

Consensus methods: review of original methods and their main alternatives used in public health

Fanny Bourrée¹, Philippe Michel^{1,2*}, Louis Rachid Salmi^{3,4}

¹ Laboratoire Santé Travail Environnement (LSTE) Université Victor Segalen - Bordeaux II : EA3672 , IFR99 , Institut de Santé Publique, d'Épidémiologie et de Développement (ISPED) , 146, rue Léo-Saignat 33076 Bordeaux, FR

² CCECQA, Comité de Coordination de l'Évaluation Clinique et de la Qualité en Aquitaine Hôpital Xavier Arnoz , Avenue du Haut Lévêque 33604 Pessac, FR

³ Épidémiologie et Biostatistique INSERM : U897 , Université Victor Segalen - Bordeaux II , Institut de Santé Publique, d'Épidémiologie et de Développement (ISPED) , 146, rue Léo-Saignat 33076 Bordeaux, FR

⁴ Service d'Information Médicale CHU Bordeaux , Groupe hospitalier Pellegrin , Place Amélie Raba-Léon 33000 Bordeaux, FR

* Correspondence should be addressed to: Philippe Michel <philippe.michel@ceccqa.asso.fr >

Abstract

Summary

Background

Consensus-based studies are increasingly used as decision-making methods, for they have lower production cost than other methods (observation, experimentation, modelling) and provide results more rapidly. The objective of this paper is to describe the principles and methods of the four main methods, Delphi, nominal group, consensus development conference and RAND/UCLA, their use as it appears in peer-reviewed publications and validation studies published in the healthcare literature.

Methods

A bibliographic search was performed in Pubmed/MEDLINE, Banque de Données Santé Publique (BDSP), The Cochrane Library, Pascal and Francis. Keywords, headings and qualifiers corresponding to a list of terms and expressions related to the consensus methods were searched in the thesauri, and used in the literature search. A search with the same terms and expressions was performed on Internet using the website Google Scholar.

Results

All methods, precisely described in the literature, are based on common basic principles such as definition of subject, selection of experts, and direct or remote interaction processes. They sometimes use quantitative assessment for ranking items. Numerous variants of these methods have been described. Few validation studies have been implemented. Not implementing these basic principles and failing to describe the methods used to reach the consensus were both frequent reasons contributing to raise suspicion regarding the validity of consensus methods.

Conclusion

When it is applied to a new domain with important consequences in terms of decision making, a consensus method should be first validated.

MESH Keywords Consensus ; Consensus Development Conferences as Topic ; Delphi Technique ; Humans ; Public Health ; Validation Studies as Topic

Author Keywords Consensus ; Méthode Delphi. Groupe nominal ; Méthodes qualitatives ; Conférence de consensus

Position du problème

La production de connaissance en santé publique est fondée sur quatre approches : l'expérimentation, l'observation (avec ou sans maîtrise des déterminants de l'intervention évaluée), la modélisation et le consensus. Un consensus est défini comme un accord général, tacite ou exprimé d'une manière formelle, parmi les membres d'un groupe [1]. Le consensus correspond à un accord sur quelque chose, mais ne signifie pas nécessairement l'accord total de tous sur tout, c'est-à-dire l'unanimité. Le consensus tend à faire cohabiter les différences, non à les éliminer. Le consensus, comme méthode de production de connaissance ou de prise de décision, met l'accent sur l'importance de l'opinion de chaque participant et permet d'exprimer un résultat qui ne fait pas l'objet d'une opposition formelle. Seules les méthodes formalisées, regroupées sous le nom de méthodes de consensus, permettent de limiter les biais issus de la confrontation des acteurs réunis pour générer ou synthétiser des connaissances.

En santé publique, ces méthodes sont le plus souvent utilisées pour prendre des décisions, générer des idées ou hiérarchiser, dans des situations où les informations fondées scientifiquement manquent ou lorsqu'il y a un surplus d'informations contradictoires [2]. Bien que

ces méthodes soient fondées sur l'implication d'un groupe de participants souvent définis comme « experts », les méthodes de consensus ne sont pas réservées à la consultation d'autorités scientifiques du domaine concerné [3–5]. L'expert est toute personne ayant une bonne connaissance pratique, politique, légale ou administrative d'un sujet précis et ayant une légitimité suffisante pour exprimer un avis représentatif du groupe d'acteurs auquel elle appartient.

Les objectifs principaux de cet article étaient d'identifier les textes fondateurs décrivant les quatre méthodes de consensus employées dans le domaine de la santé (Delphi, groupe nominal, conférence de consensus et RAND/UCLA), de décrire les domaines d'application de chacune des méthodes ainsi que les variantes utilisées au travers d'articles illustratifs présentant des applications des méthodes.

Méthode

Stratégie de recherche

Une recherche systématique a été réalisée dans les bases de données Pubmed/MEDLINE, Banque de Données Santé Publique (BDSP), The Cochrane Library, Pascal et Francis, sans restriction de dates. Ces bases de données correspondaient à la couverture thématique du sujet étudié. Les articles sélectionnés étaient publiés en français ou en anglais. Les articles ne traitant pas d'un sujet de santé publique n'ont pas été retenus. Nous avons défini des termes et expressions correspondants au sujet de notre étude : *Delphi*, « *nominal group* », « *consensus conference* », « *RAND UCLA* ». A partir des thésaurus utilisés dans chaque banque de données, nous avons déterminé les descripteurs correspondants aux termes et expressions définis dans notre recherche. Dans ces thésaurus, nous n'avons pas trouvé de descripteurs pour les expressions « *nominal group* » et « *RAND UCLA* ». L'expression « *consensus conference* » est définie comme un type d'article dans la base de données Pubmed/MEDLINE : *Consensus Development Conference*.

Les articles identifiés ont été sélectionnés si ces termes ou expressions apparaissaient dans le titre ou le résumé. Cette restriction a permis de cibler prioritairement les articles décrivant les méthodes de consensus étudiées et d'exclure les articles présentant des résultats d'études ayant utilisé une méthode de consensus sans que celle-ci soit suffisamment mise en avant dans la publication. La même recherche a été effectuée dans les différentes bases de données et les doublons ont été éliminés. Une recherche a aussi été effectuée sur Internet via le moteur de recherche Google Scholar avec les mêmes termes et expressions. Les références citées dans les articles sélectionnés ont été étudiées pour rechercher des articles similaires supplémentaires.

