

L'essai clinique randomisé par grappes

Bruno Giraudeau

> Dans le cadre d'un essai clinique randomisé, on est parfois amené à randomiser des groupes de sujets plutôt que les sujets eux-mêmes. Ce type de schéma expérimental a des conséquences méthodologiques importantes - notamment sur le plan biostatistique) - ainsi que sur l'extrapolation des résultats. Sa mise en œuvre n'est donc justifiée que si le contexte ou la question investiguée l'imposent et les aspects biostatistiques doivent alors être traités de manière rigoureuse. <

Un essai clinique randomisé par grappes (*cluster randomization trial*) est un essai dans lequel on ne randomise pas individuellement des sujets, mais des groupes de sujets qu'on appelle des «grappes» (*clusters*). Ces unités de randomisation peuvent être des hôpitaux, des médecins, des familles, des villages, des entreprises, etc., autant d'unités «sociales» pour lesquelles les sujets qui composent une unité ne peuvent être considérés comme indépendants les uns des autres. Ce type de schéma expérimental est apparu il y a maintenant plus de vingt ans et les aspects méthodologiques qui lui sont associés ont été depuis largement abordés, donnant lieu entre autres à deux ouvrages [1, 2] et un numéro spécial de *Statistics in Medicine* [3].

Après avoir discuté dans une première partie les motivations à randomiser des grappes plutôt que des individus, nous aborderons les implications méthodologiques et éthiques d'un tel schéma expérimental, et enfin ses conséquences en termes d'extrapolation des résultats.

Contextes justifiant une randomisation par grappes

La randomisation de grappes a des conséquences méthodologiques importantes et peut notamment conduire à une augmentation substantielle de la taille de l'échantillon nécessaire (*voir plus loin*). La



motivation à randomiser des grappes doit donc être parfaitement explicite et plusieurs justifications existent. Bien souvent, on est amené à randomiser des grappes parce que l'intervention testée ne s'applique pas à l'échelon individuel, mais à un échelon supérieur (le médecin, l'hôpital...). Ainsi, pour tester l'efficacité d'un nouveau programme de prise en charge anténatale, A. Donner *et al.* [4] ont randomisé des hôpitaux. C'est en effet au niveau du médecin que l'intervention s'applique (puisque, dans le bras expérimental de cette étude, on leur demande de modifier leur prise en charge, et notamment les recommandations faites aux femmes enceintes), même si c'est à l'échelon individuel des nouveau-nés que les retombées potentielles sont mesurées. Il est alors difficile d'imaginer qu'un médecin puisse donner des recommandations différentes à deux femmes qui auraient le même profil démographique, clinique et biologique. Il est également difficile d'imaginer que deux médecins travaillant dans le même service du même hôpital n'aient pas des attitudes homogènes. La seule unité de randomisation logique est donc l'hôpital. Une autre motivation à randomiser des grappes est la crainte d'une contamination entre les groupes (*voir*

Inserm CIC 202,
Faculté de Médecine,
2bis, boulevard Tonnellé,
37032 Tours Cedex, France.
[giraudeau@
med.univ-tours.fr](mailto:giraudeau@med.univ-tours.fr)



glossaire) à comparer. L'existence d'une contamination inter-groupes aurait en effet une double conséquence: d'une part l'effet estimé s'en trouverait biaisé, et d'autre part, les hypothèses statistiques sous-jacentes au test réalisé ne seraient plus vérifiées puisque les deux échantillons ne seraient plus indépendants. Ces contaminations peuvent survenir par exemple lorsqu'on étudie des agents contagieux (comme dans l'étude de O. Chosidow *et al.* [5] qui compare deux lotions anti-pédiculaires), ou dans les études prophylactiques (situations dans lesquelles la randomisation par grappes permet d'éviter qu'un sujet ne perde sa susceptibilité à être infecté, tout simplement parce qu'il évolue au milieu de sujets immunisés par le traitement étudié). C'est également le cas lorsque l'on souhaite étudier un programme de prévention: des interactions actives entre les sujets des deux bras de randomisation (le bras expérimental et le bras témoin) pourraient en effet largement contaminer le bras témoin.

