


couples. Avec le Registre national des FIV, l'Agence de la biomédecine développe un outil performant et spécifique ouvrant des perspectives intéressantes. Le projet mobilise fortement les professionnels qui œuvrent pour transmettre des données exhaustives et de bonne qualité, malgré les nombreuses difficultés de terrain. Long processus nécessitant la collaboration de tous les acteurs, le registre devrait être pleinement opérationnel en 2012.

L'Agence de la biomédecine réfléchit aux modalités possibles de suivi de l'état de santé des enfants, des couples qui ont eu recours à l'AMP et des donneuses d'ovocytes, comme la loi le prévoit. La faisabilité de croiser les données du registre avec différentes bases de données médicales est en cours d'étude.

À ce jour, l'agence utilise les données d'activité fournies par les établissements pour délivrer des informations en toute transparence au gouvernement, au Parlement, aux ARS, aux professionnels, aux médias et enfin au public. Elle s'efforce de communiquer de façon compréhensible et juste. Enfin, elle met en œuvre des stratégies d'information des couples entrant dans une démarche d'AMP garantissant la bonne compréhension des soins qui leur sont délivrés.

Avec toutes ces actions, l'agence se place au service des patients afin qu'ils puissent recevoir des soins les plus efficaces, dans les meilleures conditions possibles de prise en charge. 

Peut-on comparer les centres d'AMP entre eux ? Les pays entre eux ?

De plus en plus fréquemment, les résultats de FIV sont publiés par centre ou par région, soit par les centres eux-mêmes, soit par des agences gouvernementales, qui proposent d'ailleurs parfois en même temps un « classement » des centres de FIV, réalisé à partir des taux de succès obtenus dans chaque centre. C'est la marque d'une demande à la fois des autorités de santé, pour leur besoin d'agrément des centres, et du public pour orienter ses choix. D'une certaine façon, ce constat répond à la question posée : les résultats des centres sont, de fait, comparés. Le problème est cependant de savoir ce que représentent les différences constatées : correspondent-elles à des écarts réels ou à des fluctuations statistiques ? À des qualités différentes de prise en charge des couples ? À des recrutements différents ?

Bien qu'il s'agisse d'un aspect important de la question, nous laisserons de côté ici le choix de l'indicateur de résultat considéré pour comparer les centres : grossesses ou naissances vivantes ? Rapportées à une tentative, à une tentative et aux transferts d'embryons congelés associés ou aux tentatives successives d'un couple ? Cet article est consacré à deux questions communes à tous les indicateurs : comment tenir compte de l'incertitude liée au nombre de tentatives ? Comment tenir compte des recrutements différents des centres ?

Prise en compte du nombre de tentatives

La valeur observée (ou estimée) d'un taux de succès dans un centre d'AMP donné dépend, d'une part, de ce qu'on peut appeler son taux « réel » de succès et, d'autre part, de fluctuations dues au hasard. Hasard de ce que telle ou telle tentative a été comptabilisée

parce que le couple s'est présenté cette année et pas l'année précédente ou l'année suivante, hasard qu'une tentative donnée a conduit à un succès alors que, réalisée identiquement, elle aurait pu donner lieu à un échec. Les statistiques permettent de tenir compte de cette part de hasard en donnant, non pas un taux unique de succès, mais un intervalle (l'intervalle de confiance) qui a de très bonnes chances de contenir le taux réel. Cet intervalle est d'autant plus étroit que le nombre de tentatives est plus grand. Ainsi, 10 succès sur 50 tentatives ou 200 sur 1 000 représentent tous les deux 20 % de succès, mais les intervalles de confiance correspondants sont très différents, respectivement [10 % ; 34 %] et [18 % ; 23 %]. C'est ce qui est fait en Grande-Bretagne avec la publication des résultats de l'AMP sous forme de « League tables » (figure 1) [19]. Dans ce graphique, les traits verticaux représentent les intervalles de confiance. Cela veut dire, par exemple, que le taux de succès du centre 43, estimé à 12 %, doit être considéré comme compris quelque part entre 3 % et 25 %. Cela veut dire aussi qu'il faut se garder d'interpréter des écarts entre des centres dont les intervalles de confiance ont une zone de recouvrement importante. Par exemple, il n'y a aucune évidence qu'il y ait une réelle différence entre les centres 1 et 17, malgré des taux de succès estimés à 38 % et 18 %.

