou du moins à la définition des besoins. Les connaissances non académiques sont mieux reconnues.
- 3. Évaluation des données.** En santé, les données sont évaluées selon des critères précis qui conviennent mal à la SST, vu la variabilité des domaines considérés. Il serait intéressant de développer une approche évaluative transversale.
- 4. Connaissances et situations.** En santé, le contexte est surtout abordé en tant que facteur lié à l'utilisation des guides de pratique. En SST, les situations et les contextes, qui évoluent à grande vitesse, interpellent directement les connaissances à produire. Il y aurait lieu de développer un cadre théorique permettant de situer plus clairement la place du contexte dans la production des connaissances. Cette question est reprise au chapitre suivant.

- 5. Données probantes.** La notion est centrale en santé. La définition ne convient pas forcément bien à la SST. Ce point est discuté plus amplement au chapitre suivant. Il y aurait avantage à mieux définir ce que seraient des données probantes en SST.

Approches et partenariats

- 6. Approche prescriptive.** Malgré l'ampleur des ressources allouées, l'ensemble des études montre que les prescriptions développées sont difficiles à implanter, même si les gens y sont favorables. Sans intégration ni appropriation en lien avec la réalité, l'efficacité des approches prescriptives demeure limitée. Les études montrent cependant que les gens veulent être outillés. Une clé est sans doute la flexibilité que doit offrir un guide.
- 7. Élaboration du consensus.** Une bonne variété de procédures est proposée. Le point de vue est cependant très procédural. Les approches en SST sont plus intuitives. Comme le consensus est une clé importante en SST, les chercheurs gagneraient à réfléchir et à documenter davantage les processus utilisés. Les approches actuelles devraient être mieux connues et mieux diffusées.
- 8. Formation des groupes.** Les GsG proposent surtout l'ajout de représentants, ce qui suppose que les personnes ciblées sont organisées en associations. La SST est très structurée au Québec. Cependant, la tradition est plus de travailler avec des individus et de multiplier les allers-retours. La formation des groupes est posée peut-être davantage en termes de choix d'expertise. Les développeurs de guides pourraient formaliser plus précisément leurs expériences dans ce domaine.
- 9. Rôle des organisations.** En santé, la production des guides tend à être professionnalisée. En SST, les chercheurs sont souvent étroitement associés au processus ; ils en sont régulièrement les instigateurs. La professionnalisation de la production des guides n'y est pas souhaitable. En revanche, la production de guides nécessite un support institutionnel.

Les guides

- 10. Guide sur les guides.** Les propositions sont *a priori* intéressantes. Il n'y a cependant pas de démonstration probante que cela permette d'obtenir des guides de qualité supérieure. La vision est très «pâtisserie» (mesures et ordre précis, on improvise peu). La question de la mise en forme des guides est peu abordée. En SST, la variété des contextes, des situations, des interlocuteurs, nécessite des approches plus flexibles. La vision est plus «pot au feu», ce qui n'exclut pas qu'il y ait des secteurs à pâtisserie.
- 11. National vs local.** Les guides internationaux/nationaux sont considérés de qualité supérieure aux productions locales. Ces derniers sont cependant réputés plus faciles à implanter et mieux adaptés aux contextes. La question soulevée est aussi d'ordre économique, la production de guides étant onéreuse. En SST, un même sujet est rarement couvert par un grand nombre de guides différents. Le rôle du contexte est souvent tellement important qu'il ne serait pas approprié de vouloir centraliser la production, pour ensuite viser à l'adaptation « locale ». Ce qui n'empêche pas l'adaptation locale de guides à large diffusion.

- 12. Mises à jour des guides.** L'enjeu est important en santé où les connaissances vieillissent vite. En SST, ce sont les contextes et les situations qui évoluent très vite. Les transformations ont parfois des impacts majeurs sur la signification et l'applicabilité des connaissances produites. C'est une question importante sur laquelle il faudrait plus réfléchir en SST : pour quels contextes faut-il produire des connaissances et des guides?
- 13. Longueur du guide.** Les guides plus longs respectent mieux les règles méthodologiques... mais ils sont moins utilisés. Il y a peu de données à ce sujet en SST.

Application et implantation

- 14. Applicabilité.** En santé, la nécessité de mieux considérer la question d'applicabilité fait consensus. Cette question est aussi centrale en SST. Il serait pertinent de mieux définir les attentes en matière d'applicabilité ou du moins de se questionner assez systématiquement sur l'applicabilité des recommandations.
- 15. Rôle des patients.** En santé, la nécessité de consulter les patients à différentes étapes et de prendre en compte leurs valeurs ou leurs points de vue fait consensus. On considère que les guides produits n'y accordent pas assez d'importance. La dimension relationnelle y est vue comme un élément central. En SST, le travailleur peut être vu comme la contrepartie du patient. Les guides en SST n'abordent peut-être pas suffisamment cette question. Il y a des réflexions à mener sur le processus de consultation.
- 16. Proportion des recommandations applicables.** L'approximation situant aux deux tiers la proportion de recommandations applicables est intéressante. Elle implique d'emblée l'importance de juger d'une situation avant d'appliquer une directive. Il serait intéressant de situer pour différents guides produits la proportion (recommandations, normes, etc.) qui est en fait applicable.
- 17. Processus de décision.** Les guides de pratique en santé ne seraient pas suffisamment organisés en fonction du processus décisionnel. Avoir à prendre une décision est aussi un enjeu important en SST. Les réflexions menées en santé sont pertinentes à la SST. C'est un aspect qui pourrait être plus étroitement considéré dans le développement de guides.
- 18. Implantation.** Le consensus en santé est que cet aspect a été sous-estimé, sous-financé. En SST, les processus de développement par itérations et boucles successives prennent plus en compte l'implantation. La composition des groupes consultés est considérée comme ayant un rôle majeur à cet égard. La littérature en santé fait prendre conscience de l'importance de l'enjeu et de l'importance de savoir mieux intégrer d'emblée les perspectives d'implantation.
- 19. Rôle du contexte.** Cela ressort comme un élément clé en santé dans le succès ou l'échec des guides. La production de connaissances contextualisées est bien reconnue en SST. Il y aurait cependant lieu de mieux définir et structurer le rôle des contextes.
- 20. Passage des données aux recommandations.** Cette question est très discutée en santé. Elle mériterait plus d'attention et de développement en SST. Il serait intéressant de développer

une typologie des raisons qui empêchent l'application d'une recommandation et donc en fait, le passage des données aux recommandations.

- 21. Connaissances nécessaires à l'utilisation d'un guide.** Cette question est soulevée en santé pour expliquer la sous-utilisation des guides. Il serait intéressant en SST de situer plus précisément le niveau de connaissances/savoir-faire nécessaire pour utiliser un guide.
- 22. Bénéfices visibles.** Une recommandation qui ne montre pas de bénéfice rapide serait moins appliquée. Cela est vrai aussi en SST. Il serait utile de préciser ce qui peut amener des bénéfices visibles assez rapidement pour en faire des éléments moteurs.
- 23. Approches d'implantation.** Les approches passives sont peu efficaces. Les résultats obtenus avec les diverses stratégies d'implantation sont décevants. En SST, il est important de construire nos propres stratégies. Le guide développé par Shiffman *et al.* (2005), quoiqu'intéressant, s'appliquerait mal au contexte de la SST. Le processus de réflexion est cependant pertinent pour la SST.

Évaluation

- 24. Outils d'évaluation.** Les outils proposés couvrent des items pertinents, mais ils vérifient surtout soit le processus de développement soit sa description. Il n'y a cependant pas de démonstration claire établissant qu'un guide ayant suivi ces règles de l'art sera d'emblée de bonne qualité ou mènera à un contenu utile. Il serait pertinent de développer des outils appropriés aux guides en SST, mais possiblement plus simples et plus synthétiques. À trop décomposer, on finit peut-être par passer à côté de l'essentiel ou de ce qui ne peut s'évaluer que globalement.
- 25. Rigueur vs pertinence.** Les conflits entre les objectifs de pertinence et de rigueur, ainsi que l'écart entre les connaissances scientifiques et les pratiques organisationnelles constituent un problème auquel sont confrontées autant la santé que la SST. En SST, il serait utile de baliser plus clairement cette question qui est *de facto* au cœur des guides.
- 26. Attribut des recommandations qui influencent leur adoption.** *Grol et al.* (1998) proposent des attributs pour évaluer le potentiel d'adoption des recommandations. Il serait intéressant de développer aussi une telle liste en SST.
- 27. Instruments d'évaluation des guides.** Lorsqu'on développe un guide, il peut être intéressant d'utiliser ces outils pour réviser les guides, mais de façon souple, comme pistes de réflexion. Dépendamment du type de guide, il est évident qu'un bon nombre d'items ne s'appliqueraient pas. Certains items sont pertinents mais peu opérationnels (ex. clarté). Il n'y a pas d'équivalent en SST où des repères issus du *Human Factors* peuvent cependant être pris en compte.

7 PERSPECTIVES ET RÉFLEXIONS

Cette recension des écrits a mis l'accent sur les guides de pratique clinique, largement utilisés en santé. Ce type de guides – ses caractéristiques principales – ne peut toutefois pas être représentatif et s'appliquer tel quel à l'ensemble des guides en SST, l'éventail de ces derniers étant très large. Nonobstant cette limite, nous croyons possible d'établir des parallèles et tirer des enseignements de cette vaste littérature sur les guides de pratique, à la fois pour le développement, l'implantation et l'évaluation des guides en SST. D'ailleurs, une retombée non négligeable de cette recherche est d'avoir pu mettre en évidence la « logique médicale » qui concourt au développement des guides de pratique en santé, à leur utilisation et à leurs impacts attendus. Nous estimons cet apport central, puisqu'il nous a permis de mieux comprendre la grande influence qu'a pu – et que continue d'avoir – ce courant médical, entre autres dans notre propre contexte en SST. Bien que cette influence était soupçonnée à l'amorce des travaux, l'analyse des écrits a permis de faire ressortir des traits spécifiques à cette logique, d'en comprendre les fondements et surtout les implications concrètes.

Dans la suite de ce chapitre, nous désirons revenir sur certaines de ces dimensions critiques, afin de tenter de mieux les définir et d'en apprécier les apports éventuels dans le processus de développement, d'implantation et d'évaluation des guides en SST. Ces réflexions sont complémentaires aux efforts consacrés tout au long de ce rapport pour faire des parallèles entre santé et SST, d'insister sur la spécificité de la SST, efforts qui ont cependant été circonscrits pour ne pas alourdir le rapport. Notre intention ici est de voir s'il est réaliste et pertinent en SST de s'inspirer de ces façons de faire qui ont cours en santé. En fait, bien que nous soyons convaincus que des transferts sont possibles, nous sommes aussi d'avis qu'il existe une « logique SST » qui devrait être mieux définie, les réflexions incluses dans ce chapitre étant une amorce en ce sens. Il s'agit de trouver une zone de rapprochement entre les deux logiques – de santé et de SST – pour voir quelles adaptations sont possibles et ainsi maximiser les retombées de cette étude. Cette mise en parallèle contribuera à orienter nos choix dans la conception des guides et leur utilisation projetée. Elle pourra aussi déboucher sur les propositions d'avenues de recherche pour combler les manques.

Considérant la quantité importante de dimensions mises en évidence dans cette revue, nous en développons quelques-unes qui apparaissent plus critiques. Ce faisant, nous sommes conscients de laisser de côté des aspects qu'il aurait pu être intéressant de développer. Toutefois, nous croyons qu'à travers ces quelques dimensions, nous pourrions déjà esquisser un portrait de ce qui est transférable, et surtout prendre le pouls des adaptations requises.

Le centrage sur les données probantes, sorte de preuves issues d'études scientifiques rigoureuses, comme source majeure des contenus des guides de pratique en santé, n'est pas sans avoir des impacts hors du champ de la santé. L'importance de disposer de données probantes en SST est de plus en plus présente dans le discours d'organisations œuvrant en prévention, avec le risque de voir se développer une position dogmatique sur l'utilisation exclusive ou préférentielle de ces données. Les études recensées illustrent de façon éloquente tous les efforts consentis dans le développement des guides de pratique à cette étape d'identification de données probantes, à leur validation, à leur transformation sous forme de recommandations, etc. En fait, la recension a montré à quel point tout le cycle de vie d'un guide de pratique en santé - de sa conception à son évaluation - est conditionné par cette notion de données probantes. De la qualité du contenu axé

sur l'identification des données probantes découle tout le reste : c'est cette notion qui structure la démarche, qui imprègne sa logique et sa cohérence. En santé, les données probantes portent en majorité sur l'efficacité des traitements, des démarches de soins qu'il est possible d'offrir aux patients. L'utilisation d'un protocole avec groupe contrôle pour tester l'efficacité du traitement permet de contrôler un maximum de variables confondantes et d'attribuer l'effet bénéfique à l'intervention. Il est donc clair qu'en santé, « probante » fait référence à la rigueur méthodologique, à des recherches de haute qualité.

Or, en est-il de même dans le domaine de la SST ? La notion de données probantes en SST existe-t-elle ? Si oui, a-t-elle la même signification, la même importance ? Peut-on s'en inspirer et jusqu'à quel point ? En SST, parle-t-on de connaissances « vraies » ou « fausses », ou plutôt de connaissances plus ou moins acceptables ? En SST, il nous semble que la notion structurante soit le contexte d'application et son complément direct, l'adaptation qu'il requiert. C'est aussi une préoccupation en santé, mais qui est moins centrale. D'une logique de causalité directe – propre à la médecine – nous passons à des relations complexes et multi déterminées, par ailleurs très difficiles à démontrer. La notion de relation probabiliste caractérise bien davantage la SST, qu'elle s'applique aux causes et/ou au niveau d'exposition. En SST, le traitement fait référence aux transformations des situations de travail et de ses composantes : nous sommes dans des milieux vivants, en perpétuel changement. Ces milieux évoluent à un rythme qui surpasse souvent celui permettant le renouvellement des connaissances scientifiques. Les commentaires qui vont suivre ont donc pour dénominateur commun la prise en compte ou non des contextes d'application des guides produits et les conséquences qui en découlent. Le contexte peut être tant celui des personnes appelées à utiliser les guides que le contexte dans lequel ils doivent les appliquer, le plus souvent les milieux de travail.

Un premier commentaire a trait à la méthode scientifique utilisée pour produire les données probantes et qui a généralement pour caractéristique de simplifier la réalité. Du fait du contrôle des variables indépendantes, il s'en trouve que les problématiques étudiées sont moins complexes que ce que l'on retrouve dans les milieux de travail, ce qui rend les propositions difficilement applicables. La validité interne l'emporte sur la validité externe. Peuvent être pointés aussi les contextes de laboratoires desquels est issue la majorité des données probantes et qui, particulièrement en SST, ne peuvent qu'être une simplification des situations de travail réelles.

Par ailleurs, le pragmatisme est une notion dominante en SST. La prévention des risques professionnels, bien que préoccupante pour les organisations, n'est pas leur principale raison d'être. La prévention en SST se déroule dans un contexte social où le bien-être des travailleurs peut difficilement être dissocié de la santé financière des entreprises. Bien sûr que les travailleurs sont concernés au premier plan, mais les actions débordent largement ces derniers. Pour l'intervenant en SST, ce constat implique qu'il recherche constamment le juste équilibre entre la prévention des risques et la productivité des organisations au sein desquelles il intervient, entre rigueur et convenance. Or, le premier de ces deux aspects est le plus prégnant lorsqu'il est question des données probantes. Il n'est pourtant pas question de faire l'impasse sur la qualité des savoirs à inclure dans un guide en SST, mais d'accorder de l'importance à leur potentiel d'application en contexte, ce qui peut nous amener à sélectionner certaines connaissances au détriment de d'autres, à rechercher le meilleur compromis entre la qualité des données à recueillir et leur pertinence pour faire avancer son intervention. Cet équilibre est évidemment

variable entre les différentes sphères de la SST. Par exemple, en hygiène industrielle, la mesure de l'exposition des travailleurs à des produits potentiellement cancérigènes s'approche sans doute davantage de la logique propre à la santé.

L'origine des savoirs à inclure dans les guides est une autre dimension qui distingue la SST. C'est le savoir scientifique qui est valorisé en santé, la connaissance théorique. Dans les guides de pratique, l'intention est d'ailleurs de faire la promotion auprès des intervenants de ces savoirs afin d'uniformiser les pratiques et ainsi privilégier les « meilleures pratiques ». Bien que les savoirs théoriques soient aussi importants en SST, les savoirs pratiques, ceux issus de l'expérience des intervenants, voire des travailleurs en milieu de travail, ont aussi une grande valeur. Cela s'explique par l'importance accordée en SST au contexte de travail, dont les « spécialistes » sont d'abord ceux qui y œuvrent. Cette connaissance issue du terrain a autant de valeur que la connaissance issue des sciences : elles sont complémentaires. Peut-on y voir là un lien entre la considération de ces savoirs expérimentiels et la résistance aux changements constatée dans l'appropriation des guides en santé ? N'y a-t-il pas un écart parfois trop important entre le savoir scientifique proposé ou imposé aux utilisateurs des guides et les savoirs qu'ils ont développés en situation et qui leur permet d'être performants ?