Critères d'inclusion

La recherche était fondée sur une sélection d'articles suivant deux types distincts de publications. Tous les articles présentant une description détaillée de la méthode et des fondements des méthodes de consensus ont été inclus (articles méthodologiques). Parmi les articles dont l'objectif principal était de présenter une application et les résultats de la méthode de consensus utilisée dans un domaine particulier (articles de méthodologie appliquée), seuls ceux dont le texte était en accès libre, c'est-à-dire les articles dont les textes étaient mis à disposition en intégralité et sans contrepartie financière, ont été étudiés et ont servi à illustrer les différents domaines d'application et les différentes variantes des méthodes étudiées. Lorsque cela était nécessaire, notamment pour la description de variantes des méthodes de consensus, les articles concernés ont été acquis même s'ils n'étaient pas en accès libre.

Analyse des articles et lecture critique de la qualité des articles

La pertinence des articles, compte tenu des objectifs, a été évaluée par un des auteurs. Celle-ci a été appréciée en fonction du niveau de description de la méthode. Un article était analysé lorsque le processus d'atteinte du consensus était suffisamment décrit pour pouvoir juger de l'écart par rapport à la méthode décrite dans les articles de référence. Compte tenu des objectifs, nous n'avons pas pris en compte les résultats proprement dits des études mais seulement la description et l'utilisation des méthodes de consensus.

Analyse des données et présentation des résultats

En accord avec les objectifs et avec la nature des données, les résultats sont présentés sous format textuel. Les articles méthodologiques ont permis de présenter les fondements et les étapes méthodologiques des méthodes de consensus. Les articles de méthodologie appliquée ont permis d'illustrer les domaines d'utilisation de ces méthodes ainsi que les variantes utilisées. Seuls quelques articles de méthodologie appliquée identifiés pour chacune des méthodes de consensus sont référencés comme exemples dans cet article.

Résultats

La méthode Delphi

Historique

La méthode Delphi tient son nom de l'oracle de Delphes qui, représenté par la Pythie, permettait aux grecs de venir interroger les divinités au travers de la prophétesse [6]. La méthode Delphi a été initiée par la RAND Corporation (une contraction de l'expression *Research and Development*), organisme californien sans but lucratif, dont la mission est d'accompagner les décideurs de différents

domaines en appliquant des outils d'analyse et de recherche. La méthode a été initialement appliquée dans le cadre du « projet de Delphes » qui visait à décrire les tendances de développement à long terme de la science et des technologies (les percées scientifiques, le contrôle des populations, les progrès de l'espace, la prévention des guerres et les systèmes d'armes) [7].

Objectifs et grands principes

L'objectif de la plupart des applications de la méthode Delphi est d'apporter l'éclairage des experts sur des zones d'incertitude, en vue d'une aide à la décision [8]. La méthode Delphi est un exercice de communication de groupe, qui permet de rassembler et de synthétiser la connaissance d'un groupe de participants géographiquement dispersés et qui ne se réunissent jamais [9]. L'interrogation individuelle des participants se fait par des questionnaires successifs, rédigés par les organisateurs de l'étude ; dans le cas de questions à réponse fermée, une échelle de cotation de l'accord peut être utilisée [10].

Les caractéristiques fondamentales de la méthode sont [11]:

- l'anonymat : les participants n'ont pas connaissance de l'identité des autres, afin d'éviter les phénomènes de dominance, d'autorité ou d'affiliation.
- l'itération avec remontées contrôlées : les experts sont sollicités à plusieurs reprises. Avant chaque tour, ils reçoivent les résultats obtenus au précédent tour afin de confronter leur avis à celui des autres participants.
- l'analyse quantifiée des réponses du groupe : les membres ayant des avis correspondants aux extrêmes (totalement en accord ou totalement en désaccord) sont invités à donner une justification supplémentaire aux autres membres du groupe.

Étapes de réalisation

La méthode suit quatre grandes étapes [12]:

- *La formulation du problème* : c'est une étape fondamentale tant pour la rédaction des questionnaires que pour le choix des experts [13]. Cette phase consiste à définir avec rigueur et précision l'objet (défini comme le problème à examiner et les grands questionnements liés à ce problème) sur lequel portera l'étude, pour éviter d'entraîner les experts dans un processus dont le thème évolue au fur et à mesure.
- *Le choix des experts* : pour éviter la mise en question future du processus, les experts sont choisis pour leur connaissance de l'objet et pour couvrir la variété des parties prenantes. Le nombre minimum de participants pour assurer la validité du résultat dépend de l'objet de l'étude : un groupe de 4 participants suffit parfois [14] mais des groupes de 10 à 15 participants sont habituellement constitués.
- *L'élaboration du questionnaire* : les questions doivent être ciblées, précises et permettre des réponses fermées. Dans ce cas, chaque participant répond à chaque question à l'aide d'une échelle de cotation allant classiquement de 1 à 9, de l'existence « d'un désaccord complet » ou d'une « absence totale de preuve » ou d'une « contre indication formelle » (coté 1) à l'existence « d'un accord complet » ou « d'une preuve formelle » ou « d'une indication formelle » (coté 9).
- *L'administration du questionnaire et le traitement des résultats* : le questionnaire de départ est administré par courrier (physique ou électronique) aux experts. Le même questionnaire servira de fil conducteur à tout l'exercice; il sera enrichi, à chaque tour, des résultats et commentaires générés par le tour précédent. Au deuxième tour de questionnaire, les experts reçoivent les résultats du premier tour et se prononcent à nouveau sur le questionnaire, en ayant connaissance des réponses des autres membres du groupe. Les participants peuvent maintenir leur réponse ou bien la modifier ; si leur nouvelle réponse dévie fortement de la moyenne du groupe, ils doivent la justifier. Ce tour peut être réitéré autant de fois que nécessaire pour obtenir une convergence des réponses ou bien d'identifier les discordances. Le retour d'information aux participants comprend les opinions consensuelles médianes et la dispersion des opinions autour de cette médiane, cette dispersion pouvant être interprétée à l'aide des justifications et commentaires recueillis auprès des experts.