Enfin, des considérations logistiques peuvent également conduire à la randomisation de grappes, même s'il est théoriquement possible de randomiser des sujets. Ainsi, A. Sommer *et al.* [6] ont randomisé 450 villages indonésiens pour tester l'efficacité d'une supplémentation en vitamine A, parce que la randomisation des sujets a été considérée comme non «politiquement et administrativement réalisable». De la même façon, pour évaluer un régime alimentaire, il est plus aisé de randomiser des familles que des sujets, ce qui évite d'avoir deux menus à préparer quotidiennement.

Implications méthodologiques

En randomisant des grappes, on se met en situation de recueillir des données corrélées ce qui conduit à une double conséquence biostatistique. Premièrement, le calcul du nombre de sujets nécessaire doit prendre en compte cette corrélation et la taille des échantillons s'en trouve augmentée. En effet, les résultats observés chez deux sujets d'une même grappe auront tendance à être plus «similaires» que si les résultats étaient associés à deux sujets provenant de deux grappes distinctes (donc indépendantes). L'information apportée par une grappe de k sujets est donc moindre que s'il s'agissait de k sujets indépendants et cette perte de puissance statistique doit donc être compensée par une augmentation du nombre de sujets à inclure. En pratique, le principe appliqué est donc le suivant: on calcule la taille de l'échantillon nécessaire en utilisant les formules classiques pour des données indépendantes, et on applique un facteur d'inflation défini par $(1 + [\bar{m} - 1] \rho)$ où \bar{m} est la taille moyenne d'une grappe et ρ est

égal au niveau de corrélation entre les réponses des sujets d'une même grappe. Ce dernier coefficient ρ , appelé coefficient de corrélation intraclasse, prend des valeurs qui dépendent de l'unité de randomisation (la corrélation sera plus forte si on randomise des médecins que si on randomise des hôpitaux) mais qui sont classiquement de l'ordre de 0,001 à 0,05 [7]. Ces corrélations intraclasses peuvent paraître faibles mais ne doivent pas pour autant être négligées. À titre d'exemple, A. Donner *et al.* [4], en retenant un coefficient à 0,001 et une taille moyenne de 450 sujets par grappe obtenaient un coefficient d'inflation de 1,45 ce qui signifie donc une augmentation du nombre de sujets à inclure de 45 %. Par ailleurs, l'analyse des données recueillies doit recourir à des méthodes statistiques *ad hoc*, sous peine d'obtenir des résultats abusivement significatifs. En effet, si l'on ne tenait pas compte de la corrélation induite par le schéma expérimental, les degrés de signification se trouveraient sous-estimés, les étendues des intervalles de confiance faussement diminuées et les conclusions de l'étude erronées. Différentes méthodes permettent d'analyser correctement les données parmi lesquelles l'utilisation de modèles mixtes ou marginaux [1, 2] qu'il est possible d'estimer à l'aide des logiciels statistiques classiquement utilisés tel SAS (SAS Institute Inc, Cary, NC, USA).

Par ailleurs, contrairement aux essais où l'on randomise les sujets individuellement, la randomisation par grappes conduit de façon quasi systématique à une situation où la randomisation est antérieure à l'inclusion des sujets. En pratique, on commence en effet par randomiser les grappes (médecins, hôpitaux...), charge ensuite à chaque «responsable de grappe» d'inclure les sujets. Lorsque l'essai est réalisé en ouvert (ce qui est nécessairement le cas si l'intervention étudiée prend la forme d'un programme de prise en charge, d'un régime alimentaire...), l'inclusion des sujets se fait alors en ayant connaissance du bras dans lequel il sera inclus. Cette situation est potentiellement source de biais, et la comparaison *a posteriori* des groupes issus de la randomisation est alors d'autant plus importante.

Aspects éthiques - consentement

Dans les essais randomisés «classiques», le recueil du consentement du sujet est un préalable nécessaire à son inclusion. Dans le cadre d'un essai randomisé en grappes, les choses sont parfois plus complexes: faut-il recueillir le consentement du sujet ? du «responsable de grappe» ? Concernant ce dernier, son consentement

est incontournable, ne serait-ce que pour assurer le recrutement des sujets de l'étude. On peut parfois être amené à informer les sujets inclus dans l'étude, sans pour autant obtenir un consentement signé: quand la recherche concerne un traitement de l'eau, l'utilisation à large échelle d'insecticides, d'agents prophylactiques ou immunisants, le consentement individuel n'est pas possible et la décision ultime revient alors au « responsable de grappe » [8].