Toujours dans la figure 1, les traits horizontaux correspondent au taux de succès moyen (sur l'ensemble des centres) et à son intervalle de confiance. Ce dernier est bien sûr beaucoup plus étroit que les intervalles de confiance par centre car calculé sur plus de tentatives, mais sa fonction est la même que précédemment : il n'y a aucune évidence que les centres dont l'intervalle

Jean Bouyer

Directeur de recherche, Centre de recherche en épidémiologie et santé des populations Inserm, Ined, université Paris-Sud

Penélope Troude

Assistante hospitalo-universitaire, Service de santé publique et économie de la santé, hôpital Lariboisière, université Denis Diderot



de confiance recoupe les traits horizontaux soient réellement différents de la moyenne. C'est le cas de la plupart des centres de la figure 1.

On voit donc que ce type de graphique, s'il est interprété correctement, permet de relativiser les comparaisons entre les valeurs observées entre les centres. Il donne cependant une impression de classement, puisque les centres figurent dans l'ordre de leur taux de succès, qui peut être très trompeuse. On a d'ailleurs montré que ce classement pouvait varier de façon assez importante d'une année à l'autre, même si rien d'autre que la date n'avait changé [38].

Le problème majeur est surtout que ces League tables ne tiennent pas compte des différences de recrutement des centres [37], qui peuvent correspondre à des facteurs pronostiques différents des couples traités.

Prise en compte du recrutement des centres

Nous allons présenter ici deux méthodes. La première consiste à définir une population de référence, identique pour chaque centre, sur laquelle les taux de succès sont calculés. La seconde recourt à des techniques statistiques plus sophistiquées pour calculer un taux de succès dit « ajusté sur les caractéristiques des couples pris en charge ».

Définir une population de référence

Puisque les caractéristiques des couples pris en charge peuvent avoir une influence sur le taux de succès et peuvent différer d'un centre à l'autre, une possibilité est de limiter le calcul des taux de succès et leurs comparaisons à une sous-population de couples ayant des caractéristiques homogènes [49]. L'objectif est de

pouvoir exclure au mieux les différences de recrutement comme explication à des différences de taux de succès.

Le principe général est résumé sur la figure 2 en se limitant à la prise en compte de l'âge des femmes. On voit bien qu'en restreignant l'étude à la bande horizontale, l'âge des femmes ne peut pas être invoqué comme cause de différences entre les centres (du moins si on admet que « 25-35 ans » représente une catégorie homogène quant aux chances de succès).