Application de la méthode dans la littérature

Nous avons retrouvé 2 051 articles dans les bases de données avec le terme Delphi (descripteurs identifiés par MeSH Database *Delphi Technique* ; *Delphi Study* ; *Delphi Technic* et par Thesaurus Santé Publique version 4 *Méthode Delphi*). Parmi ces 2051, 1 986 ont été exclus car insuffisamment précis sur la méthode d'obtention du consensus ou n'étaient pas en accès libre. Parmi les 65 articles retenus, 9 étaient des articles méthodologiques [2, 6–13] et 56 des articles de méthodologie appliquée.

La méthode Delphi est utilisée dans le cadre de la production de connaissance et dans la prise de décision, dans des domaines variés (orientation des stratégies en santé publique [15], éducation liée à la santé [16], priorités de prévention [17], définition de pratiques professionnelles et leur amélioration [18], évaluation de la qualité des soins et des pratiques médicales [19]) ou dans le domaine de l'épidémiologie et la recherche clinique [20].

Parmi les 56 articles de méthodologie appliquée, 27 études utilisaient une variante de la méthode (terme utilisé : « *modified Delphi* »). Les étapes soumises à des adaptations sont principalement : le nombre de tour de questionnaires d'emblée limité à deux [21], le fait d'intégrer une rencontre des participants (comme le mini-Delphi qui propose une application en temps réel de la démarche [22]), les objectifs à atteindre (aboutir à un consensus, à la plus large gamme d'opinions possible, à la prise de décisions des intervenants malgré leurs intérêts divergents ou à la création d'idées et d'objectifs communs comme la méthode de l'Imen-Delphi [23, 24]).

La méthode du groupe nominal

Historique

La technique du groupe nominal a été créée par deux chercheurs américains, André Delbecq et Andrew Van de Ven, à la fin des années 60 [25]. La méthode a été initialement préconisée pour des problèmes nécessitant la genèse ou la hiérarchisation d'informations [26].

Objectifs et grands principes

Cette méthode a été utilisée pour analyser un problème, explorer un champ de connaissance, faire une synthèse globale d'une question, planifier et mettre en place des activités, procéder à l'évaluation d'une intervention, établir des priorités ou hiérarchiser des actions. Certains préconisent d'utiliser la technique du groupe nominal lorsqu'il est possible de réunir les participants et que l'urgence des problèmes ou le rapprochement des échéances exigent des réponses immédiates [2].

Le principe est de récolter des informations en classant une série d'items ou de questions au cours d'une réunion. Cette réunion est animée par un coordinateur [27] soit professionnel du sujet traité [25] soit sans expertise dans le domaine [28]. Les interactions se font essentiellement entre le coordinateur et chaque membre du groupe pour éviter les phénomènes de dominance. Cinq à neuf personnes participent, mais il est possible d'aller jusqu'à quinze, si le sujet de discussion n'est pas trop compliqué. S'il y a plus de participants, il est préférable de les diviser en plusieurs groupes.

Étapes de réalisation

Génération silencieuse des idées

L'animateur distribue aux participants une feuille sur laquelle la question est écrite. Après quelques minutes, chacun note sur sa feuille le plus grand nombre d'idées durant environ cinq minutes. Le travail se fait individuellement et en silence.

Énoncé des idées

L'animateur demande à chaque participant, à tour de rôle, de communiquer la première réponse inscrite sur sa feuille. Au fur et à mesure, il écrit les réponses sur un tableau. Il y a autant de tours de table que de réponses à exprimer. Son tour venu, chaque participant communique une seule idée. Si une personne considère qu'une de ses idées a déjà été exprimée par un autre participant, elle passe à la suivante. Lors de cette deuxième étape, il faut éviter les critiques et limiter les discussions et commentaires. Les idées deviennent la propriété du groupe et l'animateur numérote les idées en les écrivant.

Discussion et tour de table

Tous les énoncés étant bien en vue au tableau, les idées sont clarifiées les unes après les autres. Il s'agit de vérifier que tous les participants attribuent le même sens aux énoncés et que la logique qui sous-tend chaque opinion est bien comprise. Il appartient au groupe de clarifier les idées exprimées et non à la personne qui les a énoncées. Il est important de répartir le temps le plus également possible entre les énoncés.

Cotation et hiérarchisation des idées

La discussion de groupe est suivie de la cotation individuelle de l'importance relative des énoncés. L'objectif étant de retenir un nombre préalablement défini d'énoncés (par exemple 5), chaque participant attribue cinq points à l'énoncé jugé le plus important des cinq et un point au moins important. Les énoncés qui reçoivent le plus grand nombre de points constituent les priorités du groupe.

Compilation des résultats

L'animateur comptabilise les points obtenus par chaque énoncé. Deux étapes peuvent être ajoutées (notamment lorsque le nombre de participants est important) : une discussion des résultats du vote ou une ultime cotation pour départager des énoncés ex aequo.

Application de la méthode dans la littérature

Notre recherche a identifié 1 243 articles (terme utilisé : « *nominal group* »). Parmi eux, 57 répondaient aux critères d'inclusion et ont été identifiés comme des articles de méthodologie appliquée. Six articles supplémentaires ont été identifiés comme des articles méthodologiques présentant en détail les principes, la technique et les objectifs de la méthode [1, 2, 25–28]. Les 57 articles ont été inclus

comme travaux de méthode appliquée en utilisant l'expression « *modified nominal group* ». Les domaines d'application étaient la définition des priorités en recherche [29, 30], la formation médicale [31] et l'évaluation des pratiques [32].

Des variantes ont été mises en évidence dans 29 des 57 articles. Par exemple, la méthode a été limitée aux étapes de discussion et tour de table, de vote et de hiérarchisation des idées et de compilation des résultats (les étapes précédentes sont conduites au préalable par le coordinateur à distance) ou encore, en amont de la réunion, les participants ont incorporé le détail de la revue de la littérature pour étoffer la discussion sur les différents points [33, 34].