Au nombre des préjudices possibles d'une approche basée sur les données probantes, on trouve le danger de minimiser, voire d'ignorer le jugement professionnel de l'intervenant. Le jugement professionnel devient primordial quand un nombre plus important de savoirs entre en jeu dans l'intervention : on entre dans un mode de pensée « heuristique », au détriment d'un mode de pensée linéaire. L'intervenant doit faire des choix, résoudre des problèmes, et ce, en mobilisant et en combinant une série de savoirs : on parle alors de la compétence de l'intervenant. Ainsi, le ou les savoirs à considérer ne sont pas meilleurs dans l'absolu, ils le deviennent en fonction des multiples configurations que peuvent prendre les situations de travail. Se pose alors la question de la différenciation des données probantes qui peuvent être traduites sous forme de recommandations – de prescriptions dirait-on en SST – de celles qui doivent demeurer générales et constituer plutôt des repères au processus de décision de l'intervenant. Ces dernières pourraient alors prendre la forme de règles et/ou de principes.

Dans le même ordre d'idée, la plus grande intégration ou considération en SST des études de nature qualitative peut – partiellement du moins – s'expliquer par cette considération des savoirs situés et qui nécessitent une approche propre aux sciences sociales/humaines. De plus, on peut aussi y voir un impact sur la démarche menant au développement des guides. Les étapes suggérées par les GsG en santé sont séquentielles, le point de départ étant justement les données probantes à inclure dans le guide. En SST, on remarque beaucoup plus un déroulement à base d'itérations, d'aller-retour entre les concepteurs, les utilisateurs, et ce, à toutes les étapes, que ce soit pour le développement, l'implantation ou le suivi. Ce faisant, on intègre alors les savoirs d'expérience. De plus, cette façon de faire permet de coconstruire le contenu de telle sorte qu'il soit en phase avec les réalités pratiques qu'imposent les contextes d'application : on peut alors espérer une diminution de la résistance aux changements, des changements au niveau des représentations dominantes, etc.

Finalement, une autre dimension concerne les clientèles ciblées par les guides et leurs caractéristiques. En santé, les utilisateurs des guides sont des spécialistes de leur domaine : oncologue, pneumologue, gynécologue, etc. La réalité en SST est tout autre. Bien sûr, il y a de

grands domaines de spécialisation : hygiène, sécurité des machines, ergonomie, etc. Mais la réalité, c'est que les intervenants en SST sont pour la plupart des généralistes. Leur portefeuille de compétences est large et peut couvrir plusieurs problématiques : espace clos, TMS, cadenassage, ventilation, etc. Cela s'explique par le fait qu'il y a moins d'intervenants en SST, bien que les besoins soient criants. On peut y voir là un paradoxe entre cette caractéristique des clientèles des guides en SST et les remarques précédentes sur l'importance de considérer les multiples contextes dans lesquels ils interviennent : à des généralistes, ne devrait-on pas proposer des recommandations strictes pour mieux les guider dans leurs interventions ? Bien que cette dimension puisse avoir un impact sur la phase de développement des guides, c'est particulièrement au niveau de l'implantation que nous y voyons les plus grands défis. L'accompagnement, la formation à l'utilisation des guides apparaissent être une condition de succès. Bien des guides en SST ne sauraient être aussi « autonomes » que ceux produits en santé.

CONCLUSION

Cette recension des écrits s'est centrée sur les guides de pratique en santé, une production écrite répandue, mais qui ne constitue tout de même qu'une portion des diverses formes que peuvent revêtir les guides dans ce domaine. Qu'à cela ne tienne, l'exercice s'est avéré enrichissant et instructif et a permis d'apporter des éléments de réponse aux questions identifiées au début de ce projet. Force est de constater cependant qu'il n'est toujours pas évident de statuer sur l'une des questions dominantes formulée initialement, à savoir «qu'est-ce qu'un bon guide ?». En fait, les écrits consultés insistent surtout sur les caractéristiques de la démarche, du processus menant à la création du guide, plutôt que sur le guide lui-même. De même, les outils d'évaluation développés concernent aussi les étapes à suivre pour le développement du guide, mais pas le produit comme tel. L'accent est clairement placé sur le contenu par l'intermédiaire des données probantes, le contenant étant peu considéré.

Il est clair qu'un transfert direct entre ce qui se fait en santé au monde de la SST n'est pas envisageable ni réaliste, ni même souhaitable. Il ne faut cependant surtout pas en conclure que l'exercice auquel nous nous sommes prêtés a été vain. Cette étude a permis de soulever un nombre important de questions et nous a amenés à réfléchir à ce qui fait de la SST un univers singulier, et qui demande à être mieux caractérisé. Comment doit-on se positionner par rapport au courant dominant sur l'importance des données probantes ? Jusqu'à quel point doit-on formaliser les recommandations à inclure dans un guide en SST, et en fonction de quels facteurs ? L'aspect séquentiel du processus de développement des guides en santé colle-t-il bien à nos façons de faire et à nos préoccupations en SST ? Ce ne sont là que quelques-uns des questionnements auxquels nous avons tenté dans ce rapport de formuler des amorces de réponses. Nous sommes conscients que les réflexions doivent se poursuivre.

Une des conclusions à laquelle en arrivent les auteurs des écrits consultés est l'importance de mettre plus d'accent sur la phase d'implantation des guides, qui demeure par ailleurs dans le domaine de la santé une phase distincte de celle qui concerne le développement. Cette préoccupation n'est pas étrangère au fait que des études montrent que malgré des investissements substantiels, un pourcentage décevant d'utilisateurs applique les recommandations dépendamment des domaines. Il serait intéressant de réfléchir plus systématiquement aux raisons qui font obstacle à l'application d'une recommandation et de développer une typologie conséquente en SST. En ce sens, nous sommes d'avis qu'une meilleure intégration des phases de développement et d'implantation est une voie prometteuse.

BIBLIOGRAPHIE¹²⁹

Adams J.L., Fitzmaurice D.A., Heath C.M., Loudon R.F., Riaz A., Aterne A., Thomas C.P. (1999). A novel method of guideline development for the diagnosis and management of mild to moderate. *British Journal of General Practice* ; 49 (440) : 175-179.

Aasen T.M.B., Amundsen O., Mølnvik M., Tangen G. (2010). Innovation in transdisciplinary and heterogeneous collaborations: Exploring new ways of organizing environment-friendly energy research. Work in progress. The COMBI2010 conference in October 2010, in Vantaa, Finland.

Alkin, M.C. (2003). Evaluation theory and practice: Insights and new directions. *New Directions for Evaluation*; 97 : 81-89.

Alkin M.C. (2004) *Evaluation Root-Tracing theorists' views and influences*. University of California, L.A. ISBN : 9780761928942.

American Psychological Association (2002). Criteria for practice guideline development and evaluation. *American Psychologist*; 57 (12) : 1048-1051.

<http://www.apa.org/practice/guidelines/criteria.pdf>

American Psychological Association (2002). Criteria for evaluating treatment guidelines. *American Psychologist*; 57 : 1052-1059.

<http://www.apa.org/practice/guidelines/evaluating.pdf>

Armstrong P.W. (2003). Do guidelines influence practice? *Heart*; 89 (3) : 349- 352.
doi : 10.1136/heart.89.3.349.

<http://heart.bmj.com/content/89/3/349.full.pdf>

Armstrong P.W., Bogaty P., Buller C.E., Dorian P., O'Neill B.J. (2004). The 2004 ACC/AHA Guidelines: A perspective and adaptation for Canada by the Canadian Cardiovascular Society Working Group. *Can J Cardiol*; 20 (11) : 1075-9.

Aram, J. D., Salipante, P. F. (2003). Bridging scholarship in management: Epistemological reflections. *British Journal of Management*; 14 : 189–205. doi : 10.1111/1467-8551.00374.

Association médicale canadienne (1994). Principes directeurs concernant les guides de pratique clinique au Canada. Ottawa : 1-6.

Audet A.-M., Greenfield S. and Field M. (1990). Medical practice guidelines: Current activities and future directions. *Annals of Internal Medicine*; 113 (9) : 709-714.

Avenier M.-J., Bartunek J. (2010). Bridging a supposedly unbridgeable gap: elaborating scientific knowledge from and for practice. *Cahier de recherche* n° 2010-02 E4; Collection recherche. Centre d'Études et de Recherches Appliquées à la Gestion (CERAG)/CNRS. 32 p.

¹²⁹ Inclut les sources citées aux annexes.

Beaulieu M.D., Battista R.N., Blay R. (2001). À propos de l'evidence-based medicine. *Ruptures, revue transdisciplinaire en santé*; 7 (2) : 120-134.

Berg A.O., Atkins D., Tierney W. (1997). Clinical practice guidelines in practice and education. *J Gen Intern Med*; 12 (Suppl 2) : S25-S33.

Aasen B.T.M., Amundsen O., Molnvik M., Tangen G. (2010) Innovation in transdisciplinary and heterogeneous collaborations: Exploring new ways of organizing environment-friendly energy research. Call for R & D & I rendez-vous, work in progress.

Bero L.A., Grilli R., Grimshaw J.M., Harvey E., Oxman A.D., Thomson M.A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ*; 317 : 465-468.

Berti E., Grilli R. (2003). Practice guidelines and clinical governance: do the means match with the end? A quality appraisal of local practice guidelines. *Clinical Governance: An International Journal*; 8 (4) : 312-317.

Bonilauri Ferreira, A.P., Ferreira, R.F., Rajgor, D., Shah, J., Menezes, A., Pietrobon, R. (2010). Clinical reasoning in the real world is mediated by bounded rationality: implications for diagnostic clinical practice guidelines. *PLoS One*. Apr 20; 5 (4) : e10265.

Brouwers M., Charette M. (2001). Evaluation of clinical practice guidelines in chiropractic care: a comparison of North America guideline reports. *J Can Chiropr Assoc*; 45 : 141-153.

Browman G.P. (2000). Improving clinical practice guidelines for the 21st century - Attitudinal barriers and not technology are the main challenges. *International Journal of Technology Assessment in Health Care*; 16 (4) : 959-968.

Browman G.P. (2001). Development and aftercare of clinical guidelines. The balance between rigor and pragmatism. *JAMA*; 12 : 1509-1511.

Brozek J., Jankowski M., Pllaczkiwicz-Jankowska E., Jaeschke R. (2009). International diabetes federation document concerning postmeal glycemic control : assessment of quality of clinical practice guidelines using AGREE instrument. *Polskie Archiwum Medycyny Wewnetrznej*; 119 (1-2) : 18-25.

Bureau International du Travail (2002). Principes directeurs concernant les systèmes de gestion de la sécurité et de la santé au travail, ILO-OSH, Genève.

Burgers J.S., Bailey J.V., Klazinga N.S., Van der Bij A.K., Grol R., Feder G. (2002). Inside guidelines. Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. *Diabetes Care*; 25 : 1933-9.

Burgers J.S., Cluzeau F.A., Hanna S.E., Hunt C., Grol R. (2003). Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Health Care*; 19 (1) : 148-157.

- Burgers J.S., Fervers B., Haugh M., *et al.* (2004). International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation [AGREE] Instrument. *J Clin Oncol*; 22 (10) : 2000–2007.
- Burgers J.S., Grol R., Klazinga N.S., Mäkelä M., Zaat J. for the AGREE collaboration (2003). Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care*; 15 (1) : 31–45.
- Burgers J.S., Grol R.P., Zaat J.O., Spies T.H., van der Bij A.K., Mokkink G.A. (2003). Characteristics of effective clinical guidelines for general practice. *British Journal of General Practice*; 53 : 15-19.
- Cabana M.D., Rand C.S., Powe N.R., Wu A.W., Wilson M.H., Abboud P.A.C., Rubin H.R. (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*; 282 (15) :1458-1465. doi : 10.1001/jama.282.15.1458.
- Calder L., Herbert P.C., Carter A.O., Graham I.D. (1997). Review of published recommendations and guidelines for the transfusion of allogeneic red blood cells and plasma. *Canadian Medical Association Journal*; 156 (II Suppl.) : S1-8.
- Canadian Task Force on the Periodic Health Examination. (1979). The periodic health examination. *Canadian Medical Association Journal*; 121 :1193-1254.
- Carlsen B., Glenton C., Pope C. (2007). Thou shalt versus thou shalt not : a meta-synthesis of GPs' attitude clinical practice guidelines. *Br J Gen Pract*; 57 (545) : 971-978.
- Carlsen B., Kjellberg P.K. (2010). Guidelines; From foe to friend? Comparative interviews with GPs in Norway and Denmark. *BMC Health Service Research*; 10 : 17
www.biomedcentral.com/1472-6963/10/17.
- Cates J.R., Young D.N., Guerriero D.J., *et al.* (2001). Evaluating the quality of clinical practice guidelines. *J Manip Physiol Therap*; 24 (3) : 170–176.
- Cates J.R., Young D.N., Bowerman D.S., Porter R.C. (2006) An independent AGREE evaluation of the Occupational Medicine Practice Guidelines. *The Spine Journal*; 6 : 72-77.
- Choudry N. K., Stelfox H. T., Det-Shy A. S. (2002). Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* ; 287 (5) : 612-617.
- Christiaens T., De Backer D., Burgers J., Baerheim A. (2004). Guidelines, evidence, and cultural factors : comparison of four European guidelines on uncomplicated cystitis. *Scand J Prim H Care*; 22 (3) : 141-145.
- Christie C.A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. In C.A. Christie (Ed.), *The practice-theory relationship in evaluation. New directions for evaluation*; n° 97 : p. 7-35. San Francisco, CA: Jossey-Bass.

Cluzeau F.A., Littlejohns P., Grimshaw J.M., Feder G., Moran S.E. (1999). Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care*; 11 (1) : 21-28. doi : 10.1093/intqhc/11.1.21.

Comité européen de la santé (2001) Projet de recommandation 2001 du comité des ministres des États membres sur le développement d'une méthodologie dans l'élaboration des lignes directrices pour de meilleures pratiques médicales. CM (2001)136 Addendum 2.

Communauté européenne (2005). Dissémination et valorisation des résultats. Texte émis par la Direction générale Éducation et Culture, Direction Communication et Culture, Unité C3. Document, 2 pages. Commission européenne.

http://ec.europa.eu/dgs/education_culture/valorisation/index_en.html

Conseil de l'Europe (2002). Le développement d'une méthodologie dans l'élaboration de lignes directrices pour de meilleures pratiques médicales - Recommandations Re (2001)13 et exposé des motifs. ISBN 978-92-871-4787-5.

Cook D., Giacomini M. (1999). The trials and tribulations of clinical practice guidelines [editorial; comment]. *JAMA*; 281 (20) : 1950-1.

Cook D.J., Mulrow C.D., Haynes R.B. (1997). Systematic reviews : Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*; 126 (5) : 376-380.

www.annals.org/content/126/5/376.full

Cook D.J., Greengold N.L., Ellrodt A.G., Weingarten S.R. (1997) The relation between systematic reviews and practice guidelines. *Annals of Internal Medicine*; 127 : 210-216.

Cranney A., Waldegger L., Graham I.D. Man-Son-hing M., Byszewski A., Ooi D.S. (2002). Systematic assessment of the quality of osteoporosis guidelines. *BMC Musculoskeletal Disorders*; 3(20) www.biomedcentral.com/1471/2474/3/20

Crombie I.K. (1996). *The pocket guide to critical appraisal: a handbook for healthcare professionals*. London : BMJ Publishing Group.

Culleton B. (2009). Development and limitations of clinical practice guidelines. *Methods in molecular biology*; 473 : 251-261.

Dahan R., Reis S., Borkan J., Brown J.-B., Hermoni D., Mansor N., Harris S. (2008). Is knowledge a barrier to implementing low back pain guidelines? Assessing the knowledge of Israeli family doctors. *Journal of Evaluation in Clinical Practice*; 14 (5) : 785-791.

Davis D.A., Taylor-Vaisey A. (1997). Translating guidelines into practice : A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Can Med Assoc J*; 157 : 408-416.

Denis D., Lortie M., St-Vincent M., Gonella M., Plamondon A., Delisle A., Tardif J. (2011). Programme de formation participative en manutention manuelle - Fondements théoriques et approche proposé. Études et recherches / Rapport R-690, Montréal, IRSST, 172 pages.