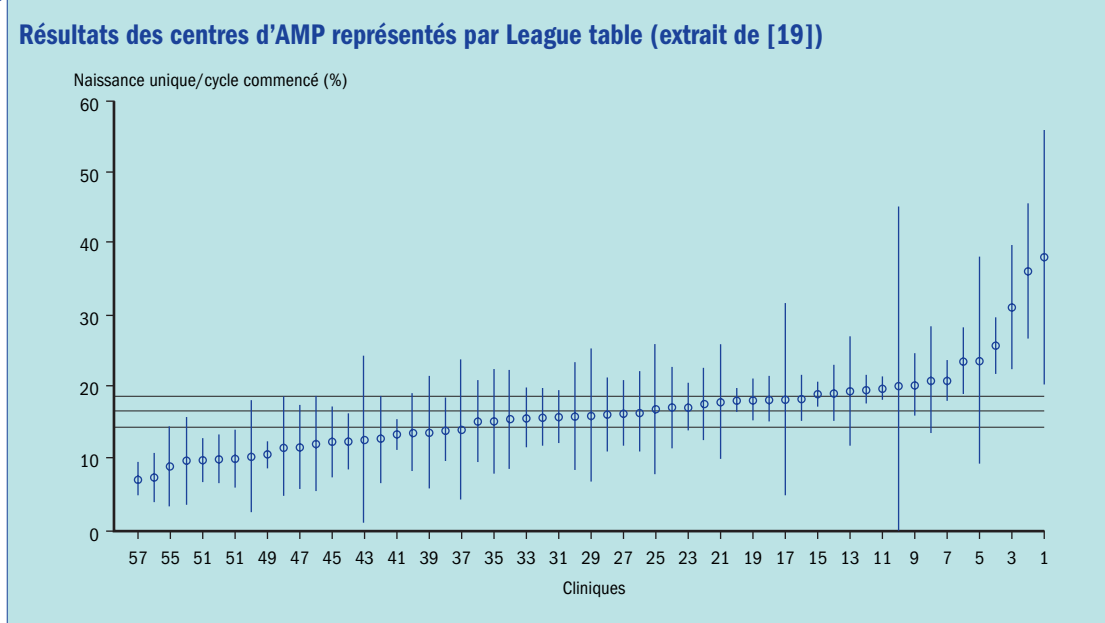
On peut bien sûr étendre ce procédé. Par exemple avec plusieurs catégories d'âge (et donc plusieurs populations de référence) comme le fait HFEA britannique (Human Fertilisation and Embryology Authority). Ou en définissant la population de référence par plusieurs caractéristiques : femmes de moins de 35 ans, dont c'est la première tentative, ayant un dosage de FSH « satisfaisant » et un BMI compris entre 19 et 35 [24].

Cette méthode a l'avantage d'être facile à mettre en œuvre, mais aussi à comprendre par un public non spécialisé. Elle ne demande pas de disposer des données individuelles des tentatives pour chaque centre. Elle a cependant l'inconvénient de ne prendre en compte que partiellement les différences de recrutement, notamment si la population de référence n'est pas assez homogène. Elle n'utilise de plus qu'une partie (éventuellement petite) de l'activité des centres, puisque toutes les tentatives ne faisant pas partie de la population de référence ne sont pas prises en compte.

Ajustement sur les caractéristiques des couples pris en charge

Le principe général est de calculer pour chaque centre un taux de succès (dit « ajusté ») qui est le taux de succès que

figure 1



devrait avoir ce centre si la distribution des caractéristiques des couples était la même que celle de l'ensemble des centres réunis. De cette façon, la part des différences de recrutement entre les centres dans la comparaison des taux de succès est « éliminée » autant que possible.

Sur le plan technique, il s'agit d'une standardisation indirecte, dont le principe est utilisé depuis longtemps pour comparer les taux de mortalité par cause entre les pays, en tenant compte du fait que les pays ont des structures d'âge et de répartition par sexe différentes.

Sur le plan pratique, cette méthode nécessite les données individuelles des tentatives et de leurs caractéristiques de façon à pouvoir déterminer (sur l'ensemble

des centres) les variables liées au succès ou ayant une pertinence clinique, et qui doivent donc intervenir dans le calcul du taux de succès ajusté. Cela peut être des caractéristiques des sujets (âge, état des trompes, Hormone folliculostimulante — FSH —, spermogramme...), des tentatives (antécédents de grossesses, rang, nombre d'ovocytes, d'embryons, fécondation *in vitro* avec micro-injection (ICSI)...), le niveau social, la consommation de tabac, l'IMC...

Les résultats obtenus peuvent être représentés par un graphique appelé « Funnel plot » (figure 3). Il est constitué d'une sorte d'entonnoir, autour du taux de succès moyen de l'ensemble des centres, dont les

figure 2

Comparaison des résultats des centres sur une population de référence (ici, les femmes de 25 à 35 ans)