La conférence de consensus

Historique

La méthode de conférence de consensus a été utilisée dans le secteur de la santé aux Etats Unis au cours des années 1970 et développée par les National Institutes of Health (NIH). En 1977, la 1^{ère} conférence de consensus des NIH portait sur le dépistage du cancer du sein. Elle appartenait à un programme de recherche de consensus répondant « au double souci du Congrès américain de mettre en place un système formalisé d'évaluation des innovations et d'autre part de voir les NIH se préoccuper du transfert vers les médecins praticiens des résultats issus de la recherche, lui donnant une responsabilité nouvelle dans la validation et la diffusion des innovations médicales. » Le principe reposait sur la réunion d'un jury sélectionné qui écoutait en public des experts présenter les données disponibles. Un bilan des acquis a été fourni par le jury qui a rédigé sur place un rapport et émis des recommandations ensuite largement diffusées. Des groupes d'experts médicaux s'attachent à définir par consensus la bonne utilisation de nouvelles thérapeutiques ou de nouveaux médicaments. Ces experts donnent des recommandations destinées à être reprises par l'ensemble de la profession pour le bon usage d'une innovation médicale [35].

Objectifs et grands principes

Une conférence de consensus vise à dégager au sein de la communauté concernée les points d'accord et de divergence relatifs à une intervention de santé, qu'il s'agisse d'une procédure diagnostique, d'une stratégie thérapeutique ou d'aspects liés à l'organisation du système de santé [36]. A l'instar de toute recommandation professionnelle, elle a pour objectifs d'établir une synthèse des connaissances et d'être une aide à la décision.

La conférence de consensus consiste en la rédaction de recommandations par un jury au terme d'une présentation publique de rapports d'experts faisant la synthèse des connaissances sur un thème précis. Le déroulement de la séance publique tient à la fois de la conférence scientifique, du débat démocratique et du modèle judiciaire.

Étapes de réalisation

Dès 1990, l'Agence Nationale pour le Développement de l'Évaluation Médicale (ANDEM) a décrit la méthode de conférence de consensus dans un guide à l'usage des professionnels [37], puis la Haute Autorité de Santé (HAS) a proposé une base méthodologique pour leur réalisation en France [38].

Le *promoteur*, souvent une société savante ou une institution publique, définit le thème et fournit les moyens financiers pour la mise en œuvre de la conférence de consensus. Le *comité d'organisation* assume la responsabilité de l'ensemble du processus et en garantit la bonne tenue. Il recrute le jury, nomme le président, prépare et organise les réunions intermédiaires, recrute les experts qualifiés, met en œuvre le plan de communication pour une couverture médiatique et organise enfin la conférence publique. Le *jury* est multidisciplinaire et multi professionnel, constitué de huit à seize personnes, choisies parmi des médecins, des chercheurs, des professionnels de santé non médecins, des méthodologistes, des représentants des domaines éthiques, économiques et législatifs, des représentants du grand public (associations de malades, médias...). Une analyse critique de la littérature leur est fournie par un *groupe bibliographique*. Le panel des *experts* est constitué de 12 à 15 personnes (scientifiques, ingénieurs, juristes, économistes, sociologues...) choisis par le comité d'organisation. Leur rôle est de répondre, pendant la conférence publique, aux questions du jury qui leur ont été communiquées à l'avance. La conférence publique est le point d'orgue du processus. Animée par un président de séance, elle se déroule en trois étapes, en présence et avec la participation du public : les experts répondent aux questions du jury ; le jury pose les réponses aux experts ; des questions complémentaires sont posées par le public. Puis le jury se réunit à huis clos pour élaborer le document final qui privilégie la recherche du consensus sur les questions en débat, il établit les recommandations de la manière la plus indépendante et la plus objective possible, en distinguant ce qui relève des faits scientifiques, de la présomption et de la pratique usuelle. Suite à la lecture publique du document final, les experts peuvent intervenir pour corriger une imprécision ou une erreur factuelle, mais sans droit à influencer les opinions exprimées.

Application de la méthode dans la littérature

La conférence de consensus est la méthode la plus utilisée en santé ; 2 238 articles ont été identifiés lors de notre recherche (Pubmed/MEDLINE : Limits ; Type of Article *Consensus Development Conference* ; *Consensus Development Conference, NIH*, descripteurs identifiés par MeSH Database : *Consensus Development Conference* ; *Consensus Development Conference, NIH* ; *NIH*

Consensus Development Conference ; Consensus Development Conferences as Topic ; Conferences, Consensus Development as Topic ; Consensus Development Conferences, NIH as Topic ; NIH Consensus Development Conferences as Topic, descripteurs identifiés par Thésaurus Santé Publique version 4 *Conférence consensus*). 135 articles répondaient aux critères d'inclusion. Seulement 3 articles étaient des articles présentant la méthode et ses fondements [35, 37, 38]. La quasi-totalité des articles présentaient les résultats de conférences de consensus, majoritairement utilisées par des institutions de soins pour définir des recommandations de pratiques professionnelles [39, 40] ou faire le point sur les connaissances du domaine [41, 42].

Des évolutions et des variations dans la mise en œuvre de conférence de consensus ont été identifiées [43–46]. Au Danemark, le Parlement danois souhaitait se démarquer du modèle américain dans le domaine de l'évaluation des choix technologiques. L'un des objectifs était de réduire le décalage entre les experts, les responsables politiques et la population générale. Il s'agissait donc de concevoir des méthodes à même de prendre en compte les préoccupations, l'expérience quotidienne et les visions des citoyens, en même temps que les compétences des experts, tout en tenant compte des contraintes propres aux politiques. Les thèmes traités avec cette variante, appelée conférence de citoyens [47], sont plus vastes, ils doivent être à la fois d'intérêt public et faire l'objet de controverses aussi bien dans la communauté scientifique que dans la société en général (ex : « pollution de l'air », « thérapie génique »...). D'autres méthodes ont vu le jour comme le PubliForum, version suisse des conférences de citoyen.