<http://www.irsst.qc.ca/-publication-irsst-programme-formation-manutention-manuelle-fondements-theoriques-approche-proposee-r-690.html>

Desmarais L., Lortie M. (2011). *La dynamique du transfert des connaissances – Perspectives centrées sur l’usager*. Collection Santé et sécurité du travail, Presses de l’Université Laval.

Desmarais L., Parent R., Vézina N., Leclerc L. (2011). Le parcours collectif d’apprentissage organisationnel (*Learning History*) un outil formateur? Le cas d’un transfert de programme de formation entre deux équipes de recherche dispersées géographiquement (Québec vers Terre-Neuve). In : Desmarais L., Lortie M. *La dynamique du transfert des connaissances -Perspectives centrées sur l’usager*. Collection Santé et sécurité au travail. Presses de l’Université Laval.

Divac N., Maric N.P., Damjanovic A., Jovanovic, A.A., Jsovic-Gasic M., Prostran M. (2009). Use or underuse of therapeutic guidelines in psychiatry? *Psychiatria Danubina*; 21 (2) : 224-229.

Dowie R., Gregory R.P.F., Rowsell K.V., Annis S., Gick A.D., Harrison C.J. (1998). A decision analytic approach to commissioning ambulance cardiac services. *Journal of Management in Medicine*; 12 (2) : 81-91. doi : 10.1108/02689239810227100.

Dowie J. (1996). 'Evidence-based', 'cost-effective' and 'preference-driven' medicine: decision analysis based medical decision making is the pre-requisite. *J Health Serv Res policy*; 1 (2) : 104-113.

Eddy D.M. (1990). Comparing benefits and harms: the balance sheet. *JAMA*; 263 (18) : 2498-2501.

Eddy D.M. (1992). *A manual for assessing health practices & designing practice policies*. In collaboration with the Council of Medical Specialties Societies Task Force on Practice Policies. Library of Congress. ISBN 0-943126-18-5.

Edwards A. (2010). A story from abroad. Journée de réflexion de l’AETMIS. Les guides de pratique dans le secteur des services sociaux- du pourquoi au comment. 30 Avril 2010, Montréal.

Egilman D.S., Druar N.M. (2011). Corporate versus public interests: Community responsibility to defend scientific integrity. *Int J Occup Environ Health*; 17(2) : 181-185.
<http://www.law-lib.utoronto.ca/ghostwriter/ghostwriting%20ediitorial2.pdf>

Eichacker P.Q., Natanson C. (2008). Medical guidelines and performance measures: The need to keep them free of industry influence. *Mens Sana Monograph (MSM)*; 6 (1) : 22-28. doi : 10.4103/0973-1229.33005. <http://www.msmonographs.org/article.asp?issn=0973-1229;year=2008;volume=6;issue=1;page=22;epage=28;aulast=Eichacker>

Eichacker P.Q., Natanson C., Danner R.L., Lilly E. (2006). Surviving sepsis – practice guidelines, marketing campaigns. *N. Engl. J. Med*; 355 : 1640-1642.

Engelbrecht R., Courte-Wienecke S. (1999). A survey on the current state of development, dissemination and implementation of guidelines of clinical practice in European countries. Neuherberg : GSF - National Research Center for Environment and Health.

Farges G. (2004). Premiers retours d'expérience du « Guide des bonnes pratiques biomédicales en établissement de santé ». *ITBM-RBM News*, Éditions scientifiques et médicales Elsevier SAS; 25 (1) : 5- 9.

Farges G. (2007). « Guichet unique » sur le guide des bonnes pratiques biomédicales : un projet interassociatif. *IRBM News*; 8 (3) : 9-13.

Farmer A.P., Légaré F., Turcot L., Grimshaw J., McGowan J.L., Wolf F. (2009). Printed educational materials : effects on professional practice and health care outcomes (Review). *The Cochrane Library*; Issue 3.

Faye, C., Lortie, M., Desmarais, L. (2007). Guide sur le transfert des connaissances - à l'intention des chercheurs en santé et sécurité du travail. Réseau de recherche en santé et sécurité du travail (RSSTQ) - Communauté transfert des connaissances. <http://www.rrsstq.qc.ca/>

Feder G., Eccles M., Grol R., Griffiths C., Grimshaw J. (1999). Clinical guidelines: using clinical guidelines. *BMJ*; 318 : 728-30.

Fervers B., Burgers J.S., Haugh M.C., Latreille J., Mlika-Cabanne N., Paquet L. *et al.* (2006). Adaptation of clinical guidelines : literature review and proposition for a framework and procedure. *International Journal for Quality in Health Care*; 18 (3) : 167-176.

Field M.J., Lohr K.N. (1990). Clinical practice guidelines: Directions for a new program. Institute of Medicine, Washington, DC : National Academy Press.

Field M.J., Lohr K.N. (1992). Guidelines for clinical practice: from development to use. Institute of Medicine, Washington, DC : National Academy Press.

Fischer Walker C.L., Rimoin A.W., Hamza H.S., Steinhoff M.C. (2006). Comparison of clinical prediction rules for management of pharyngitis in settings with limited resources. *The Journal of Pediatrics*; 149 (1) : 64-71.

Fretheim A., Schünemann H.J., Oxman A.D. (2006). Improving the use of research evidence in guideline development : 15. Disseminating and implementing guidelines. *Health Res Policy Syst*; 4 : 27. doi : 10.1186/1478-4505-4-27.

Gartlehner G. , West S.L., Lohr K., Kahwati L., Johnson J.G., Harris R.P., Whitener L., Voisin C.E., Sutton S. (2004). Assessing the need to update prevention guidelines : a comparison of two methods. *International Journal for Quality in Health Care*; 16 (5) : 399-406. doi : 10.1093/intqhc/mzh081.

Gates P.E. (1995). Think globally, act locally : an approach to implementation of clinical practice guidelines. *Jt Comm J Qual Improv*; 21 (2) : 71-84.

Giacomini M.K., Cook D.J. (2000). Users' guides to the medical literature. XXIII. Qualitative Research in Health Care. B. What are the results and how do they help me care for my patients? *JAMA*; 284 (4) : 478-482. doi : 10.1001/jama.284.4.478.

Gibbons M., Limoges C., Nowotny H., Schwartzman S., Scott P. and Trow M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. London : Sage. Réimpression en 2004 : 179 p.

Gibbons M. (2001). Innovation and the developing system of knowledge production. Paper presented at Centre for Policy Research on Science and Technology at Simon Fraser University Summer Institute on innovation, competitiveness and sustainability in the North American region. <http://edie.cprost.sfu.ca/summer/papers/Michael.Gibbons.html>

Giraud L., Massé S., Dubé J., Schreiber L., Turcot A. (2004). Sécurité des convoyeurs à courroie : généralités, protection contre les phénomènes dangereux - Guide de l'utilisateur. Guide R-385, Montréal, IRSSST, 85 pages.

Gold N. (1983). Stakeholders and program evaluation. *New Directions for Program Evaluation*; 17 : 63–72.

Graham I.D., Calder L.A., Hébert P.C., Carter A.O., Tetroe J.M. (2000). A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care*; 16 (4) : 1024-1038.

Graham I.D., Harrison M., Brouwers M. (2003). Evaluating and adapting practice guidelines for local use: a conceptual framework. In Pickering S, Tomlinson J., eds. *Clinical Governance and Best Value: Meeting the Modernisation Agenda*, London, Churchill Livingstone; 213–229.

Graham I.D., Harrison M.B. (2005). Evaluation and adaptation of clinical practice guidelines. *Evid. Based Nurs*; 8 : 68-72. doi : 10.1136/ebn.8.3.68.

Gravel S., Lortie M., Bilodeau H., Dubé J. (2010) La résilience des entreprises face aux dilemmes de gestion des problèmes de santé et sécurité du travail ayant une incidence sur la gestion des ressources humaines. 44^{ème} Congrès de la Société d'Ergonomie de Langue Française (SELF), Liège, pp.188-193.

Greenhalgh T. (1997) How to read a paper: Papers that summarise other papers (systematic reviews and meta-analysesou analysis?). *BMJ*; 315 (7109) : 672.

Greer N., Mosser G., Logan G., Halaas G.W. (2000) A practical approach to evidence grading. *Joint Commission Journal on Quality Improvement*; 26 : 700-712.

Grilli R., Lomas J. (1994). Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Med Care*; 32 (4) : 202-213.

Grilli R., Magrini N., Penna A., Mura G., Liberati A. (2000). Practice guidelines developed by specialty societies: the need for a critical appraisal. *The Lancet*; 255 (9188) : 103-106. doi : 10.1016/S0140-6736(99)02171-6.

- Grimshaw J.M., Russell I.T. (1993). Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *The Lancet*; 342 (8883) : 1317–1322. doi :10.1016/0140-6736(93)92244-N.
- Grimshaw J.M., Eccles M.P., Russell I. (1995). Developing clinically valid practice guidelines. *Journal of Evaluation in Clinical Practice*; 1(1) : 37-48.
- Grimshaw J.M., Thomas R.E., MacLennan G., Fraser C., Ramsay C.R., Vale L. *et al.* (2004). Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology assessment*; 8 (6) : 94 p. <http://www.hta.ac.uk/execsumm/summ806.htm>
- Grimshaw J.M., Thomas R.E., MacLennan G., Fraser C., Ramsay C.R., Vale L. *et al.* (2005). Effectiveness and efficiency of guideline dissemination and implementation strategies. *International Journal of Technology Assessment in Health Care*; 21 (1) : 149-149. doi : 10.1017/S0266462305290190.
- Grol R. (1997). Beliefs and evidence in changing clinical practice. *BMJ*; 315 : 418-421.
- Grol R., Dalhuijsen J., Thomas S., In't Veld C., Rutten G., Mookink H. (1998). Attributes of clinical practice guidelines that influence use of guidelines in general practice: observational study. *BMJ*; 317 : 858-861.
- Grol R., Eccles M.P., Maisonneuve H., Woolf S.H. (1998). Developing clinical practice guidelines. The European experience. *Dis Manage Health Outcomes*; 4 : 255–66.
- Grol R., Grimshaw J. (1999). Evidence-based implementation of evidence-based medicine. *Jt Comm J Qual Improv*; 25 : 503–510.
- Guyatt G.H., Sackett D.L., Cook D.J. (1993). Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA*; 270 : 2598-2601.
- Guyatt G.H., Sackett D.L., Hayward R., Cook D.J., Cook (1994). Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA*; 271 (1) : 59-63.
- Guyatt G.H., Sackett D.L., Sinclair J.C., Cook D.J., Cook R.J. and Evidence-Based medicine Working Group. (1995). Users' guides to the medical literature. Users' guides to the medical literature. IX. A Method for Grading Health Care Recommendations *JAMA*; 274(22):1800-1804.
- Hannes K., Van Royen P., Aertgeerts B., Buntinx F., Ramaekers D., Chevalier P. (2005). La validation systématique de guides de pratique clinique : l'instrument AGREE, élaboration et diffusion d'une liste internationale de critères. *Rev Med Liège*; 60 (12) : 949-956.
- Harbour R., Miller J. (2001). A new system for grading recommendations in evidence based guidelines. *BMJ*; 323 : 334-6.

Harpole L.H., Kelley M.J., Schreiber G., Toloza E.M., Kolimaga J., McCrory D.C. (2003). Assessment of the scope and quality of clinical practice guidelines in lung cancer. *Chest*; 123 (1 Suppl) : 7S– 20S.

Harris R.P., Helfand M., Woolf S.H., Lohr K., Mulrow C.D., Teutsch S.M., Atkins D. (2001). Current methods of the US Preventive Services Task Force. *Am J Prev Med*; 20 : 21-35.

Harris G., Roberts J. (2007). Doctor's ties to drug makers are put on close view. *The New York Times*; 21 march.
<http://www.nytimes.com/2007/03/21/us/21drug.html?ex=1332129600and%20en=8ab21926768519a3andei=5090andpartner=r>.

Hayward R.S., Wilson M.C., Tunis S.R., Bass E.B., Guyatt G.H. (1995). Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *JAMA*; 274 : 570–574.

Hayward R.S., Guyatt G.H., Moore K.A., McKibbon A., Carter A.O. (1997). Canadian physicians' attitudes about and preferences regarding clinical practice guidelines. *Canadian Medical Association Journal*; 156 (12) : 1715-1723.

Hayward R.S., Wilson M.C., Tunis S.R., Bass E.B., Rubin H.R., Haynes R.B. (1993). More informative abstracts of articles describing clinical practice guideline. *Annals of Internal Medicine*; 118 (9) : 731-737. <http://www.annals.org/content/118/9/731.full.pdf>

Helou A., Ollenschläger G., (1998). Goals, possibilities and limits of quality evaluation of guidelines. A background report on the user manual of the 'Methodological quality of guidelines' check list. *Z Arztl Fortbild Qualitatssich*; 92 : 361-365

Hibble A., Kanka D., Pencheon D., Pooles F. (1998). Guidelines in general practice : the new Tower of Babel? *BMJ*; 317 (7162) : 862-863.

Hill K.M., Lalor E.E. (2009). How useful is on line tool to facilitate guideline implementation? Feasibility study of using GLIA by stroke clinicians in Australia. *Qual and Saf in Health Care*; 18 (2) : 157-159.

Hutchinson A., McIntosh A., Anderson J., Gilbert C., Field R. (2003). Developing primary care review criteria from evidence-based guidelines: coronary heart disease as a model. *Br J Gen Pract*; 53 (494) : 690-696.

Institut national de santé publique du Québec (2001). Dissémination des guides de pratiques chez les médecins. 51 p.

Institut national d'excellence en santé et en services sociaux (INESSS). (2011). Processus d'élaboration d'un guide de pratique dans le secteur des services sociaux : Abrégé. Document préparé par S. Beauchamp et J.-P. Duplantie avec la collaboration de C. Mercier.
<http://www.inesss.qc.ca/>

Jackson R., Feder G. (1998). Guidelines for clinical guidelines - Editorials. *BMJ*; 317 : 427-428.

Jaeschke R., Guyatt G.H., Sackett D.L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*; 271 (9) : 703-707.

Jaeschke R, Guyatt G .H., Sackett D.L. (1994). Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA*; 271 (5) : 389-391.

Jobson K.O., Potter W.Z. (1995). International psychopharmacology algorithm project report. *Psychopharmacol Bull*; 31 : 457-459.

Kassirer J. (2004). *On the take: How medicine's complicity with business can endanger your health*. New York : Oxford University Press.

Katrak P., Bialocerkowski A.E., Massy-Westropp N., Kumar S., Grimmer K.A. (2004). A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*; 4 4 : 22. doi : 10.1186/1471-2288-4-22. <http://www.biomedcentral.com/1471-2288/4/22>

Kellett J., Clarke J. (1995). Comparison of 'accelerated' tissue plasminogen activator with streptokinase for treatment of suspected myocardial infarction. *Medical Decision Making*; 15 : 297-310. doi: 10.1177/0272989X9501500401.

Kryworuchko J., Stacey D., Bai N., Graham I.D. (2009). Twelve years of clinical practice guideline development, dissemination and evaluation in Canada (1994 to 2005). *Implementation Science*; 4 (49) doi :10.1186/1748-5908-4-49.

Laroche, E. (2009). *Étude de la production de la recherche et de ses mécanismes de transfert – une application au domaine de la santé et de la sécurité du travail au Canada*. Thèse de doctorat, Département de management, Faculté des sciences de l'administration, Université Laval, Québec (Québec).

Levy P. (1994) *L'intelligence collective - Pour une anthropologie du cyberspace*. Ed.L la découverte/Sciences et Société : 244 p.

Liddle J., Williamson M., Irwig L. (1996). Method for evaluating research and guideline evidence. Sydney : New South Wales Department of Health; 54 p.
<http://www.health.nsw.gov.au/pubs/1996/pdf/mergetot.pdf>

Ling T., Villalba van Dijk L. (2009). The performance audit handbook. Routes to effective evaluation. RAND Europe. Technical Report TR-788-RE : 221 p.
http://www.rand.org/pubs/technical_reports/TR788.html

Littlejohns P., Cluzeau F., Bale R., Grimshaw J., Feder G., Moran S. (1999). The quantity and quality of clinical practice guidelines for the management of depression in primary care in the UK. *Br. Med. J. GP*; 49 : 205–210.

Leentjens A.F., Burgers J.S. (2008). What factors are important for the successful implementation of guidelines. *Tijdschr Psychiatr*; 50: 329-335.

Lohr K.N., Field M.J. (1992). A provisional instrument for assessing clinical practice guidelines. In : Field, M.J., Lohr, K.N. (eds). *Guidelines for clinical practice. From development to use*. Washington DC : National Academy Press.

Lomas J. (1993). Diffusion, dissemination and implementation: who should do what? *Annals of the New York Academy of Sciences*; 703 : 226-257.