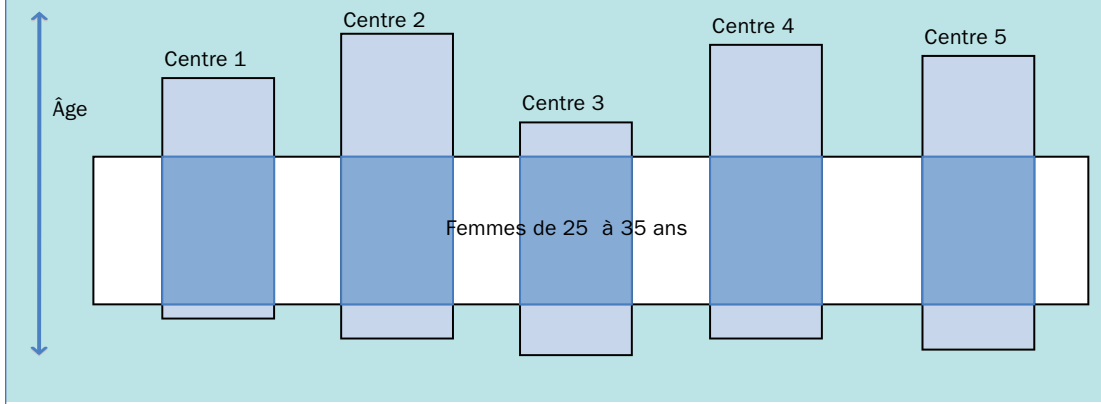
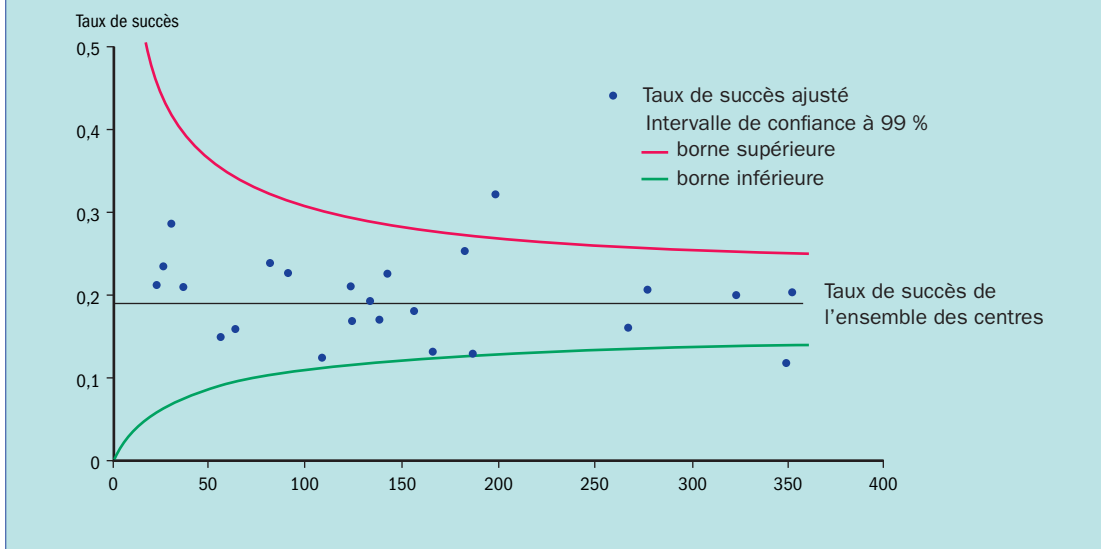


figure 3

Résultats des centres d'AMP représentés par Funnel plot





bornes inférieure et supérieure limitent une zone où les taux de succès des centres peuvent être considérés comme non différents de la moyenne. Les taux de succès (ajustés) de chaque centre sont représentés par des points sur le graphique permettant de détecter ceux qui ont des résultats « hors norme », de façon à entreprendre des études complémentaires pour en comprendre les raisons.

Cette méthode doit donc être comprise comme un outil de détection et non de « diagnostic ». La représentation graphique a par ailleurs l'avantage de ne pas proposer de façon évidente un classement entre les centres, dont on a vu la fragilité.