La méthode RAND/UCLA et le consensus formalisé

Historique

Développée dans les années 80 aux Etats-Unis, principalement par la RAND Corporation et l'Université de Californie à Los Angeles [48], la méthode RAND/UCLA est bien codifiée et largement utilisée en santé publique. La méthode de consensus formalisé, publiée par la HAS, en est dérivée [49]. Elle diffère de cette méthode principalement par l'incorporation d'un groupe de lecture et de la réunion du groupe de cotation à l'issue de la consultation du groupe de lecture.

Objectifs et grands principes

La méthode RAND/UCLA est une association d'éléments des méthodes Delphi et groupe nominal [1]. L'objectif de la méthode RAND/UCLA est de dégager clairement les points de concordance, fondements des futures recommandations, et les points de discordance ou d'indécision, permettant d'orienter d'éventuels travaux de recherche ultérieurs.

Le rôle des experts est prédominant puisque l'analyse, la synthèse de la littérature et la rédaction des propositions de recommandations pour chaque champ d'application sont confiées aux experts. Une cotation itérative est utilisée par tous les experts pour favoriser la pertinence, l'objectivité et l'homogénéité du classement des propositions. La cotation individuelle des propositions de recommandations a l'avantage de permettre aux avis minoritaires de s'exprimer, de favoriser la communication et l'interactivité au sein du groupe, et d'éviter l'effet de dominance d'un expert.

Étapes de réalisation

Cette méthode suit différentes étapes [50]:

- Le promoteur choisit le thème et initie la réalisation de la méthode.
- Le comité d'organisation définit le thème et les principaux champs à traiter (moins de 10) suite à une revue de littérature sur le sujet et désigne des experts
- Le groupe d'experts, lors d'une première réunion, valide le choix des différents champs à traiter et répartit les champs entre les sous-groupes d'experts.
- Chaque sous-groupe d'experts travaille de façon indépendante et effectue l'analyse, la synthèse de la littérature, rédige un texte d'argumentaire et une liste de propositions de recommandations dans le champ dont il a la responsabilité, pour l'établissement du document global qui sera ensuite soumis à l'ensemble du groupe.
- Le groupe d'experts, lors d'une seconde réunion, élabore le document global. Il s'agit d'examiner les libellés des propositions de recommandations, qui doivent être clairs, non ambigus et compréhensibles, de supprimer les redondances et de vérifier la cohérence globale et l'absence d'oublis majeurs. En principe, à cette étape, il n'y a pas de discussion sur le fond ni sur la pertinence des propositions de recommandations.

Le document global est envoyé à tous les experts et fait l'objet de cotation. Les règles de cotation sont prédéfinies et déterminent d'une part l'accord (ou le désaccord) et d'autre part le degré de convergence des avis des experts. Les réponses à chaque question ou proposition sont analysées en tenant compte d'abord de la médiane, puis de la dispersion des cotations sur l'échelle de 1 à 9. Trois zones

sont définies en fonction de la place de la médiane calculée sur toutes les notes : les cotes 1 à 3 correspondent à la zone de désaccord, les cotes 4 à 6 à la zone d'indécision et les cotes 7 à 9 à la zone d'accord. Pour chaque question, le degré de convergence des avis du groupe est apprécié par la position sur l'échelle de l'intervalle borné par les cotations minimales et maximales. L'accord (ou le désaccord) est dit fort si l'intervalle est situé à l'intérieur des bornes. Si l'intervalle empiète sur une borne, l'accord (ou le désaccord) est dit faible (intervalle de 1 à 4 ou de 6 à 8 par exemple) [51]. Dans la méthode RAND/UCLA, une valeur extrême haute et une valeur extrême basse sont écartés.

- Pour le deuxième tour de cotation, les experts ont la connaissance de l'ensemble des réponses du groupe à chaque proposition (médiane, distribution et extrêmes avec rappel de leur propre note) et de l'argumentaire de chaque sous-groupe. A chaque tour, l'expert est libre de changer d'avis ou non (selon le principe de la méthode Delphi).

- Le groupe d'experts est réuni pour la troisième fois. L'objectif est d'identifier les propositions de recommandations pour lesquelles il y a accord, désaccord ou indécision pour aboutir au texte final de recommandations. Les résultats du deuxième tour de cotation sont présentés. Les propositions pour lesquelles il y a accord faible, désaccord ou indécision du groupe sont discutées à partir de l'argumentaire du sous-groupe concerné. A l'issue de cette discussion, une troisième cotation est réalisée individuellement par chaque expert qui est libre de changer d'avis ou non. Cette dernière cotation est le fondement de la formalisation des recommandations sur laquelle le document final est élaboré. Les points de désaccord ou d'incertitude sont identifiés, permettant notamment de définir des perspectives de recherche.

- Le texte de recommandations est rédigé par le comité d'organisation à partir des résultats du dernier tour de cotation et des argumentaires.

Les structures nationales de santé (NIH aux Etats-Unis, HAS en France...) ont développé d'autres méthodes de consensus fondées sur les méthodes Delphi, groupe nominal et RAND/UCLA. La méthode de consensus formalisé [49] fait intervenir différents acteurs : le promoteur, le groupe de pilotage, le groupe de cotation et le groupe de lecture qui donne un avis sur le fond et la forme des recommandations retenues, en particulier sur leur applicabilité, leur acceptabilité et leur lisibilité.

La méthode du consensus formalisé s'effectue en quatre phases:

- Une phase de préparation ayant pour but de délimiter le thème, d'effectuer une revue de la littérature et de produire une première série d'items (groupe de pilotage).
- Une phase de cotation (groupe de cotation) : première cotation individuelle des propositions puis réunion du groupe permettant de discuter des résultats, seconde cotation individuelle selon une échelle de cotation.
- Une phase d'évaluation (groupe de lecture).
- Une phase de finalisation du texte (groupe de pilotage).