Lomas J., Anderson G.M., Domnick-Pierre K., Vayda E., Enkin M.W., Hannah W.J. (1989). Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med*; 321 (19) : 1306–1311.

Lomas J., Culyer T., McCutcheon C., McAuley L., Law S. (2005). Rapport final : Conceptualiser et regrouper les données probantes pour guider le système de santé. Ottawa : Fondation canadienne de la recherche sur les services de santé.

Lugtengberg M., Zegers-van Schaick J.M., Westert G.P., Burgers J.S. (2009). Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implementation Science*; 4 : 54 doi :10.1186/1748-5908-4-54.

Manna D.R., Bruijnzeels M.A., Mokkink H.G., Berg M. (2003). Ethnic specific recommendations in clinical practice guidelines: a first exploratory comparison between guidelines from the USA, Canada, the UK, and the Netherlands. *Qual Saf Health Care*; 12 : 353–358.

Mäkelä M., Thornsen T.A. (1999). A framework for guidelines implementation studies. In : Changing professional practice : Theory and practice of clinical guidelines implementation. Copenhagen; Danish Institute for Health Service; 23-53.

Marshall J.K. (2000) A critical approach to clinical practice guidelines. *Can J Gastroenterol*; 14 (6) : 505-509.

Marriott S., Palmer C., Lelliott, P. (2000). Disseminating healthcare information : getting the message across. *Quality and Safety in Health Care*; 9 : 58-62.

McKibbon A., Eddy A., Marks S. (1999). *PDQ: Evidence-based Principles and Practice*. B.C. Decker, Hamilton, Ontario.

Mendelson E.B. (1995). The development and meaning of appropriateness guidelines. *The Radiologic clinics of North America*; 33 (6) : 1081–1084.

Mercier C. (2010). Le guide de pratique : de quoi s'agit-il? Présentation lors de la journée de réflexion AETMIS : Les guides de pratique dans le secteur des services sociaux : du pourquoi au comment. Montréal, 30 avril 2010.

<http://btech2.syspark.net/interteddi/publications/presentations/mercier-2010>

- Mottur-Pilson C. (1995). Internists' evaluation of guidelines: the IMCARE Practice Guidelines Network. *International Journal for Quality in Health Care*; 7 (1) : 31–37. doi: 10.1093/intqhc/7.1.31.
- Mottur-Pilson C. (1993). User-friendly guidelines: The missing link? *Journal of Continuing Education in the Health Professions*; 13 (3) : 221-228.
- Mulrow C.D. (1987). The medical review article: state of the science. *Ann. Inter Med*; 106 : 485-488.
- Mulrow C.D. (1994). Systematic Reviews: Rationale for systematic reviews. *British Medical Journal*; 309 : 597-599.
- Murphy M.K., Black N.A., Lampin D.L., McKee C.M., Sanderson C.F., Ashkam J., Marteau T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*; 2 (3) : 87 p.
- National Health and Medical Research Council (1999). A guide to the development, implementation and evaluation of clinical guidelines - Endorsed 16 November 1998. Commonwealth of Australia 1999. www.ausinfo.gov.au/general/gen_hottobuy.htm. ISBN 1864960485
- National Institute for Health and Clinical Excellence (2009). The guidelines manual. First issue (2004) www.nice.org.uk
- Nuffield Institute for Health (University of Leeds) Center for Health Economic and NHS Centre for Reviews and Dissemination (U. of York) (1994). Implementing Clinical Practice guidelines : Can guidelines be used to improve clinical practice? *Effective Health Care*; Bulletin n° 8 : 12 p.
- New Zealand Guidelines Guide (2001). Handbook for the preparation of explicit evidence-based clinical practice guidelines. Wellington, Nouvelle-Zélande. ISBN : 0-473-08181-4. <http://www.nzgg.org.nz>
- Ohno-Machado L., Gennari J.H., Murphy S.N., Jain N.L., Tu S.W., Oliver D.E., Pattison-Gordon E., Greenes R.A., Shortliffe E.H., Barnett G.O. (1998). The GuideLine Interchange Format: A model for representing guidelines. *JAMA*; 5 (4) : 357-372.
- Ollenschläger G., Marshall C., Qureshi S., Rosenbrand K., Burgers J., Mäkelä M., Slutsky J., Board of Trustees 2002, Guidelines International Network (G-I-N) (2004). Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guidelines International Network (G-I-N). *Qual Saf Health Care*; 13 (6) : 455-460. doi :10.1136/qshc.2003.009761.
- Ongenaes F., De Backere F., Steurbaut K., Colpaert K., Kerckhove W., Decruyenaere J., De Turck F. (2010). Towards computerizing intensive care sedation guidelines: design of a rupe-based architecture for automated execution of clinical guidelines. *BMC Medical Informatics and Decision Making*; 10 (3) : 22 p. doi :10.1186/1472-6947-10-3. www.biomedcentral.com/1472-6947/10/3

Organisation de coopération et de développement économique (2002). Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats. Série : L'évaluation et l'efficacité de l'aide; n° 6.

Organisation mondiale de la santé (OMS). Guidelines for WHO guidelines. Genève, Suisse : OMS; 2003. Disponible à : http://whqlibdoc.who.int/hq/2003/EIP_GPE_EQC_2003_1.pdf.

Oxman A.D., Guyatt G.H. (1993). The science of reviewing research. *Ann N Y Acad Sci*; 703 : 125-133.

Oxman A.D., Sackett D.L., Guyatt G.H. (1993). Users' Guides to the Medical Literature I. How to Get Started. *JAMA*; 270 (17) : 2093-2095.

Oxman A.D., Thomson M.A., Davis D.A., Haynes R.B. (1995). No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Canadian Medical Association Journal*; 153 : 1423-1431.

Oxman A.D., Shünemann H.J., Fretheim A. (2006) Improving the use of research evidence in guideline development : 16. Evaluation. *Health Research Policy and Systems*, 4 : 28. doi : 10.1186/1478-4505-4-28.

Parker G. , Fletcher K. (2007). Treating depression with the evidence-based psychotherapies: a critique of the evidence. *Acta Psychiatr Scand*; 115 : 352-359.

Pencharz J.N., Grigoriadis E., Gwenderlyn F.J., Bombardier C.A. (2002). A critical appraisal of clinical practice guidelines for the treatment of lower-limb osteoarthritis. *Arthritis Res*; 4 (1) : 36-44. doi:10.1186/ar381. <HTTP://ARTHRITIS-RESEARCH.COM/CONTENT/4/1/36>

Perrenoud P. (1999). Transférer ou mobiliser ses connaissances? D'une métaphore à l'autre : implications sociologiques et pédagogiques. Texte remanié et complété d'une communication au colloque « Raisons éducatives sur les compétences ». Faculté de psychologie et des sciences de l'éducation, Université de Genève; mars 1999. 24 p. site web : www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_1999/1999_28.rtf

Pressman J. L., Wildavsky A. (1973). *Implementation: How Great Expectations in Washington are Dashed in Oakland*. Los Angeles : University of California Press.

Reddy K.S. (1999). Implementation of international guidelines on hypertension: the Indian experience. *Clin Exp Hypertens*; 21 (5-6): 693-701.

Reed G.M., McLaughlin C.J., Newman R. (2002). American Psychological Association policy in context. The development and evaluation of guidelines for professional practice. *Am Psychol*; 57 (12) : 1041-1047.

Rogers E.M. (1995). *Diffusion of Innovations*. Fourth edition. New-York : Free Press.

Rosser W.W., Palmer W.H. (1993). Dissemination of Guidelines on cholesterol. *Canadian Family Physician*; 39 : 280-284.

Roy M., Guindon J.-C., Fortier L. (1995). Transfert de connaissances– revue de littérature et proposition d'un modèle. Études et recherches, IRSST; R-099 : 53 p.
http://www.irsst.qc.ca/fr/publicationirsst_460.html

Roy S., Durand M.-J., Corriveau H. (2011). L'évaluation des capacités reliées au travail pour une clientèle présentant une déficience physique - Guide de pratique à l'intention des ergothérapeutes. Études et recherches / Guide technique RG-703, IRSST, 76 p.

Sackett D.L. (ed) (1995). *The Cochrane Collaboration Handbook*. Oxford : Cochrane Collaboration.

Sackett D.L., Rosenberg, W.M., Gray J.A., Haynes R.B., Richardson W.S. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ*; 312 (7023) : 71.

Sanders G.D., Nease R.F. Jr, Owens D.K. (2000). Design and pilot evaluation of a system to develop computer-based site-specific practice guidelines from decision models. *Medical Decision Making*; 2000; 20 (2) : 145-159.

Sankaran G. 2006. Knowledge diffusion from DBA research. *Online Journal of Knowledge Management*; 3 (1) : 25-33. ISSN: 1834-3554.
<http://www.actkm.org/userfiles/File/actKMjnl/2006/Knowledge%20diffusion%20from%20DBA%20research.pdf>

Savoie I., Green C., Bassett K. *et al.* (1996) Critical appraisal criteria for clinical practice guidelines. Vancouver : BC. Office of technology Assessment (cité par Vlayen *et al.* 2005).

Selker H.P. (1993). Criteria for adoption in practice of medical practice guidelines (editorial). *Am. J. Cardiol*; 71 : 339-341.

Scottish Intercollegiate Guidelines Network (1995). Clinical guidelines: criteria for appraisal for national use. Edinburgh : SIGN; 1995.

Shaneyfelt T.M., Mayo-Smith M.F., Rothwangl J. (1999). Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*; 281 (20) : 1900-1905. doi : 10.1001/jama.281.20.1900. <http://jama.ama-assn.org/content/281/20/1900.short>

Shekelle P.G., Ortiz E., Rhodes S., Morton S.C., Eccles M.P., Grimshaw J.M. Woolf S.H. (2001). Validity of the agency for healthcare research and quality clinical practice guidelines. How quickly do guidelines become outdated. *JAMA*; 286 (12) :1461-1467.

Shekelle P.G., Kravitz R.L., Beart J., Marger M., Wang M., Lee M. (2000) Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. *Health Serv Res*; 34 (7) : 1429-1448.

Shekelle P.G., Woolf S.H., Eccles M.P., Grimshaw J.M. (1999). Developing clinical guidelines. *West J Med* ; 170 (6) : 348-351.

Shiffman R.N., Shekelle P., Overhage J.M., Slutsky J., Grimshaw J., Deshpande A.M. (2003). Standardized reporting of clinical practice guidelines: a proposal from the conference on guideline standardization. *Ann Intern Med*; 139 (6) : 493–498.

Shiffman R.N., Michel G., Essaihi A., Thornquist E. (2004). Bridging the guideline implementation gap : A Systematic, document-centered approach to guideline implementation. *Journal of the American Medical Informatics Association*; 11 (5) : 418-426.

Shiffman R.N., Dixon J., Brandt C., Essaihi A., Hsiao A., Michel G., O'Connell R. (2005). The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Medical Informatics and Decision Making*; 5 : 23. doi : 10.1186/1472-6947-5-23.

Shünemann H.J., Fretheim A., Oxman A.D. (2006). Improving the use of research evidence in guideline development : 1. Guidelines for guidelines. *Health Research Policy and Systems*; 4 : 13. doi : 10.1186/1478-4505-4-13. <http://www.health-policy-systems.com/content/4/1/13>.

Smith, J.K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*; 12 (3) : 6-13.

Stern M., Brennan S. (1994). Medical audit in the hospital and community services. London : Department of Health.

Stricker, G., Abrahamson, D.J., Bologna, N.C., Hollon, S.D., Robinson, E.A, Reed, G.M. (1999). Treatment guidelines: The good, the bad, and the ugly. *Psychotherapy*; 36 : 69-79.

Tcaciuc R., Lortie M., Patry L. (2011) Interventions à la Commission des lésions professionnelles : utilisation des connaissances sur le troubles musculo-squelettiques et questions soulevées par les intervenants. In : Desmarais L., Lortie M. *La dynamique du transfert des connaissances - Perspectives centrées sur l'utilisateur*. Collection Santé et sécurité au travail. Presses de l'Université Laval.

The AGREE Collaboration (2001). Instrument. 20 p.
<http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf>

The AGREE Collaboration (2002). Grille d'évaluation de la qualité des recommandations pour la pratique clinique. Version en langue française de : Appraisal of Guidelines for Research & Evaluation Instrument. 22 p. <http://www.agreecollaboration.org/pdf/fr.pdf>

The AGREE Collaboration (2003). *Instrument Training Manual*. 69 p.
<http://www.agreecollaboration.org/pdf/aitraining.pdf>

The AGREE Collaboration (2003). Writing Group: Cluzeau F.A., Burgers J.S., Brouwers M., Grol R., Mäkelä M., Littlejohns P., Grimshaw J., Hunt C. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care*; 12 (1): 18-23.

The AGREE Collaboration (2000). Writing Group : Thomason M., Cluzeau F., Littlejohns P. Guideline development in Europe: an international comparison. *International Journal of Technology Assessment in Health Care*; 16 (4) : 1036-1046.

The GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. *BMJ*; 328 : 1490-4.

Trépanier J., Desmarais L., Giraud L., Bélanger J. (2011) Évaluation des impacts du guide de sécurité sur les convoyeurs à courroies: un portrait de son utilisation par les gens du milieu. In Desmarais L. et Lortie M. (2011) *La dynamique du transfert des connaissances - Perspectives centrées sur l'usager*. Collection Santé et sécurité au travail. Presses de l'Université Laval.

Turner T., Misso M., Harris C., Green S. (2008). Development of evidence-based clinical practice guidelines (SPGs) : comparing approaches. *Implementation Science*; 3 : 45.
doi :10.1186/1748-5908-3-45. <http://www.implementationscience.com/content/3/1/45>

Twohig P.L., Putnam W. (2002). Group interviews in primary care research : advancing the state of the art or ritualized research? *Family Practice*; 19 (3) : 278-284. doi : 10.1093/fampra/19.3.278.

Ulvenes L.V., Aasland O., Nylenna M., Kristiansen I.S. (2009) Norwegian physicians' knowledge of and opinions about evidence-based medicine : cross-sectional study. *PlosONE*, 4 (11) e7828. www.plosone.org

US Preventive Services Task Force (1989). *Guide to clinical preventive services: an assessment of the effectiveness of 169 interventions* (ed M. Fisher), Williams and Williams, Baltimore.

Vlayen J., Aertgeerts B., Hannes K., Sermeus W., Ramaekers D. (2005). A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *Int J Qual Health Care*; 17 (3) : 235-242. doi : 10.1093/intqhc/mzi027.

Ward J.E., Grieco V. (1996). Why we need guidelines for guidelines: a study of the quality of clinical practice guidelines in Australia. *Med J Aust*; 165 (10) : 574-576.

Watine J., Friedberg B., Nagy E., Onody R., Oosterhuis W. *et al.* (2006). Conflict between guideline methodologic quality and recommendation validity: A potential problem for practitioners. *Clinical Chemistry*; 52 (1) : 65-72.

Wilson J. R., Corlett, N. C. (1995). Evaluation of human work. CRC Press : Boca Raton (Florida) ; 1026 pages. ISBN : 978-0-415-26757-1.

Woolf S.H. (1990). Practice guidelines: a new reality in medicine. I. Recent developments. *Arch Intern Med*; 150 (9) : 1811-1818.

Woolf S.H. (1992) Practice guidelines : a new reality in medicine II. Method of developing guidelines. *Arch Intern Med*; 152 : 946-952.

Woolf S.H. (1995). Practice guidelines : What the Family Physician Should Know. *American Family Physician*; 51 (6) : 1455-1463.

Woolf S.H., Grol R., Hutchinson A., Eccles M.P., Grimshaw J.M. (1999). Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ*; 318 (7182) : 527–530.

WHO World Health Organization (2003). Guidelines for the WHO guidelines.
http://whqlibdoc.who.int/hq/2003/EIP_GPE_EQC_2003_1.pdf

ANNEXES

Annexe 1.1 – Survol du rôle des États dans la production des guides à l’aube des années 2000¹³⁰ et de quelques acteurs clés.

Canada - Le Canada a joué un rôle important à travers les travaux menés par le groupe d'étude canadien sur l'examen médical périodique (créé en 1976) qui ont joué un rôle précurseur important dans l'élaboration de directives fondées sur les preuves, la formalisation du processus et le concept de niveau de preuve. Ce groupe a publié ce qui est considéré comme le ou un des premiers guides fondés sur des données probantes et dont les recommandations allaient à l'époque à l'encontre des croyances médicales. Il a été le premier à développer une échelle de gradation de la qualité des preuves scientifiques. L'Association médicale canadienne collabore par ailleurs fortement au développement des guides. Elle a mis en place une base de données sur les guides qui propose aussi un guide sur le développement et l'implantation des guides de pratiques cliniques (Kryworuchko *et al.* 2009). (Beaulieu *et al.* 2001). http://www.cma.ca/index.php/ci_id/54316/la_id/1.htm.