Comparaison entre pays


Les résultats de l'AMP sont régulièrement publiés au niveau européen [43] ou au niveau mondial [42]. La question de leurs comparaisons éventuelles se pose dans des termes similaires à ce qui a été dit plus haut. Il faut cependant ajouter que les différences entre pays, au-delà des différences de caractéristiques des couples pris en charge évoquées plus haut, se situent à d'autres niveaux liés aux conditions de prise en charge, d'accès aux soins, de législation (nombre d'embryons implantés, possibilité de congélation, utilisation de sperme de donneur...). Il paraît donc beaucoup plus difficile de tenir compte des différences de recrutement entre pays, même de niveau de développement économique comparable, et donc de comparer leur taux de succès en FIV. En ce sens, comparer, par exemple, les taux bruts de succès entre la France et d'autres pays (d'Europe ou d'Amérique du Nord) pour en tirer des enseignements sur la qualité relative de la prise en charge des couples paraît pour le moins périlleux.

Notons enfin que les données internationales sont collectées de façon agrégée (et non pas tentative par tentative) et sont publiées « brutes », sans tenir compte donc du recrutement des centres, ce qui ajoute aux réserves sur leurs comparaisons. Il est cependant envisagé que les données européennes soient fournies dans le futur par classe d'âge des femmes, ce qui peut faire évoluer les choses [43].

Conclusion

Il est donc possible de comparer les centres d'AMP entre eux, à condition d'utiliser des méthodes appropriées. La comparaison entre pays est plus difficile car un plus grand nombre de facteurs de différences, notamment socio-économiques et de législation, intervient. La méthode de la population de référence, même si elle n'est pas la meilleure, est un premier pas, particulièrement utile en l'absence de données individuelles. La méthode du taux ajusté et de sa représentation sur un graphique de type « Funnel plot » est préférable quand on dispose des données individuelles des tentatives et de leurs caractéristiques.

Dans tous les cas, la publication de résultats par centre demande de la pédagogie pour expliquer ce qui est fait, à l'image de la section « Comprendre les taux de succès des centres de FIV » du site internet de la HFEA (<http://www.hfea.gov.uk/fertility-clinics-success-rates.html>).

Enfin, même si cela n'a pas été discuté ici, il ne faut pas oublier l'importance du choix de l'indicateur de succès. En soulignant qu'il n'est pas nécessairement unique, car la notion de succès a plusieurs facettes. Retenir plusieurs indicateurs a par ailleurs l'avantage d'empêcher un classement simpliste et peu fiable entre les centres. 

Les enfants de l'AMP

Sylvie Epelboin

Gynécologue
responsable du
Centre d'assistance
médicale à la
procréation, Hôpital
Bichat, Paris

Tandis qu'en 1978 Louise Brown était la seule enfant née par fécondation *in vitro* (FIV) parmi les 128 000 000 enfants venus au monde cette année-là, actuellement, en France, près de 3 % des naissances, soit 1 sur 40, sont consécutives à une AMP. D'un fait divers qui souleva en son temps bien des passions, sans cesse renouvelées à chaque avancée technologique, les naissances par AMP sont donc devenues une réalité sociale. Modes de conception connus et admis du grand public, les procréations assistées sont cependant encore à l'origine de représentations fantasmagiques, de l'ordre paradoxal de l'excès, des dérives ou des manques, et les enfants qui en sont conçus encore bien souvent considérés comme « hors-norme », prodiges sélectionnés, ou victimes d'une technologie abusive. Les médias relaient l'information en misant sur le sensationnel, d'où l'intérêt fondamental d'études

scientifiques menées sur de larges cohortes d'enfants issus de la technique, portant sur leur santé néonatale, et leur développement.

La démarche de suivi des enfants n'a pas été immédiate après les débuts de la FIV, mais l'a été dès les premières ICSI en raison des interrogations suscitées par le caractère intrusif de la technique. Il existe plusieurs dizaines de milliers de publications sur l'AMP, plusieurs centaines de papiers fournissent des données sur les enfants à la naissance, mais moins d'une cinquantaine concernent l'évolution des enfants de plus de 2 ans.

Ces suivis sont extrêmement précieux pour l'évaluation des techniques, mais il faut souligner qu'ils nécessitent la collaboration des parents qui acceptent de s'enrôler dans des enquêtes par questionnaires, ou de soumettre leurs enfants à des examens itératifs, alors qu'ils pourraient souhaiter oublier le parcours d'infertilité qui fut le