Application des méthodes dans la littérature

68 articles ont été identifiés (expression « *RAND UCLA* ») dont 19 répondant aux critères d'inclusion. 2 articles ont été identifiés comme étant des articles méthodologiques [48, 50]. La méthode RAND/UCLA est utilisée dans le domaine de la santé publique notamment pour la définition de pratiques professionnelles [52, 53] et l'évaluation de nouvelles technologies en santé [54]. 34 articles concernaient le consensus formalisé (expression « *formalized consensus* »), dont 4 répondaient aux critères d'inclusion. Un seul article était un article méthodologique [49]. La méthode de consensus formalisé est principalement utilisée pour évaluer des technologies de santé [55] et pour des recommandations de pratiques professionnelles [56].

Validité des méthodes de consensus

Le nombre d'études visant à valider les méthodes de consensus, par rapport aux autres approches d'aide à la décision, est extrêmement faible relativement à leur fréquence d'utilisation. Au début des années 1970, des études ont montré la plus grande validité d'un consensus de groupe par rapport à l'élaboration individuelle de connaissances [57]. Des travaux ont comparé les méthodes de consensus à des groupes non structurés fonctionnant sur des interactions directes. Une recherche psychométrique menée par Delbecq et Van de Ven auprès de 420 participants [2] a démontré une nette supériorité de la technique Delphi et du groupe nominal sur le groupe non structuré et une légère supériorité du groupe nominal sur Delphi par rapport à la quantité des idées exprimées et à la satisfaction éprouvée par les participants dans les processus de prise de décision en groupe. Les 420 participants étaient répartis en 20 groupes de 7 personnes pour chacune des trois techniques. Ils venaient du même milieu, une université du Middle West, et ils avaient une égale connaissance du problème qui leur était soumis. Les groupes hétérogènes étaient constitués d'étudiants, de professeurs et d'administrateurs. Les deux critères d'efficacité étaient la quantité d'idées exprimées et la satisfaction des participants.

D'autres études ont aussi mis en avant l'efficacité des méthodes Delphi et du groupe nominal [26, 58, 59]. Nous n'avons pas retrouvé d'études traitant de la validité des méthodes Delphi et groupe nominal dans le domaine de la santé à l'inverse de travaux concernant la méthode RAND/UCLA [54, 60]. Pour cette dernière, un article étudie la reproductibilité de la définition du caractère approprié de soins médicaux par des groupes de médecins appliquant de façon indépendante cette méthode [61] ; un autre met en avant les incertitudes concernant la validité de la méthode lorsqu'il s'agit de décider de la prise en charge de patients [62]. Un dernier enfin conclut que, malgré ses faiblesses, la méthode permet de définir des recommandations cliniques [63]. L'exemple de l'endoscopie digestive a été particulièrement bien étudié [64, 65].

Discussion

Toutes les méthodes sont précisément décrites dans la littérature et comprennent des fondements communs tels que la définition du sujet, la sélection des experts et un processus d'interactions interpersonnelles directes ou à distance. La cotation, pour hiérarchiser les propositions, est parfois utilisée. Il existe de nombreuses variantes aux définitions originelles de ces méthodes mais peu de travaux de validation.

Pour répondre à nos objectifs, une revue exhaustive ne paraissait pas nécessaire et pour des raisons pratiques, seuls les articles aux textes en accès libre ont été étudiés. Ceux-ci étaient suffisants pour couvrir les variantes connues et vraisemblablement les différents types d'applications des méthodes de consensus. Une seule commande a été nécessaire, concernant les articles décrivant la méthode de l'Imen-Delphi. Nos objectifs ne rendaient pas non plus utile la réalisation d'une double lecture indépendante.

Les modalités de mise en œuvre des méthodes de consensus sont précisément décrites dans les articles originaux ou des textes de référence. Les différentes étapes de réalisation des méthodes sont présentées dans le tableau 1. Même si leur utilisation doit s'adapter au contexte, aux objectifs et aux contraintes locales, leurs grands principes doivent être respectés. Or des exemples dans la littérature montrent que les méthodes peuvent être galvaudées à tel point que ces travaux n'ont de commun avec les méthodes d'origine que leur dénomination. On peut même s'interroger sur le fait que la littérature actuelle ne soit pas « l'arbre qui cache la forêt », car la plupart des résultats issus de consensus n'est pas publiée. Malgré le fait que nous n'ayons pas fait la revue exhaustive de la littérature, les quelques articles cités illustrent que les méthodes de consensus sont utilisées au prix de distorsions parfois importantes de la méthode originale, alors même qu'elles s'en prévalent directement. Ce constat pose la question de la validité des résultats.

Cette hétérogénéité des pratiques, qui contraste avec celles prévalant dans les autres méthodes d'élaboration de connaissance, entretient certainement la suspicion de la communauté scientifique sur ces méthodes. Plusieurs raisons concourent à cela : le manque de rigueur dans l'utilisation des méthodes, l'absence de description précise du processus d'obtention du consensus dans ces publications et le manque de travaux de validation. La validité de ces méthodes est en effet peu étudiée alors que leur intérêt apparaît de plus en plus grand. Proposées comme légitimes lorsque les autres approches sont inappropriées, elles pourraient être préférées aux autres approches même si ces dernières sont faisables, car elles sont moins coûteuses et plus rapides [66]. Pour favoriser leur utilisation, il serait souhaitable que des travaux de validation soient menés dans les grands domaines d'application de ces méthodes et particulièrement dans le domaine de la santé

References:

Références

1. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995 ; 311 : 376 - 80
2. Editor: Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning: A guide to nominal group and Delphi processes. Glenview Scott, Foresman ; 1975 ;
3. Martino J. Technological Forecasting for Decision Making. New York McGraw-Hill ; 1993 ;
4. Johnson DE, Meiller LR, Miller LC, Summers GF. Needs assessment: Theory and methods. Ames Iowa State University Press ; 1987 ;
5. Wely G. Problems of Selecting Experts for Delphi Exercises. *Academy of Management Journal*. 1972 ; 15 : 121 - 4
6. Kaplan A, Girshick M, Skogstad A. The Prediction of Social and Technological Events. *Public Opinion Quarterly*. 1950 ; XIV : 93 - 110
7. Gordon TJ, Helmer O. Report on a Long Range Forecasting Study. Santa Monica, California Rand Corporation ; 1964 ; Report No.: R-2982
8. Dalkey N. The Delphi Method: an experimental study of group opinion. Santa Monica Rand Corporation ; 1969 ; Report No.: RM-5888-PR
9. Adler M, Ziglio E. Gazing into the Oracle: the delphi method and its application to social policy and public health. London Jessica Kingsley Publisher ; 1996 ;
10. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast*. 1999 ; 15 : 353 - 75
11. Fink A, Koscoff J, Chassin M, Brook R. Consensus methods: characteristics and guidelines for use. *Am J Public Health*. 1984 ; 74 : 979 - 83
12. Fowles J. Handbook of futures research. Connecticut Greenwood Press ; 1978 ;
13. Linstone H, Turoff M. The Delphi Method, techniques and applications. Addison Wesley Publishing. 1975 ;
14. Brockoff K. The performance of forecasting groups in computer dialogue and face-to-face discussion dans Linstone H.A. et Turoff M. Addison-Wesley. The Delphi Method :Techniques and Applications. 1975 ;
15. Annells, DeRoche, Koch, Lewin, Lucke. A Delphi study of district nursing research priorities in Australia. *Appl Nurs Res*. 2005 ; 18 : 36 - 43
16. De Villiers MR, De Villiers PJ, Kent AP. The Delphi technique in health sciences education research. *Medical Teacher*. 2005 ; 27 : 639 - 43
17. Schaller G, Fournier P. Les priorités de prévention en santé mentale à Genève : une enquête par méthode Delphi. *Santé Publique*. 1998 ; 10 : 191 - 201
18. Cook C, Brismee JM, Sizer J, Phillip S. Subjective and objective descriptors of clinical lumbar spine instability: A Delphi study. *Manual Therapy*. 2006 ; 11 : 11 - 21
19. Richard S, Rawal S, Martin D. An ethical framework for cardiac report cards: a qualitative study. *BMC Medical Ethics*. 2005 ; 6 : 3 -
20. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M. Global prevalence of dementia: a Delphi consensus study. *The Lancet*. 2005 ; 366 : 2112 - 7
21. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *J Adv Nurs*. 2006 ; 53 : 205 - 12