Royaume-Uni — Un des pays où la tradition est la plus ancienne. Les guides sont devenus un instrument pour implanter les méthodes de soins fondées sur des preuves d'efficacité, décrites ou sélectionnées ou désignées ou répertoriées à partir de méthodes explicites pour les déceler plutôt que sur la base de conférences de consensus. Le National Health Service utilise un instrument systématique d'évaluation. On y retrouve deux acteurs clés dans le développement des guides sur les guides :

- **NICE** : Le National Institute for Health and Clinical Excellence. Il s'agit de la fusion en 2005 du National Institute for Clinical Excellence, mis en place en 1999, avec le Health Development Agency. NICE fait partie du National Health Center. Son rôle est de produire des avis-directives indépendants, exécutifs (*authoritative*) et fiables quant aux questions portant sur la promotion de la santé, ainsi que la prévention et le traitement des maladies pour les patients et les soignants, les professionnels de la santé et le public élargi. Son rôle vient d'être transformé pour n'être que consultatif.
- **SIGN** : Scottish Intercollegiate Guideline Network. Il a été mis en place en 1993 par l'Academy of Royal Colleges and their faculties in Scotland. Son mandat est de développer des guides cliniques basés sur les données probantes pour le National Health Service in Scotland.

États-Unis — Voir le chapitre 1.

Nouvelle-Zélande et Australie — On y retrouve deux acteurs clés :

- **NZGG** : New Zealand Guidelines Group. Mis en place en 1996 par le National Health Committee en tant que réseau informel d'expertise et d'information sur le développement et l'implantation des guides, il s'est incorporé en 1999 comme société indépendante sans but lucratif. Il demeure cependant financé en partie par le ministère de la Santé et à partir de contrats provenant des diverses agences gouvernementales.
- **NHMRC** : National Health and Medical Research Council of Australia. Il a développé un guide sur les guides.

Hollande — L'organisation la plus importante est le Dutch College of General Practitioners qui produit des guides depuis 1987, au rythme de 8 à 10 par année. Les procédures d'analyse de littérature sont standardisées, mais elles demeurent combinées à des approches par consensus réunissant experts et praticiens. Un programme d'implantation systématique a été mis en place.

¹³⁰ Sources principales : Woolf *et al.* 1999; Grol *et al.* 1998. Le portrait est donc sans doute incomplet par rapport à 2011, malgré quelques ajouts.

Finlande et Suède — Les organisations nationales et locales finlandaises ont produit plus de 700 « directives » depuis 1989. En Suède, l'organisation dominante, le Swedish Council on Technology Assessment in Health Care, est centré sur les questions de technologies.

France — L'Agence nationale de l'accréditation et d'évaluation en santé qui, à la fin des années 90, avait publié plus de 100 guides, privilégiait le développement à partir de conférences de consensus ou l'adaptation de guides provenant d'autres pays. Ils sont évalués sur la base d'audits. L'ANAES a été remplacée en 2004 par la Haute autorité de santé (HAS). www.has-sante.fr

Québec — La Loi sur l'Institut national d'excellence en santé et en services sociaux adoptée le 11 juin 2010 précise ainsi sa mission : Évaluer les avantages cliniques et les coûts des technologies médicales, des médicaments et des interventions. Préparer des recommandations et préparer des guides de pratique clinique pour s'assurer de leur usage optimal. (Source : Mercier 2010) <http://translate.google.ca/translate?hl=fr&langpair=en/fr&u=http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type%3D5%26file%3D2010C15A.PDF> Consulté le 20 juillet.

Annexe 2.1 - Les attributs des guides proposés par l'Institute Of Medicine et les ajouts proposés par divers groupes de travail

Attributs proposés par l'IOM (1992) et leurs ajouts subséquents	Inclusions proposées par divers groupes de travail
Validité	
<p>Valide si génère de meilleurs résultats (santé et coût)</p> <p>Évaluation prospective devra considérer les</p> <ul style="list-style-type: none"> - résultats anticipés en santé - coûts des alternatives - relations preuves – recommandation - qualité des preuves scientifiques/cliniques, consensus professionnel - moyens utilisés pour évaluer preuves - description des coûts - description des bénéfices, des risques - conflits potentiels entre les différentes directives 	<p>AMC, 1994 Tenir compte des problèmes d'éthique dans toutes les phases du processus d'élaboration de CPG</p> <p>NHMRC - Choix de la méthode utilisée pour la revue systématique : doit tenir compte des coûts et du temps (rigueur vs pragmatisme) - The GRADE working group fournit une approche structurée pour graduer la qualité des évidences et faciliter la comparaison des recommandations.</p>
Fiabilité/reproductibilité	
<p>Si, avec les mêmes preuves et méthodes que pour le développement des GdP, d'autres groupes d'experts arrivent aux mêmes conclusions</p> <p>Si, avec les mêmes conditions cliniques, les GdP sont interprétés et appliqués de façon consistante</p> <p>Évaluation prospective de la fiabilité pourrait considérer 1) résultats d'une évaluation externe indépendante; 2) pré-tests</p>	<p>AHCPR Large distribution géographique des experts</p>
Applicabilité clinique	
<p>Définir de façon appropriée la population des patients; préciser la population cible</p> <p>Discussion des problèmes cliniques complexes</p> <p>Raisons pour exclure des patients</p>	<p>Vlayen Préciser but, objectifs, les raisons et les sujets Discriminer patients inclus vs exclus Préciser les pourvoyeurs (<i>provider</i> population)</p> <p>AHCPR Ex. : groupes ethniques, économiques ou cliniques</p> <p>Vlayen et APA Aspects éthiques APA Compatibilité avec les politiques de l'APA</p>
Flexibilité clinique	
<p>Relever les exceptions cliniques (connues ou attendues) aux recommandations</p> <p>Trouver les exceptions non cliniques</p> <p>Discuter des préférences des patients</p>	<p>AHCPR Attribut jugé le plus important</p>
Clarté	
Langage non ambigu	
Processus multidisciplinaire	
<p>Participation des représentants des groupes affectés, pour fournir des évidences ou points de vue, pour reviser la version brouillon des directives)</p> <p>Développement groupe multidisciplinaire Tenir compte des biais et conflits d'intérêts</p>	<p>OMS</p>

Balancer biais, conflits potentiels	Fait signer formulaire de déclaration de conflit d'intérêts Sources de financement et biais potentiels
Méthodes pour tenir compte de points de vue autres que ceux du comité	
Plan de révision	
Préciser quand le guide sera révisé de façon à garantir la révision advenant nouvelles preuves cliniques ou changement consensus	V Validité et ratio coût-bénéfice devraient être scrutés sur une base périodique (environ aux trois ans) AHCPR L'applicabilité des nouveaux développements peut requérir révision du guide de pratique clinique
Documentation	
Description méticuleuse des : procédures utilisées; participants impliqués; preuves utilisées; hypothèses et argumentaires acceptés; méthodes analytiques employées.	
Diffusion (ajout Cluzeau)	
Comment les guides seront distribués aux utilisateurs	AMC Stratégies adaptées pour que patients, médecins et autres s'impliquent volontairement lors de la diffusion et adoption des guides
Implantation (ajout Cluzeau)	
Stratégies pour implanter les guides	V Faisabilité : politiques et implications administratives de l'utilisation
Évaluation (ajout Graham)	
Comment le guide sera évalué une fois implanté. Ex. : santé économique, satisfaction des patients et des pourvoyeurs, résultats (<i>outcomes</i>)	
<p>AHCPR : Agency for Health Care Policy and Research AMC : Association médicale canadienne (1994) APA: American psychological Association GRADE : The Grading of Recommendations Assessment, Development and Evaluation Working Group NHMRC : National Health and Medical Research Council (1999) OMS : Organisation mondiale de la santé V: Vlayen <i>et al.</i> 2005 Cluzeau <i>et al.</i> (1999) Graham et Harrison (2005)</p>	

Annexe 2.2 – Les instruments d'évaluation des guides

La première section situe les quatre instruments clés qui ont été développés. La seconde présente les attributs et les critères utilisés pour développer ces quatre instruments ainsi que quatre autres propositions.

1. Historique des principaux instruments développés pour évaluer les guides

Institute of Medicine (1990)	Mise au point d'un instrument provisoire, adopté par l' <i>Agency for Health Care Policy and Research</i> Structure : Huit attributs portant sur contenu (validité, fiabilité, applicabilité, flexibilité clinique) et le processus de développement (clarté, multidisciplinarité, révision, documentation) couvert par 142 items. Jugé trop lourd et peu pratique à appliquer. La plupart des instruments développés par la suite s'en sont inspirés : Cluzeau <i>et al.</i> (1999); Hayward <i>et al.</i> (1995); Mendelson (1995); Savoie <i>et al.</i> (1996); SIGN - Scottish Intercollegiate Guidelines Network (1995); Shaneyfelt <i>et al.</i> (1999); Ward and Grieco (1996).
Cluzeau <i>et al.</i> (1999)	Développé à la demande de l'IOM. Structure : 3 dimensions, 37 items. Évalué comme étant valide et fiable. Utilisé par le National Health Service Executive au Royaume-Uni. Est à la base de trois autres instruments ¹³¹ : Hutchison <i>et al.</i> (2003); Helou et Ollenschlager (1998); AGREE <i>et al.</i> (2003). Ce dernier est le seul des trois à avoir été validé.
The AGREE collaboration (2003)	Version simplifiée du Cluzeau. Structure : 6 domaines, 23 items. Développé et testé (valide et fiable) au niveau international. 1 ^{ère} phase : 11 pays participants ¹³² (10 européens+Canada). Diffusé, promu et évalué par 18 pays (7 états européens se sont joints). Promu dans 6 pays de l'Europe de l'Est ¹³³ . Utilisé par des agences nationales (ex. Royaume-Uni, France, Allemagne, Écosse) et l'OMS.
Shaneyfelt <i>et al.</i> (1999)	Élaboré à partir de la méthodologie proposée par l'American Medical Association, l'Institute of Medicine et le Canadian Medical Association; validité confirmée. Structure : 3 regroupements, 25 items.

2. Résumé des attributs, dimensions et critères proposés par les différents auteurs

2.1 Institute of Medicine (Field and Lohr 1992)

Attribute	Explanation
Validity	Practice guidelines are valid if, when followed, they lead to the health and cost outcomes projected for them. A prospective assessment of validity will consider the substance and quality of the evidence cited, the means used to evaluate the evidence, and the relationship between the evidence and recommendations.
Strenght of evidence	Practice guidelines should be accompanied by descriptions of the strength of the evidence and the expert judgment behind them.
Estimated outcomes	Practice guidelines should be accompanied by estimates of the health and cost outcomes expected from the interventions in question, compared with alternative practices. Assessments of relevant health outcomes will consider patient perceptions and preferences.
Reliability/	Practice guidelines are reproducible and reliable (1) if—given the same evidence and methods for

¹³¹ Vlayen *et al.* (2005).

¹³² Deux autres pays sont ajoutés dans le manuel de formation (AGREE 2003) : E.-U. et Nouvelle-Zélande.

¹³³ Hannes *et al.* (2005), AGREE (2003).

Attribute	Explanation
reproducibility	guidelines development—another set of experts produces essentially the same statements and (2) if—given the same clinical circumstances- the guidelines are interpreted and applied consistently by practitioners (or other appropriate parties).
Clinical applicability	Practice guidelines should be as inclusive of appropriately defined patient populations as evidence and expert judgment permit, and they should explicitly state the population(s) to which statements apply.
Clinical flexibility	Practice guidelines should identify the specifically known or generally expected exceptions to their recommendations and discuss how patient preferences are to be identified and considered.
Clarity	Practice guidelines must use unambiguous language, define terms precisely, and use logical and easy-to-follow modes of presentation.
Multidisciplinary process	Practice guidelines must be developed by a process that includes participation by representatives of key affected groups. Participation may include serving on panels that develop guidelines, providing evidence and viewpoints to the panels, and reviewing draft guidelines.
Scheduled review	Practice guidelines must include statements about when they should be reviewed to determine whether revisions are warranted, given new clinical evidence or professional consensus (or the lack of it).
Documentation	The procedures followed in developing guidelines, the participants involved, the evidence used. The assumptions and rationales accepted, and the analytic methods employed must be meticulously documented and described.

2.2 Cluzeau *et al.* (1999)

1st DIMENSION : RIGOR OF DEVELOPMENT (7 groups, 20 items)

Responsibility for guideline development (3 items)

1. Is the agency responsible for the development of the guidelines clearly identified?
2. Was external funding or other support received for developing the guidelines?
3. If external funding or support was received, is there evidence that the potential biases of the funding body (or bodies) were taken into account?

Guideline development group (2 items)

4. Is there a description of the individuals (*e.g.* professionals, interest groups—including patients) who were involved in the guidelines development group?
5. If so, did the group contain representatives of all key disciplines?

Identification and interpretation of evidence (4 items)

6. Is there a description of the sources of information used to select the evidence on which the recommendations are based?
7. If so, are the sources of information adequate?
8. Is there a description of the method(s) used to interpret and assess the strength of the evidence?
9. If so, is(are) the method(s) for rating the evidence satisfactory?

Formulation des recommandations (4 items)

10. Is there a description of the methods used to formulate the recommendations?
11. If so, are the methods satisfactory?
12. Is there an indication of how the views of interested parties not on the panel were taken into account?
13. Is there an explicit link between the major recommendations and the level of supporting evidence?

Peer review (4 items)

14. Were the guidelines independently reviewed prior to their publication/release?
15. If so, is explicit information given about methods and how comments were addressed?
16. Were the guidelines piloted?
17. If the guidelines were piloted, is explicit information given about the methods used and the results adopted?

Updating (2 items)

18. Is there a mention of a date for reviewing or updating the guidelines?
19. Is the body responsible for the reviewing and updating clearly identified?

Overall assessment of development process (1 item)

20. Overall, have the potential biases of guideline development been adequately dealt with?

2nd DIMENSION : CONTEXT AND CONTENT (4 groups, 12 items)

Objectives (2 items)

21. Are the reasons for the developing the guidelines clearly stated?

22. Are the objectives of the guidelines clearly defined?

Context (3 items)

23. Is there a satisfactory description of the patients to which the guidelines are meant to apply?

24. Is there a description of the circumstances (clinical or nonclinical) in which exceptions might be made in using the guidelines?

25. Is there an explicit statement of how the patient's preferences should be taken into account in applying the guidelines?

Clarity (3 items)

26. Do the guidelines describe the condition to be detected, treated, or prevented in unambiguous terms?

27. Are the different possible options for management of the condition clearly stated in the guidelines?

28. Are the recommendations clearly presented?

Likely costs and benefits (4 items)

29. Is there an adequate description of the health benefits that are likely to be gained from the recommended management?

30. Is there an adequate description of the potential harms or risks that may occur as a result of the recommended management?