- 22. Tugwell P, Sitthi-Amorn C, Hatcher-Roberts J, Neufeld V, Makara P, Munoz F. Health Research Profile to assess the capacity of low and middle income countries for equity-oriented research. *BMC Public Health*. 2006; 6 : 151 -
- 23. Passig D. Imen-Delphi : A Delphi variant procedure for emergence (L'Imen-Delphi : émergence d'une variante de la procédure Delphi). *Human Organ*. 1997; 56 : 53 - 63
- 24. Passig D, Sharbat A. The Imen-Delphi procedure in practice. *Systems Research and Behavioral Science*. 2004; 21 : 187 - 91
- 25. Delbecq AL, Van de Ven AH. A group process model for problem identification and program planning. *J Appl Behav Sci*. 1971; 7 : 466 - 92
- 26. Van de Ven A, Delbecq A. Nominal versus interacting group processes for committee decision-making effectiveness. *Acad Manage J*. 1971; 14 : 203 - 11
- 27. Fitzpatrick R, Boulton M. Qualitative methods for assessing health care. *Qual Health Care*. 1994; 3 : 107 - 13
- 28. Glaser EM. Using Behavioral Science Strategies for Defining the State-of-the-Art. *J Appl Behav Sci*. 1980; 16 : 79 - 92
- 29. Elwyn G, Edwards A, Iredale R, Davies P, Gray J. Identifying future models for delivering genetic services: a nominal group study in primary care. *BMC Family Practice*. 2005; 6 : 14 -
- 30. Vella K, Goldfrad C, Rowan K, Bion J, Black N. Use of consensus development to establish national research priorities in critical care. *BMJ*. 2000; 320 : 976 - 80
- 31. Lloyd-Jones G, Fowell S, Bligh JG. The use of the nominal group technique as an evaluative tool in medical undergraduate education. *Med Educ*. 1999; 33 : 8 - 13
- 32. Potter M, Gordon S, Hamer P. The nominal group technique : a useful consensus methodology in physiotherapy research. *Journal of Physiotherapy*. 2004; 32 : 126 - 30
- 33. Levine DA, Saag KG, Casebeer LL, Colon-Emeric C, Lyles KW, Shewchuk RM. Using a modified nominal group technique to elicit director of nursing input for an osteoporosis intervention. *J Am Med Dir Assoc*. 2006; 7 : 420 - 5
- 34. Aspinall F, Hughes R, Duncley M, Addington-Hall J. What is important to measure in the last months and weeks of life?: A modified nominal group study. *Int J Nurs Stud*. 2006; 43 : 393 - 403
- 35. Ferguson J. The NIH consensus development program. *Jt Comm J Qual Improv*. 1995; 21 : 332 - 6
- 36. Rotondi AJ, Kvetan V, Carlet J, Sibbald WJ. Consensus conferences in critical care medicine. Methodologies and impact. *Crit Care Clin*. 1997; 13 : 417 - 39
- 37. Guide pratique pour la réalisation d'une conférence de consensus. Paris ANDEM; 1992;
- 38. Carpentier F, Pazart L, Durocher A, Dosquet P, Fleurette F, HM. Les conférences de consensus : base méthodologique pour leur réalisation en France. ANAES, service recommandations professionnelles. 1999;
- 39. Sherman M, Bain V, Villeneuve JP, Myers RP, Cooper C, Martin S. The management of chronic viral hepatitis: A Canadian consensus conference 2004. *Can J Infect Dis Med Microbiol*. 2004; 15 : 313 - 26
- 40. Greig B, Oldaker T, Warzynski M, Wood B. 2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: recommendations for training and education to perform clinical flow cytometry. *Cytometry B Clin Cytom*. 2007; 72 : 23 - 33
- 41. Grivas TB, Wade MH, Negrini S, O'Brien JP, Maruyama T, Hawes MC. SOSORT consensus paper: school screening for scoliosis. Where are we today?. *Scoliosis*. 2007; 26 : 17 -
- 42. Sulkowski M, Benhamou Y. Therapeutic issues in HIV/HCV-coinfected patients. *J Viral Hepat*. 2007; 14 : 371 - 86
- 43. Kahan J, Kanouse D, Winkler J. Stylistic variations in National Institutes of Health consensus statements, 1979-1983. *Int J Technol Assess Health Care*. 1988; 4 : 289 - 304
- 44. Wortman PM, Vinokur A, Sechrest L. Do Consensus Conferences Work? A Process Evaluation of the NIH Consensus Development Program. *Journal of Health Politics Policy and Law*. 1988; 13 : 469 - 98
- 45. Ferguson J. The NIH Consensus Development Program. The evolution of guidelines. *Int J Technol Assess Health Care*. 1996; 12 : 460 - 74
- 46. McGlynn E, Koseoff J, Brook R. Format and conduct of consensus development conferences. Multi-nation comparison. *Int J Technol Assess Health Care*. 1990; 6 : 450 - 69
- 47. Bourg D, Boy D. Conférence de citoyens, mode d'emploi: Charles Léopold Mayer. 2005;
- 48. Fitch K, Bernstein S, Aguilar M, Burnand B, LaCalle J, Lazo P. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica Rand Corporation; 2001; Report No.: 1. MR-1269-DG-XII/RE
- 49. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. Paris HAS; 2006;
- 50. Brook RH. Editor: McCormick KAMS, Siegel RA. The RAND/UCLA Appropriateness Method. Methodology perspectives. Rockville (MD) Public Health Service, US Department of Health and Human Services; 1994; 59 - 70
- 51. Diehl J, Garrigues V, Guidet B, Jonquet O, Jouve P, Joye F. Méthodologie de recommandations d'experts: Bureau du Consensus Société de Réanimation de Langue Française (SRLF). 2002;
- 52. Park R, Fink A, Brook R, Chassin M, Kahn K, Merrick N. Physician ratings of appropriate indications for three procedures: theoretical indications vs indications used in practice. *Am J Public Health*. 1989; 79 : 445 - 7
- 53. Hunter D, McKee C, Sanderson C, Black N. Appropriate indications for prostatectomy in the UK--results of a consensus panel. *J Epidemiol Community Health*. 1994; 48 : 58 - 64
- 54. Shekelle PG, Park RE, Kahan JP, Leape LL, Kamberg CJ, Bernstein SJ. Sensitivity and specificity of the RAND/UCLA Appropriateness Method to identify the overuse and underuse of coronary revascularization and hysterectomy. *J Clin Epidemiol*. 2001; 54 : 1004 - 10
- 55. Consensus formalisé d'experts sur les indications de radiographie "grand champ". Paris ANAES, Service évaluation des technologies; 2003;
- 56. Prise en charge de l'hémochromatose liée au gène HFE (Hémochromatose de type 1). Paris HAS, Service des recommandations professionnelles; 2005;
- 57. Gustafson DH, Shukla RK, Delbecq A, Walster GW. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organizational Behavior and Human Performance*. 1973; 9 : 280 - 91
- 58. Van de Ven AH, Delbecq A. The effectiveness on nominal, Delphi and interacting group decision making processes. *Academy of Management Journal*. 1974; 17 : 605 - 21
- 59. Riggs WE. The Delphi technique: An experimental evaluation. *Technological Forecasting and Social Change*. 1983; 23 : 89 - 94
- 60. Shekelle P, Chassin M, Park R. Assessing the predictive validity of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care*. 1998; 14 : 707 - 27
- 61. Vader J-P, Burnand B, Froehlich F, Larequi-Laubert T, Pache I, Gonvers JJ. Indications à l'endoscopie digestive supérieure : Comparaison de critères américains et suisses. *Médecine et hygiène*. 1997; 55 : 2329 - 33
- 62. Simonet ML. Adéquation des soins : Qu'en est-il du patient ? . *Médecine et hygiène*. 1999; 57 : 2094 - 7
- 63. Sarasin FP. Méthodes de mesure de l'adéquation des soins et recommandations pour la pratique clinique. *Médecine et hygiène*. 1999; 57 : 2098 - 102
- 64. Froehlich F, Pache I, Burnand B, VADER J-P, Fried M, Kaseoff J. Underuse of upper gastrointestinal endoscopy. *Gastroenterology*. 1997; 112 : 690 - 7
- 65. Burnand B, Vader J, Froehlich F, Dubois R, Brook R, Gonvers J. Reliability of panel-based guidelines for colonoscopy : an international comparison. *Gastrointest Endosc*. 1998; 47 -
- 66. Pope C, Mays N. Qualitative Research: Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ*. 1995; 311 : 42 - 5

Tableau 1

Etapes des principales méthodes de consensus

| Etapes Méthode Delphi | | Groupe Nominal | Conférence de consensus | Méthode RAND/UCLA |
|------------------------------|--|-------------------------------------|---|--|
| 1 | Définition du thème | Définition du thème | Définition du thème | Définition du thème |
| 2 | Sélection des experts | Sélection des experts | Organisation logistique de la méthode | Sélection des experts |
| 3 | Elaboration du questionnaire | Organisation de la réunion | Sélection des experts et du jury | Répartition des sujets à traiter entre les groupes d'experts |
| 4 | Validation du questionnaire | Génération silencieuse des idées | Revue de littérature | Première réunion des groupes, revue de littérature et définition des items |
| 5 | Premier tour | Enoncé des idées | Conférence publique : question/réponse expert, jury, public | Seconde réunion des groupes rédaction du questionnaire global |
| 6 | Analyse des réponses du groupe | Discussion et tour de table | Délibération à huit clos du jury | Cotation individuelle des items |
| 7 | Second tour | Hierarchisation des idées | Rédaction des recommandations | Troisième réunion discussion des cotations |
| 8 | Analyse des réponses du groupe | Compilation des idées et validation | Communication des recommandations | Etablissement d'un rapport consensuel sur les items |
| 9 | Troisième tour | Communication des données obtenues | | Communication des résultats |
| 10 | Analyse des réponses du groupe | | | |
| 11 | Réitération autant de fois que nécessaire pour stabiliser les réponses du groupe (étapes 9–10) | | | |
| 12 | Conclusion de l'analyse, obtention d'un consensus | | | |
| 13 | Communication des informations obtenues | | | |