31. Is there an estimate of the costs or expenditures likely to incur from the recommended management?

32. Are the recommendations supported by the estimated benefits, harms, and costs of the intervention?

3rd DIMENSION : APPLICATION (3 groups, 5 items)

Guideline dissemination and implementation (1 item)

33. Does the guideline document suggest possible methods for dissemination and implementation?

Monitoring of guidelines/clinic audit (3 items)

34. Does the guideline document specify criteria for monitoring compliance?

35. Does the guideline document identify clear standards or targets?

36. Does the guideline document define measurable outcomes that can be monitored?

National guidelines only (1 item)

37. Does the guideline document identify key elements that need to be considered by local guideline groups?

2.3 The AGREE collaboration (2003)

Scope and purpose (3 items)

1. The overall objective(s) of the guideline is (are) specifically described.

2. The clinical question(s) covered by the guideline is (are) specifically described.

3. The patients to whom the guideline is meant to apply are specifically described.

Stakeholder involvement (4 items)

4. The guideline development group includes individuals from all the relevant professional groups.

5. The patients' views and preferences have been sought.

6. The target users of the guideline are clearly defined.

7. The guideline has been piloted among target users.

Rigor of development (7 items)

8. Systematic methods were used to search for evidence.

9. The criteria for selecting the evidence are clearly described.

10. The methods for formulating the recommendations are clearly described.

11. The health benefits, side effects and risks have been considered in formulating the recommendations.

12. There is an explicit link between the recommendations and the supporting evidence.

13. The guideline has been externally reviewed by experts prior to its publication.

14. A procedure for updating the guideline is provided.

Clarity of presentation (4 items)

15. The recommendations are specific and unambiguous.

16. The different options for management of the condition are clearly presented.

17. Key recommendations are easily identifiable.

18. The guideline is supported with tools for application.

Applicability (3 items)

19. The potential organizational barriers in applying the recommendations have been discussed.

20. The potential cost implications of applying the recommendations have been considered.

21. The guidelines present key review criteria for monitoring and/or audit purposes.

Editorial independence (2 items)

22. The guideline is editorially independent from the funding body.

23. Conflicts of interest of guideline development members have been recorded.

2.4 Shaneyfelt *et al.* (1999)

Standards on guideline format and development (10 items)

1. Purpose of the guideline is specified.
 2. Rationale and importance of the guideline are explained.
 3. The participants in the guideline development process and their areas of expertise are specified.
 4. Targeted health problem or technology is clearly defined.
 5. Targeted patient population is specified.
 6. Intended audience or users of the guideline are specified.
 7. The principal preventive, diagnostic, or therapeutic options available to clinicians and patients are specified.
 8. The health outcomes are specified.
 9. The method by which the guideline underwent external review is specified.
 10. An expiration date or date of scheduled review is specified.
-

Identification and summary evidence (10 items)

11. Method of identifying scientific evidence is specified.
 12. Time period from which evidence is reviewed is specified.
 13. The evidence used is identified by citation and referenced.
 14. Method of data extraction is specified.
 15. Method for grading or classifying the scientific evidence is specified.
 16. Formal methods of combining evidence or expert opinion are used and described.
 17. Benefits and harms of specific health practices are specified.
 18. Benefits and harms are quantified.
 19. The effect on health care costs from specific health practices is specified.
 20. Costs are quantified.
-

Formulation of recommendations (5 items)

21. The role of value judgments used by the guideline developers in making recommendations is discussed.
 22. The role of patient preferences is discussed.
 23. Recommendations are specific and apply to the stated goals of the guideline.
 24. Recommendations are graded according to the strength of the evidence.
 25. Flexibility in the recommendations is specified.
-

2.5 Ward et Grieco (1996)

Développé à la demande du National Health and Medical Research Council (NHMRC)

Applicability

1. State health problem.
 2. Describes patient population.
 3. Defines proposed intervention and any restrictions on providers and patients.
-

Validity

4. Lists or mentions expected health outcomes.
 5. States costings associated with the intervention or comparison of costing with other interventions.
 6. Describes method used to reach consensus.
 7. Contains discussion about potential conflict with other guidelines or controversies.
 8. States methods used to identify evidence admitted to guidelines process.
 9. States methods used to synthesize the evidence.
-

Reproducibility

10. References sources of information used to develop guidelines.
 11. States document was subjected to review by experts or outside panels.
-

Clinical flexibility

12. Provides information about specifically known, or generally accepted, exceptions.
-

Clarity

13. Is unambiguous.
 14. Uses clear headings, indexes lists and/or flowcharts.
 15. Permits easy retrieval of major recommendations.
-

Multidisciplinary

16. Lists participants who developed guidelines.
-

Documentation

17. Has a date of publication and/or version or edition.
-

Scheduled review

18. Has a specific review date.
-

2.6 Scottish Intercollegiate Guidelines Network (SIGN 2008)

Organisation of guideline development

1. Ex. SIGN council, strategy group, guideline programmes advisory group et specialty subgroups, methodology development group
2. Funding (diversity of sources).
3. Timescale : work organization and planning deadlines.
4. Influence of financial and other interests : it is not possible to completely eliminate any possible bias from this source, nor even to quantify the degree of bias with any certainty.

Selection of guideline topics

5. Selection of a specific healthcare need, appropriate to the development of the guideline.
6. Robust evidence of effective practice on which to base guideline recommendations.
7. Criteria for selection of topics : on the basis of the burden of disease, the existence of variation in practice, and the potential to improve outcome.
8. Topic selection process : specialty subgroups consider all new proposals, prioritise them using a suitability screening and scoring tool. Final approval belongs to the SIGN Council.
9. Application procedure: suggestion is discussed by the SIGN Senior Management Team. If the proposed topic is an appropriate one, the proposer is asked to complete a second, more detailed, application form. A check is made to see if any other good quality guidelines have been produced on the subject. Then, a search for existing systematic reviews is carried out.

Updating published guidelines

10. Scheduled updates : Comity : reviewing guidelines after a period of three years after publication and then after every period of Limit : evidence.
11. Update proposals.
12. Selective update procedure.
13. Withdrawing guidelines.
14. Monitoring and interim updates.

Involving patients and their representatives

15. Patient involvement in guideline development.
16. Identifying patients' views.
17. Recruitment of patients to guideline development groups.
18. Role of patient representatives on guideline development groups.
19. Support for patient representatives on guideline development groups.
20. Wider consultation with patients and carers.

The guideline development group

21. Composition of the guideline development group.
22. Responsibilities of development group members.

Systematic literature review

23. Addressing patient issues in the literature search.
24. Using existing guidelines.

Defining key questions

25. Identifying and selecting the evidence.
26. Evaluating the evidence.

Forming guideline recommendations

27. Synthesising the evidence.
28. Considered judgment.
29. Levels of evidence and grades of recommendation.
30. Resource implications.
31. Current areas for development.

Consultation and peer review

32. National open meeting.
33. Peer review.

Presentation and dissemination

34. Content and presentation of the guideline.
35. Recommendations for research.
36. Quick reference guides and key messages.
37. Electronic publishing.
38. Information for patients.
39. Dissemination.
40. Links with audit.

Implementation

41. Getting guidelines into practice.
42. Identifying barriers to implementation.

-
43. Implementation initiatives.
 44. Practical steps.
 45. Monitoring implementation.
-

2.7 Grilli *et al.* (2000)

'Description of professionals involved in developing the guideline'

Yes : if there is a description of the type of professionals and other stakeholders involved in the development process.
Partially : if only a list of names with institutional affiliation was provided.
No : only names were reported, without further information.

'Description of the sources of information used to retrieve the relevant evidence'

Yes : If it was explicitly stated that searches were undertaken, at least through MEDLINE.
No : If no information was reported.

'Explicit grading of the evidence in support of the main recommendations'

Yes : If any form of explicit grading of the quality of the supporting evidence was reported.
No : If otherwise.

2.8 Reed *et al.* (2002). American Psychological Association

Treatment guideline checklist : yes, somewhat, no, unable to assess or n/a.

Treatment efficacy

1. Is there a broad, careful, and critical selection of the empirical literature? Some considerations:
Search strategy / Inclusion-exclusion criteria / Other sources of information.
 2. Is the methodological rigor of the selected studies identified and appropriately weighted?
Clinical opinion / Expert consensus / Systematic clinical observation / Quasi-experimental studies/ uncontrolled trials /
Randomized clinical/comparative outcomes trials.
 3. Does the document indicate whether factors affecting sample characteristics (e.g. participant selection and attrition) were
addressed in the underlying studies?
 4. Are the guideline recommendations adequately supported by the data? For example:
Alternative hypotheses are considered / Patient/treatment matching is considered.
 5. Do guidelines specify and consider a range of outcomes, such as:
Quality of life / Life functioning / Long-term consequences / Indirect consequences / Patient satisfaction / Iatrogenic / side effects
/ Clinical significance.
 6. Are other threats to reliability and internal validity adequately addressed?
-

Utility

7. Do the guidelines address patient variables that may influence the clinical utility of the intervention? These might include:
Complexity and idiosyncrasy of patients' clinical presentations including severity, comorbidity, and external stressors / Culturally
relevant research and expertise / Gender (social characteristic) and sex (biological characteristic) / Age and developmental level /
Other relevant patient characteristics.
 8. Do the guidelines address provider characteristics affecting the utility of the treatment or procedure? These might include :
Training, skill, and experience / Interactions between the healthcare professional's and the patient's characteristics (e.g.,
language, ethnicity, background, sexual orientation, and sex/gender).
 9. Do the guidelines address how differences in implementation across healthcare settings may affect the utility of the treatment
or procedure?
 10. Do the guidelines address the degree to which variability in treatment protocol implementation affects outcomes?
 11. Do the guidelines address how acceptable the intervention is to the patient?
Provide information and maximize patient choice among comparable interventions.
Consider the patient's willingness and ability to participate in recommended procedures.
 12. Do the guidelines address how feasible it is for health care professionals to deliver the intervention (e.g. training, theoretical
orientation, values and beliefs, resources, costs, administrative support)?
 13. If costs are considered in making recommendations, are cost and efficacy issues addressed independently?
 14. Are direct, indirect, short-term, long-term, and non-monetary costs to the patient, professional, and healthcare system
addressed?
 15. Are costs of withholding treatment addressed?
 16. Are threats to external validity and generalizability considered?
-

Guideline development process

17. Does the document adequately identify the guideline development panel members?
18. Is a wide range of relevant expertise or experience represented on the panel? This might include:
Service delivery / Scientific methodology / Diverse communities / Patient/consumer advocacy.

A broad range of relevant disciplines.

17. Are potential, actual, and apparent conflicts of interest disclosed?

18. Is the panel process adequately described?

Selection criteria for panelists / Panel procedures and deliberations / Dissenting opinions considered and noted / Independent review process / Drafts widely distributed for review.

19. Is the purpose and rationale for the guidelines clear?

Need / Intended audience / Goals other than improving patient care, such as rationing services, limiting legal liability, and containing costs.

20. Does the document specify a time frame for review or revision of the guidelines?

Attributes

21. Is the guideline document consistent with APA's definition of guidelines? For example:

Aspirational language / Educational value / Importance of professional judgment.

22. Is the distinction between standards and guidelines clarified in a statement consistent with APA's definition?

23. Are terms defined as needed and appropriate?

24. Is an adequate diagnostic or classification system identified?

25. Are the guidelines consistent with the current APA Ethics?

26. Are the guidelines consistent with other APA policy?

27. Is there a statement that federal and state laws supercede treatment guidelines?

28. Is the scope of the guidelines adequately delineated? Target condition or problems / Patient population(s) / Clinical interventions / Applicable professional and practice setting(s).

29. Does each guideline statement provide a specific recommendation regarding clinical assessment or treatment?

30. Do the guidelines demonstrate respect for human rights and dignity?

31. Are the guidelines internally consistent?

32. Is the language clear, succinct, and unambiguous?

33. Are line numbers included in reviewdrafts?

Annexe 2.3 - Tableau compilatif des dimensions, critères ou items proposés pour évaluer les guides développés

Le tableau ci-après compile les divers critères proposés par les auteurs de guides d'évaluation. L'unité de référence choisie est celle de Cluzeau *et al.* (1999). Les variantes, les ajouts ou précisions proposés par d'autres ou des commentaires ou explications les accompagnant y sont colligées (indiqué par une lettre). Le but n'est pas de comparer les outils mais d'exposer l'ensemble des propositions. Des commentaires issus de lectures diverses sont adjoints.

A	AGREE : version 2001 : Instrument; version 2002 : Grille d'évaluation de la qualité des recommandations pour la pratique clinique (6 domaines, 23 items); version 2003. Instrument training manual.
AMC	AMC : Association médicale canadienne, 1994, Principes directeurs concernant les guides de pratique clinique au Canada.
IOM	IOM : instrument de l'IOM (Field et Lohr 1990); Field et Lohr (1992) : Guidelines for Clinical Practice, from development to use (8 attributs).
G	Grilli et al. (2000). Practice guidelines developed by specialty society - the needs for a critical appraisal.
Shi	Shiffman et al. (2003). Standardized reporting of clinical Practice guidelines : a proposal from the conference on guideline standardization.
Sha	Shaneyfelt et al. (1999) : Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature.
W	Ward et Grieco (1996). Why we need guidelines for guidelines: a study of the quality of clinical practice guidelines in Australia.
Gra	Graham et al. (2000). A comparison of clinical practice guideline appraisal instruments. A évalué 13 instruments (11 dimensions/ attributs, 44 items).
V	Vluyen et al. (2005). A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. A évalué les 13 instruments de Graham <i>et al.</i> plus 11 autres. (10 dimensions, 50 items).

Critères proposés par Cluzeau <i>et al.</i> (1999)	Retenu par	Explications, ajouts proposés ou prescriptions	Commentaires, critiques, réserves
RIGUEUR DU DÉVELOPPEMENT		Attributs qui visent à augmenter la validité et la reproductibilité.	
1.1 Organisation responsable du développement		Qui développe, y a-t-il des conflits d'intérêts?	
1. L'organisation est-elle clairement identifiée?	Shi, Gra, V	Crédibilité de l'agence qui développe?	Les guides nationaux supérieurs aux locaux (Cluzeau <i>et al.</i> 1999). Information difficile à obtenir (Burgers <i>et al.</i> 2003).
1a Endossement officiel par une organisation?	Gra, V		Item dont l'importance est généralement sous-estimée (Gra)
2. Financement : des fonds externes ont-ils été reçus pour développer le guide?	Shi	Identifier les sources de financement et les commanditaires.	
2a Décrire son rôle	Shi		
3. Si oui, les biais potentiels ont-ils été pris en compte, y a-t-il indépendance éditoriale?	A, Shi, Gra, V	Décrire le rôle. Déclarer les conflits potentiels d'intérêts (Shi). Énoncé spécifiant que l'organisation subventionnaire n'a pas influencé les recommandations (A).	
1.2 Groupe qui développe			La perception de la crédibilité du G est fortement influencée par qui développe et qui adopte (Gra).

4. Décrit-on les individus (ex. professionnels, représentants) impliqués?	G, Gra, V, W, Shi, Sha, IOM	Noms, affiliations, qualifications, conflits d'intérêts potentiels (Shi, IOM). Domaine d'expertise (Sha). Y a-t-il des représentants des patients? (V) Ex. donnés par l'IOM : clinicien, spécialiste, généraliste, infirmier, praticien non-médecin (ex. optométriste), expert en recherche et méthode d'analyse, expert en méthode de jugement de groupe, utilisateur, payeur, décideur, commanditaire, patient potentiel, expert en assurance qualité, etc.	Les développeurs ont souvent peu d'expérience (Cluzeau <i>et al.</i> 1999). L'inexpérience est un facteur de variation des guides (Shi).
4a Qui devrait présider le groupe? Spécialiste? Md?	Gra, V, IOM	Méthode utilisée pour solliciter les avis et établir un consensus est-elle décrite? Explique-t-on comment on a choisi le président du groupe (IOM)?	On sait que cet élément est important, mais il n'y a pas de consensus sur la meilleure façon de faire (cf. chap. 2).
4b Conflit d'intérêts actuel ou potentiel?	A, IOM	Les conflits d'intérêts sont-ils notés? Y a-t-il un formulaire de déclaration de conflits d'intérêts? (A) Ceux qui ont des conflits ne devraient pas être éligibles. OMS fait signer un formulaire de déclaration de conflit d'intérêts. Les biais potentiels sont-ils discutés? A-t-on pris des mesures appropriées? (IOM)	
5. Le groupe comporte-t-il des représentants de toutes les disciplines clés?	A, IOM, Gra, V	Cet item peut concerner le comité directeur, l'équipe de recherche impliquée dans le triage et l'évaluation des données probantes, le groupe qui formulera les recommandations. Exclut les réviseurs externes du guide. Le G doit fournir les informations sur la composition, l'affiliation, la discipline et l'expertise des membres (A). Tous les groupes professionnels susceptibles d'utiliser le G doivent être représentés (AT. manuel).	La multidisciplinarité donne de meilleurs guides (Cluzeau <i>et al.</i> 1999). Elle améliore la flexibilité et l'ancrage, la probabilité de s'adresser aux bons problèmes; elle favorise l'implication et l'appropriation. IOM : l'aborde en termes de processus plutôt que de description.
5a Rétroaction : le groupe de développement a-t-il obtenu des réponses des utilisateurs potentiels?	Gra IOM	Les méthodes utilisées pour solliciter les commentaires externes et pour les présenter sont décrites et elles sont adéquates. (IOM). Inclut un panel de réviseurs externes et des audiences publiques.	
1.3 Identification et interprétation des données probantes (DP)	IOM	La validité du G est liée à la qualité des DP, aux moyens utilisés pour les évaluer et pour établir des liens avec les recommandations (IOM 1992). Le G doit indiquer jusqu'à quel point il y a consensus sur le plan quantitatif (ex. en pourcentage). IOM suggère que le comité conduise sa propre révision plutôt que de la confier à des consultants externes.	
6. Description des sources d'information utilisées pour choisir les DP à la base des recommandations?	A, G, IOM, W, Sha, Gra, V	Base de données électroniques (ex. MEDLINE, EMBASE, CINAHL), base de données de revue systématique (ex. the Cochrane Library), recherche manuelle dans les revues, comptes-rendus de conférence, etc. (IOM).	
6a Stratégies de recherche de littérature.	Gra, V, IOM, Shi	Le G doit décrire la méthode utilisée pour effectuer une recherche de littérature, incluant les dates et les bases utilisées (Shi).	
6b Utilisation de méthodes systématiques pour rechercher les DP.	A	Une recherche qui ratisse un grand nombre de sources est mieux à même d'identifier les données probantes pertinentes. La description de la méthode devrait inclure la liste des mots clés. La recherche doit couvrir chaque question clé.	Critique adressée à ce type de revue : des critères sont souvent utilisés pour limiter l'ampleur de la recherche (ex. juste anglais, années récentes) et demeurer systématique.
7. Sources d'information utilisées adéquates?	IOM		

7 a. Comment a-t-on obtenu DP?	Gra, V, IOM Sha	Le G doit décrire clairement les critères de sélection des DP. Il faut équilibrer spécificité et sensibilité. Doit définir des critères d'inclusion et d'exclusion (ex. type d'étude, taille, date, lieu géographique). Les raisons pour exclusion des DP doivent être précisées. (A et AT. manual)	
8. Décrit-on la ou les méthodes utilisées pour interpréter et évaluer la force des DP?	Sha, Shi, W, IOM	Il faut décrire les critères utilisés pour coter la qualité des DP et le système utilisé pour évaluer la force des recommandations (Shi, Gra, V).	Les deux fondements des R sont : la qualité des DP et l'importance des bénéfices (ou dommages) attendus (Shi).
9. Si oui, la ou les méthodes sont-elles satisfaisantes?	Sha, Shi, IOM	IOM détaille des méthodes qui pourraient s'appliquer (ex. synthèse formelle, techniques basées sur les jugements d'experts, de consensus).	
9a Décrit-on la méthode pour synthétiser les DP?	Shi, Sha, IOM, Gra, V	Comment utilise-t-on les DP pour développer les recommandations (ex. tableaux, méta-analyse, analyse de décision...)? (Shi, Sha)	
1.4 Formulation des R			
10. Décrit-on les méthodes utilisées pour formuler les R?	A, Shi, Gra, V	Peut inclure : système de vote, technique formelle de consensus (ex. Delphi, techniques Glaser, Rand). Les domaines de désaccords et la façon de les résoudre doivent être précisés (A, Gra, V). On reconnaît que les DP doivent être extrapolées et « traduites » pour mener à des R, ce qui implique généralement un processus de consensus. Le passage des DP aux R doit être transparent. (AT. manual)	
11. Si oui, les méthodes sont-elles satisfaisantes?			
12. Indique-t-on comment les points de vue des personnes intéressées sont prises en compte?		IOM présente cet item comme faisant partie du processus multidisciplinaire.	
12a Le rôle des jugements de valeur utilisés par les développeurs lors de la discussion des R.	Sha	Peu de G (6 %) décrivent les valeurs utilisées lorsqu'on juge la désirabilité des résultats et lors de l'élaboration des R. Les R ne sont pas juste affaire de DP et de revue systématique : il faut soupeser les avantages et les désavantages attendus. Cela dépend aussi des perspectives. Elles influencent le regard porté sur les DP.	
13. Le lien entre les principales R et le niveau des DP est-il explicite?	A, Sha, Shi, Gra, V, IOM	Chaque R devrait être accompagnée de la liste de références sur lesquelles elle se fonde ou l'absence de DP claires devrait être clairement indiquée. (AT. manual).	
13a Les R sont-elles consistantes entre elles et avec DP les supportant?	Gra, V	IOM : aborde la question de consistance sous l'angle de la logique dans la section clarté.	IOM aborde les R sous l'angle de la clarté: consistance, compréhension, timing.
13b Les R sont-elles spécifiques; s'appliquent-elles aux buts énoncés dans G?	Sha		Cet item est nouveau car il demande de porter un jugement en dehors des DP.
13c Différencie-t-on les R majeures des autres?	V, IOM		
	IOM Shi	L'IOM suggère qu'une étape de révision par un externe peut aider à évaluer la fiabilité et la validité. (IOM) Comment ceux qui ont développé le G ont procédé pour le réviser ou le tester avant sa diffusion. (Shi)	On parle ici des guidelines, qui ont un sens plus large que celui de R. Shi : pose une seule question qui recouvre les items ci-dessous.
1.5 Révision pré-diffusion			

14. Les directives ont-elles été revues par un processus indépendant avant leur publication/diffusion?	A, W, Gra, V, IOM	Il s'agit de réviseurs qui n'ont pas participé au développement. Devrait inclure des experts en méthodologie. Peut inclure des représentants des patients. (A) Explique-t-on pourquoi il n'y a pas eu de revue indépendante? (IOM)	
15. Si oui, l'information est-elle explicite quant aux méthodes et moyens utilisés?	Sha, A	La méthode utilisée pour procéder à une évaluation externe devrait être décrite. Devrait inclure la liste des réviseurs et leur affiliation. (A)	
16. Y a-t-il eu des essais pilotes?	A, Gra, V, IOM	Une directive devrait avoir été pré-testée auprès des utilisateurs visés avant sa publication. Le processus devrait être documenté (A). Cela permet aussi d'obtenir des rétroactions quant à la faisabilité et l'acceptabilité des R. (A T. manual). Il peut s'agir de projet pilote ou de pré-test. (Gra, V) L'absence de pré-tests est-elle expliquée? (IOM)	Les pré-tests aident aussi à évaluer la fiabilité et la validité. (IOM)
17. Si oui, explicite-t-on les méthodes et résultats obtenus?	Shi	Les processus devraient être documentés. (A)	
1.6 Mise à jour du G			
18. Fait-on mention d'une date de révision/mise à jour?	Sha, Shi, Gra, V, W, IOM	Date d'émission, d'expiration, agenda pour la révision. (Gra, V) Ex. d'ici trois ans post publication. Il peut y avoir des revues non prévues en sus des DP publiées ou face à un problème d'adhésion professionnelle. Une organisation développant des G devrait prévoir un processus annuel pour examiner ce qui aurait besoin d'une révision, cédulée a priori ou non. (IOM)	Pour déterminer si une révision est nécessaire ou pas, cela nécessite aussi la mise en place d'un processus dont les responsables doivent être crédibles et imputables. (IOM)
18a. Procédure de mise à jour?	A, IOM	Le G doit inclure un énoncé clair sur la procédure de mise à jour. Cela peut inclure la date, mais aussi les responsables et la méthode prévue (AT. manual).	
19. L'organisation qui en sera responsable est-elle clairement identifiée?		Il pourrait aussi s'agir de charger qq'un de revoir régulièrement la littérature et d'effectuer les changements. (A)	
1.7 Évaluation du processus de développement			
20. Globalement, a-t-on bien géré les biais potentiels?		Les 12 items concernent la fiabilité, l'applicabilité, la flexibilité et la clarté.	Une étude comparative montre que les chercheurs cotent plus sévèrement cette deuxième dimension que les infirmières et les médecins et praticiens.
II. CONTEXTE ET CONTENU			
2.1 Objectifs			
21. Les raisons pour lesquelles le G a été développé sont-elles clairement énoncées?	Sha, Shi, Gra, V	Les raisons mais aussi en termes d'importance. (Sha)	
22. Les objectifs sont-ils clairement définis?	A, Sha, Shi, Gra, V	Cela concerne en fait les impacts attendus (ici sur la santé); les objectifs devraient être décrits en termes de bénéfices attendus ou de problème potentiel. (A) Suggère d'examiner les paragraphes introductifs qui décrivent les visées et les objectifs (<i>scope, purpose</i>). Parfois les raisons sont expliquées dans un document séparé, par ex. dans la proposition de développement. (AT. manual)	

2.2 Contexte

23. La description des patients à qui s'adressent les directives est-elle satisfaisante?	A, IOM, Sha, Shi, Gra, V	La population ciblée devrait être décrite précisément (ex. âge, sexe, comorbidité) (A) ainsi que les patients exclus (AT. manual). L'audience et les utilisateurs sont-ils spécifiés? (Sha) La population peut être décrite en termes de diagnostic, pathophysiologie, gravité de la condition, système de support social... (IOM)	Pourrait être remplacé par « travailleurs » ou situations. Applicabilité clinique réfère à trois aspects: 1. population de patients ciblée; 2. population de patients exclue ou non concernée; 3. notes et discussions sur toute question complexe. (IOM)
23a Les questions (cliniques) couvertes dans le G sont-elles spécifiquement décrites?	A2, IOM	Cela aide le groupe de développement à se centrer sur les DP pertinentes. Les questions, en tant que point de départ de la revue systématique doivent être aussi ciblées que possible (population ciblée, intervention, résultats attendus, comparaisons prévue, ...). (AT. manual)	
24. Y a-t-il une description des circonstances (cliniques ou non) sujettes à exception?	Sha, Shi, IOM	Ces facteurs peuvent inclure la situation/ environnement du patient, les contraintes cliniques, etc. (IOM) Critères d'exclusion. (Shi)	IOM : l'associe à la flexibilité clinique. (1992, 1990)
25. Y a-t-il un énoncé explicitant comment les préférences du patient seront prises en compte?	A, Sha, Shi, Gra, V, IOM	Il existe différentes méthodes : représentants des patients sur le groupe de développement, entrevues, revue de littérature sur les expériences des patients, expériences des membres du groupe, etc. (A) Il faut décrire la place que prennent les préférences du patient quand les R concernent de façon substantielle un élément impliquant un choix ou des valeurs personnelles. (Shi, IOM)	IOM : l'associe à la flexibilité clinique.
25a. Le G décrit les utilisateurs prévus et le cadre ou contexte d'utilisation.	A, Shi, Gra, V	L'utilisateur doit immédiatement savoir si le G est pertinent pour lui. (Ex. G sur les douleurs au dos : praticien généraliste, neurologue, orthopédiste, physiothérapeute). (A)	

2.3 Clarté

26. Le G décrit-il sans ambiguïté les conditions à détecter, à traiter ou à prévenir?	A, Sha, Shi, Gra, V	Les R doivent être spécifiques et sans ambiguïté : quoi gérer, dans quelles conditions, pour quels patients (A). Le langage utilisé a un impact majeur sur l'acceptabilité et l'applicabilité des R. Les options possibles doivent être clairement décrites. (AT. manual) Il faut décrire : la première maladie, les conditions et l'intervention, les services, la technologie (Shi). Il faut définir aussi la technologie. (Sha)	
27. Les différentes options pour gérer la condition sont-elles clairement énoncées?	A, Sha, Shi, Gra, V	Les différentes options pour sélectionner/trier, prévenir, diagnostiquer, traiter doivent être clairement présentées (A). Indiquer le type d'intervention : approche préventive alternative, diagnostic, traitement. (Shi) Spécifier la flexibilité dans les recommandations. (Sha)	

28. Les R sont-elles clairement présentées?	A, Shi	Les R clés sont-elles faciles à identifier? Doivent permettre de répondre aux principales questions cliniques formulées. Il y a différentes façons de les identifier : souligné, caractères gras, encadré, algorithme, graphique de flux (A). Le document doit être logiquement structuré avec des titres, entêtes clairs et un index. Les R peuvent inclure une structure décisionnelle et proposer un bouquet de stratégies possibles. (AT. manual) F. L distinguent le langage, la logique et la structure (8 items). (IOM)	Les items de l'IOM n'ont pas été retenus, car ils se sont avérés difficiles à objectiver.
28a <i>Format convivial</i>	IOM	Index et entêtes.	
28b <i>Termes non ambigus, définitions précises.</i>	Gra, V	Définir les termes non familiers et les termes pouvant être mésinterprétés. (Shi)	Éviter le jargon inutile et les termes indéfinis augmente la fiabilité. (IOM)
28c <i>Résumé structuré.</i>	Gra, V IOM	Devrait inclure : date de production, statut (version originale, révisée, mise à jour), sources (imprimé et électronique). (Shi)	
28d <i>Énonce clairement les actions recommandées et spécifie les circonstances.</i>	Shi		
2.4 Coûts et bénéfices			Cette estimation est vue comme un facteur de validité. (IOM)
29. Décrit-on adéquatement les bénéfices pour la santé attendus?	A, Sha, Shi, Gra, V	Les bénéfices, les effets secondaires et les risques sont-ils pris en compte dans la formulation des R? (A) Les G devraient fournir une description des effets globaux du traitement proposé comme sur la qualité/durée de vie vs des traitements alternatifs, risques ou effets secondaires. Certains G fournissent un bilan (<i>balance sheet</i>) en matière de coût, bénéfice, dommages (AT. manual). IOM distingue entre qualitatif et quantitatif ainsi qu'en terme d'espérance de vie.	
30. Y a-t-il une description adéquate des dommages et des risques potentiels?	A, Sha, Shi, Gra, V	Les bénéfices, les effets secondaires et les risques sont-ils pris en compte dans la formulation des R? (A)	
30a. <i>Les bénéfices et les dommages sont-ils quantifiés?</i>	Sha	Quand les preuves empiriques sont limitées, les effets potentiels peuvent être seulement listés. (IOM)	
31. Y a-t-il un estimé des coûts probables associés à l'application (<i>management</i>) des R?	A, Sha, Gra, V	Les R peuvent requérir des ressources additionnelles (ex. personnel spécialisé, nouvel équipement, médicament de coût élevé). L'impact potentiel des R sur les ressources devrait être discuté. (A)	
32. Les R sont-elles appuyées par des estimés sur les coûts, bénéfices, dommages (harms) associés à l'intervention.	A11, IOM	Les bénéfices, les effets secondaires et les risques sont-ils pris en compte dans la formulation des recommandations? (A)	
32a <i>Les estimés des coûts, bénéfices ou dommages sont-ils compatibles avec la force des preuves présentées?</i>	IOM		

32b Si les bénéfices sont projetés en termes d'accroissement de l'espérance de vie, le coût unitaire est-il estimé?	IOM	Cette dimension concernant l'implantation, la dissémination et les stratégies de suivi.	CPG nationaux seraient meilleurs que les locaux à cet égard. Score médian : les chercheurs cotent plus bas que les infirmières.
III. APPLICATION			
3.1 Dissémination (D) et implantation (I)		<p>L'IOM insiste sur plusieurs points :</p> <ul style="list-style-type: none"> - On accorde beaucoup trop d'attention au développement, au détriment de l'implantation et de l'évaluation. - Les décisions quant aux moyens de D (ex. journaux, conférences) sont liées à leur crédibilité et responsabilité (<i>accountability</i>). - On sous-estime l'importance du format (attrayant, accessible) auprès des praticiens et des patients (sous-financé aussi). La présentation n'est pas un simple extra mais un aspect central à l'implantation qui requiert de l'attention. - Les groupes de développement devraient avoir à l'esprit les défis et difficultés liés à l'implantation et à l'évaluation. Par ex. la tension entre des G sophistiqués, détaillés, complexes et les G rapidement traduisibles en critères de révision médicaux ou en document compréhensible pour le patient doit être abordée pendant et non après. La question d'évaluation doit être abordée aussi durant le processus de développement. La formulation des énoncés devrait être discutée pour faciliter leur implantation. 	
33. Le document suggère-t-il des méthodes de D et I?	A, Gra, V	<p>Comment le G sera-t-il distribué aux usagers?</p> <p>Prend diverses stratégies pour que les patients, les médecins et les autres participent volontairement lors de la diffusion et de l'adoption des guides (AMC). Prend différents formats. Par ex. le G complet risque d'être non digestible. Un G peut être accompagné d'exemples, d'outils éducatifs, de feuillets pour les patients, d'aide informatique (A).</p>	
Considérations sur l'implantation		Plusieurs auteurs suggèrent des ajouts.	
33a Les barrières potentielles organisationnelles sont-elles discutées?	A, Shi	<p>L'application des recommandations du G peut nécessiter des changements organisationnels. Ceux-ci peuvent être une barrière à l'utilisation des recommandations dans la pratique quotidienne. Cela doit être discuté dans le G. (AT. manual) Implications politique et administrative vs l'I : il faut vérifier la faisabilité. (Gra, V)</p>	Item rarement vu, mais important pour les cliniciens et les décideurs (<i>policy makers</i>).
Formulation des énoncés	Shi	<p>La question du format est souvent discutée, mais aucun item n'est spécifiquement proposé. La formulation des énoncés (R) devrait être discutée entre les développeurs, les disséminateurs et ceux qui implantent. La question de l'électronique devrait être discutée.</p>	Shiffman <i>et al.</i> notent que les auteurs, de par leur souci de refléter exactement les connaissances, introduisent de l'ambiguïté qui reflète leurs incertitudes. Ils utilisent des termes qui ne sont pas clairement définis et qui posent des difficultés quand il s'agit de traduire la R en décision clinique. D'où l'importance accordée à la mise en place d'un processus de discussion.
b. Les barrières potentielles sont-elles décrites?	Shi	Aussi en lien avec la formulation des énoncés.	
c Fournit-on des documents auxiliaires pour aider à l'I?	Shi		

d. Fournit-on un algorithme (si approprié), une description graphique des étapes et du processus décisionnel?

Shi

3.2 Suivi et audits

34. Le document spécifie-t-il les critères pour faire le suivi (conformité vs R; compliance)?

A

Le document présente les critères clés pour faire le suivi ou les objets des audits. Mesurer l'adhésion peut d'ailleurs encourager son utilisation. Cela nécessite de définir les critères dérivés des R. Le document devrait présenter la liste des critères sur lesquels se fera le suivi de l'I des R et les audits. Ceux-ci devraient être le plus près possible des données recueillies sur une base routinière par l'organisation. (AT. manual)

35. Le document identifie-t-il clairement les standards ou cibles à atteindre?

36. Le document définit-il des résultats qui peuvent faire l'objet d'un suivi?

Gra, V

Exemple d'évaluation post implantation : santé, économie, satisfaction des patients, des dispensateurs de soins, résultats.

36a Les usagers adhèrent-ils au CPG?

V

36b Suggère-t-on des critères pour mesurer les changements post I?

Shi

En matière de comportements.

3.3 S'il s'agit d'un guide national

37. Le document identifie-t-il les éléments clés à considérer par les groupes locaux?

AMC

Divers. Tenir compte des problèmes d'éthique à toutes les phases d'élaboration. Mais cela ne se traduit pas en critère spécifique.

Annexe 2.4 - Attributs des guides de pratique et d'évaluation des traitements de l'American Psychologists Association (2002)

Practice guidelines attributes

Respect of human rights and dignity

Need : only for areas with a clearly demonstrated and documented need

Delineation of scope in term of content, users and context

Avoidance of bias. Documenting alternative views, providing reasoning behind decisions and judgments. Educational value

Internal consistency : no parts in conflict with any other part

Flexibility; the guidelines recognize the importance of professional judgment and discretion

Basis : take into account the best available sources of current theory, research...

Feasibility

Aspirational language : avoid should an must (mandatory). Encourage, recommand, strive (aspirational)

Clarity : succinct, unambiguous

Compatibility : avec APA policy

Outline for practice Guidelines proposal

Introduction : explain the need for the proposal G and the process by which they have been developed. Rational for creating the proposed G, steps, adequate review and citations of lit

Purpose : subject matter, beneficiaries, boudaries of applicability, intended degree of specificity

Distinction between standards and Guidelines

Users : intended users

Definitions

Needs assessment : describes the impetus for the endeavor. Demonstrates the need for the Gs and their relevance to current practice

Compatibility

Practice G development process

Practice G developers

Financial support or benefit

Criteria for selection of material

Pratice G statements : a specific recommandation for prof. conduct.

Development process

Panel

Independent with expertise in the delivery of services in the subject area

Expertise in scientific methodology

Representatives of the patient community

Expertise of a broad range of relevant disciplines

Expertise on relevant issues (e.g. diversity).

Nominees for guideline panels and panel members should disclose potential, actual, and apparent conflicts of interest. Bodies with appropriate oversight authority should evaluate the conflicts of interest and take steps to eliminate or minimize them. Maintain the climate of openness and free exchange of views : 1) selection criteria and their qualifications to be described in the guidelines; 2) deliberations available for review; 3) references; 4) disagreements to be noted in the guidelines.

Guidelines panels agree on specific goals for constructing the guidelines : 1) identify the audience; 2) goals of guideline development other than improving patient care should be clearly identified.

Guideline panel define the process and methods of guideline development as carefully as possible : 1) target condition or problems; 2) patient population(s); 3) clinical interventions that will and will not be considered; 4) type of professional and the practice setting; 5) specific subsidiary goals around which literature reviews will be organized.

Specify methods and strategies for reviewing evidence.

Specify methods for evaluating the guidelines : independent evaluation of the validity, of the clinical utility, specification of a time frame for a revision.

Efficacy assessment

Guidelines should be based on broad and careful consideration of the relevant empirical literature

Take into consideration the level of methodological rigor and clinical sophistication of the research supporting the intervention

Consider clinical opinion, observation, and consensus among recognized experts representing the range of views in the field

Systematized clinical observation is weighed more heavily than unsystematized observation in evaluating treatment efficacy

The evaluation of treatment efficacy places greatest emphasis on evidence derived from sophisticated empirical methodologies, including quasi-experiments and randomized controlled experiments or their logical equivalents

Treatment conditions to which the intervention has been compared

- Consider whether the treatment gets better results than doing nothing
- Consider whether the intervention offers the patient any benefit beyond simply being in treatment
- Consider whether an intervention's results are better than the results of other interventions

Consider evidence regarding patient-treatment matching

Should specify outcomes and evidence should be provided for each outcome; list of 11 items

Utility

- The breath of patient variables that may influence the clinical utility of the intervention :
 - ✓ complexity and idiosyncrasie of patients' clinical presentation including severity, comorbidity, and external stressors
 - ✓ culturally relevant research and expertise
 - ✓ gender and sex
 - ✓ age and developmental level
 - ✓ other relevant characteristics
- Take into account differences between individual health care prof.
 - ✓ the effect of the health care prof's training, skill, and expertise on treatment outcome.
 - ✓ the effet interaction patient/health care professional's and the patient's characteristics
- Take into account information pertaining to the setting in which the setting is offered
- Take into account data on treatment robustness
- Take into account level of acceptability to the patients
 - ✓ Provide with choice among comparable interventions
- Consider patient willingness and ability to participate in recommended interventions
 - ✓ Adverse effects and benefits
- Should address the preparation of the health care prof to deliver the intervention
- Costs

Annexe 4.1 - Principaux facteurs ou déterminants du transfert (Faye et al. 2007)

Étape	Facteurs et déterminants du transfert	
Création	Caractéristiques du chercheur	La réputation, l'expérience et la crédibilité du chercheur La disponibilité et l'engagement du chercheur L'engagement du chercheur dans des réseaux
	Adéquation recherche/besoins	La pertinence, l'utilité et l'opportunité de la recherche La prise en compte des préoccupations des utilisateurs
	Ressources	Moyens (équipements) mis à la disposition du chercheur Montant des financements alloués
	Implication des utilisateurs	La participation des utilisateurs à la recherche La participation des décideurs à la recherche
	Contexte de la recherche	Politiques de soutien à la recherche Appui des organismes subventionnaires Existence d'agents et d'organismes relayeurs
Transformation	Transformation Contexte	Adaptation des connaissances aux caractéristiques, exigences, niveaux et profils des utilisateurs Clarté et accessibilité du langage Politiques de soutien à la diffusion des résultats de la recherche
Diffusion	Soutien et canaux utilisés pour le transfert	Nature, attrait, convivialité du soutien utilisé (ex. papier, fichier téléchargeable) Choix de canaux de diffusion adaptés : guides, périodiques, internet... Réunions et rencontres d'information sur les connaissances à transférer Existence d'une stratégie de diffusion
	Contexte de la diffusion des connaissances	Existence de réseaux, communautés de recherche, etc. Existence et participation d'agents et d'organismes relayeurs Politiques de soutien à la diffusion des résultats de la recherche Appui des organismes subventionnaires
Réception	Contexte	Existence d'agents de liaison/relayeurs Accès aux résultats
	Caractéristiques des utilisateurs	Intérêts à court terme Niveau d'éducation des utilisateurs Motivation des utilisateurs
Adoption	Utilité des connaissances	Utilité perçue des connaissances à transférer Correspondance entre les connaissances et les problèmes à résoudre Interactions entre producteurs et utilisateurs de connaissances
Appropriation des connaissances	Soutien reçu par les utilisateurs	Disponibilités de ressources internes (ex. personnes ressources, comités internes) Réunions et rencontres d'information
	Contexte d'appropriation	Contexte organisationnel : organisation apprenante, apprentissage organisationnel Interactions
	Caractéristiques des utilisateurs	Niveau d'éducation des utilisateurs Motivation des utilisateurs
Utilisation des connaissances	Attitudes des utilisateurs	Motivations à s'approprier et à utiliser les connaissances Perception de l'utilité des connaissances Perception de la crédibilité des résultats possibles Confiance à utiliser les connaissances Avantages escomptés
	Caractéristiques des utilisateurs	Qualifications des utilisateurs Capacité à comprendre et à utiliser les connaissances Formation reçue Expériences antérieures d'utilisation de connaissances similaires
	Contexte de l'utilisation	Soutien (ex. pairs, hiérarchie/organisation, conseillers ASP ou CSST) Ressources internes (ex. réunions et rencontres d'information, comités internes) Facteurs organisationnels (ex. charge de travail, marge de décision, climat)

Annexe 5.1 - Termes de transfert appliqués aux guides (Faye et al. 2007)

Diffusion : signifie « action de répandre ». C'est un processus à travers lequel des connaissances sont communiquées à travers certains canaux, pour une période de temps, parmi les membres d'un système social (Roy *et al.* 1995). Fondement théorique : théorie de la diffusion de Rogers (« *process by which an innovation is communicated within a social system, and adopted or rejected by its members* »¹³⁴).

Dissémination : vient du verbe *disseminare* c.-à-d. « propager, répandre ». C'est un processus planifié à travers lequel les informations et les résultats relatifs à un programme ou une initiative (ici les G) sont livrés aux acteurs clés (Communauté européenne 2005).

Mobilisation des connaissances : elle implique un certain nombre d'actions et d'opérations mentales complexes (adaptation, différenciation, intégration, combinaison, coordination), dans un contexte déterminé, en vue de transformer les connaissances. La notion de mobilisation va au-delà des termes « utiliser » ou « appliquer » (Perrenoud 1999).

Utilisation : signifie action, manière d'utiliser. Utiliser (surtout employé à partir du XIX^e siècle) : fut d'abord attesté au sens de « rendre utile », « faire servir à une cause précise ». Le verbe est devenu, par la suite, courant dans le sens de « employer ». Il est plus proche du verbe anglais « to use » que « to utilize ». En anglais, « *utilization* » est défini comme « to put to use, especially to find a profitable or practical use for. » Il a un sens plus étroit qu'avec « *to use* ». Utile est emprunté au latin *utilis*, « qui sert », « avantageux »; c'est un dérivé du verbe *uti* « se servir de ». Le mot s'applique d'abord à ce qui sert à quelque chose, à ce qui peut être avantageux pour quelqu'un. Les études sur l'utilisation visent à vérifier dans quelle mesure les connaissances transférées sont effectivement appliquées par les utilisateurs.

Appropriation : rend compte du niveau d'acquisition, d'intégration et de sédimentation des connaissances transférées par les utilisateurs réels ou potentiels, qu'ils soient des personnes physiques ou morales.

Résultat (du transfert) : mesure du niveau d'atteinte des objectifs, implicites ou explicites, visés au départ, ainsi que les changements, les prises de décision ou les formulations de politiques ayant résulté d'une action de transfert. L'étude des résultats consiste à vérifier si les effets attendus correspondent bien à ceux initialement visés. Ils peuvent se définir comme les changements induits par une intervention qui peuvent conduire à des effets directs (OCDE 2002).

Impact : la notion d'impact peut se définir comme tout « effet à long terme, positif et négatif, primaire et secondaire, induit par une intervention, directement ou non, intentionnellement ou non » (OCDE 2002). Elle déborde celle de « résultats » et recouvre les effets indirects, qu'ils soient positifs ou négatifs. On mesure donc les effets indirects ou non attendus induits par un transfert. L'évaluation des impacts revient à vérifier si la mise en œuvre effective des connaissances transférées a permis d'obtenir des effets autres que ceux initialement prévus et désirables.

Effet : un « changement escompté ou non, attribuable directement ou indirectement à une action » (OCDE 2002).

¹³⁴ Cité par Sankaran 2006.

Annexe 5.2 - Résumé des approches destinées à changer les pratiques (Grol 1997)

Focus on	Theories	Interventions	Strategy
Internal processes			Bottom up, local consensus development
Educational	Adult learning theories	Intrinsic motivation of professionals	Small group interactive learning Problem based learning
Epidemiological	Cognitive theories	Rational information seeking and decision making	Evidence-based guideline development Disseminating research findings through courses, mailing, journals
Marketing	Health promotion, innovation and social marketing theories	Attractive product adapted to needs of target audience	Needs assessment, adapting change proposals to local needs Stepwise approach Various channels for dissemination (mass media and personal)
External influences			
Behavioural	Learning theory	Controlling performance by external stimuli	Audit and feedback Reminder systems, monitoring Economic incentives, sanctions
Social interaction	Social learning and innovation theories, social influence/power theories	Social influence of significant peers/role models	Peer review in local networks Outreach visits, individual instruction Opinion leaders Influencing key people in social networks Patient mediated interventions
Organisational	Management theories, system theories	Creating structural and organisational conditions to improve care	Re-engineering care process Total quality management/continuous quality improvement approaches Team building Enhancing leadership Changing structures, task
Coercive	Economic, power, and learning theories	Control and pressure, external motivation	Regulations, laws Budgeting, contracting Licensing, accreditation Complaints/legal procedures

Annexe 5.3 - Guide Line Implementation Appraisal (Shiffman et al. 2005)

Global considerations

1. Do the organization(s) and author(s) who developed the guideline have credibility with the intended users of the G?
2. Is the patient population eligible for the guideline clearly defined?
3. Does the G document suggest possible strategies for dissemination and implementation?
4. Is the G supported with tools for application e.g., a summary document, a quick reference guide, educational tools, patients' leaflets, online resources or computer software?
5. If any guideline R are considered more important than others. Does their presentation or formatting reflect this?
6. Is it clear in what sequence the recommendations (R) should be applied?
7. Is the G internally consistent, i.e., without contradictions between R or between text, R and flowcharts, summaries, patient education materials, etc.?

For each statement

Decidability (precisely under what circumstances to do something)

8. Would the guideline's intended audience consistently determine whether each condition in the R has been satisfied? That is each and every condition described clearly enough so that reasonable practitioners would agree when the R should be applied?
9. Are all reasonable combinations of conditions accounted for, i.e., is the R comprehensive?
10. If there are more than one condition in the R, is the logical relationship among all conditions (ANDs and ORs) clear?

Executability (exactly what to do under the circumstances defined)

11. Is the recommended action (what to do) stated specifically and unambiguously? That is, would members of the intended audience execute the action in a consistent way? In situations where two or more options are offered, the executability criterion is met if the user would select an action only from the choices offered.
12. Is sufficient detail provided or referenced (about how to do it) to allow the intended audience to perform the recommended action, given their likely baseline knowledge and skills?

Effect on process of care (the degree to which the R impacts upon the usual workflow of a care setting)

13. Can the R be carried out by current non-performers without substantial increases in provider time, staff, equipment, etc.?
14. Can the R be tried without full provider commitment? For ex., buying and installing expensive equipment to comply with a R is not easily reversible.

Presentation and formatting (the degree to which the R is easily recognizable and succinct)

15. Is the R easily identifiable, e.g., summarized in a box, typed in bold, underlined, presented as an algorithm, etc.?
16. Is the R (and its discussion) concise?

Measurable outcomes (the degree to which the guideline identifies markers or endpoints to track the effects of implementation of this R)

17. Can criteria be extracted from the guideline that will permit measurement of adherence to this R? Measurement of adherence requires attention to both the actions performed and the appropriateness of the circumstances under which they are performed.
18. Can criteria be extracted from the guideline that will permit outcomes of this R to be measured?

Apparent validity (the degree to which the R reflects the intent of the developer and the strength of evidence)

- 19) Is the justification for the R stated explicitly?
- 20) Is the quality of evidence that supports the R explicitly stated?

Novelty/innovation (the degree to which the R proposes behaviors considered unconventional by clinicians or patients)

21. Can the R be performed by the guideline's intended users without the acquisition of new competence (knowledge, skills)?
22. Is the R compatible with existing attitudes and beliefs of the guideline's intended users?
23. Is the R consistent with patient expectations? In general, patients expect their concerns to be taken seriously, benefits of interventions to exceed risks, and adverse outcomes to fall within an acceptable range.

Flexibility (the degree to which a R permits interpretation and allows for alternatives in its execution)

24. Does the R specify patient or practice characteristics (clinical and non-clinical) that require (or permit) individualization?
25. Does the R consider coincident drug therapy and common co-morbid conditions?
26. Is there an explicit statement by the guideline developer regarding the strength of this R? Note: There is a difference between

quality of evidence (item 20) and stringency of a policy. Potential statements to satisfy this criterion might include “Strong R”, “Standard”, “Clinical option”, etc.

27. If patient preference is considered, does the R propose mechanisms for how it is to be incorporated?

Computability (the ease with which a R can be operationalized in an electronic information system; optional)

28. Are all patient data needed for this R available electronically in the system in which it is to be implemented?

29. Is each condition of the R defined at a level of specificity suitable for electronic implementation?

30. Is each R action defined at a level of specificity suitable for electronic implementation?

31. Is it clear by what means a R action can be executed in an electronic setting, e.g., creating a prescription, medical order, or referral, creating an electronic mail notification, or displaying a dialog box?