Par ailleurs, la randomisation étant antérieure à l'inclusion (voir ci-dessus) le consentement du sujet (s'il est possible de le recueillir) devient alors un consentement double. En effet, le sujet accepte d'une part de participer à l'étude, et d'autre part les conditions de sa participation, c'est-à-dire que le sujet sait s'il est dans le bras expérimental ou dans le bras témoin. Cette situation n'est pas sans rappeler le schéma proposé par M. Zelen [9] où le consentement du sujet est recueilli après la randomisation.

Extrapolation des résultats

Une singularité de l'essai randomisé par grappes réside dans le fait que l'extrapolation des résultats se fait souvent au niveau d'unités qui ne sont pas les unités de randomisation, voire même, qui ne sont pas les unités sur lesquelles l'intervention porte. Quand A. Donner *et al.* [4] réalisent leur essai, ils randomisent des hôpitaux, forment des médecins à un programme expérimental de prise en charge des femmes enceintes, espérant ainsi des retombées en termes de réduction des petits poids de naissance et de réduction des morbidités maternelles (pré-éclampsie ou éclampsie, anémie, infection urinaire...). C'est cette distinction entre l'unité de randomisation et l'unité d'inférence des résultats (c'est-à-dire l'unité qui sera prise en compte pour l'analyse et l'extrapolation des résultats) qui est la cause des complexifications biostatistiques évoquées plus haut; c'est cette même distinction qui limite parfois les conclusions de l'étude. En effet, si très peu de grappes sont randomisées, la question se pose alors de savoir si l'effet observé est le fait de l'intervention mise en place ou bien s'il est dû aux caractéristiques des grappes randomisées dans le bras expérimental. La règle qui prévaut est alors celle énoncée par D.M. Murray [2]: plus le nombre de grappes randomisées est important, plus on se rapproche de l'essai avec randomisation individuelle, et les conclusions tirées d'une analyse statistique *ad hoc* sont alors aussi «solides» que si des sujets avaient été randomisés.

Conclusions

Randomiser des grappes conduit à une étude moins puissante que si les sujets étaient randomisés individuellement: tailles d'échantillon accrues, analyses statistiques complexifiées, extrapolations des résultats plus limitées. Ce type de schéma expérimental ne doit donc être mis en œuvre que lorsque le contexte et l'objectif de l'étude l'imposent, et les aspects biostatistiques doivent alors être traités avec la rigueur nécessaire, sous peine de parvenir à des conclusions erronées. ♦

GLOSSAIRE

Bras. Dans un essai randomisé, un «bras» est un groupe de sujet. Le «bras» expérimental est le groupe de sujets bénéficiant du nouveau traitement (ou nouveau programme de prise en charge), alors que les sujets du bras témoin bénéficient de la prise en charge standard.

Coefficient de corrélation intraclasse. Coefficient mesurant le degré de ressemblance des patients au sein d'une grappe (appelée également classe). Un coefficient nul indique qu'il y a indépendance entre les patients d'une même grappe. À l'inverse, plus le coefficient est élevé, plus les patients d'une même grappe sont semblables.

Contamination inter-groupes. Phénomène indiquant que les patients d'un des bras de l'essai bénéficient également (tout au moins pour certains d'entre eux) de l'intervention mise en place dans l'autre bras de l'essai. L'existence d'une contamination remet en cause l'indépendance des deux bras de l'essai et biaise l'estimation de l'effet de l'intervention.

Données corrélées. On parle de données corrélées lorsque, du fait du mode de recueil de ces données, elles se «ressemblent». Par exemple, lorsque des femmes enceintes sont suivies par un même médecin, les paramètres mesurés sur les bébés seront plus «semblables» que si elles avaient bénéficié de prises en charge indépendantes par des médecins distincts.

Facteur d'inflation. Facteur par lequel la taille d'échantillon nécessaire doit être multipliée lorsque le plan expérimental prévoit une randomisation par grappes.

Grappe (cluster). Ensemble de patients définissant une entité sociale, clinique ou économique. Exemple de grappes: familles, patients vus par un même clinicien, hôpitaux, entreprises, villages ...

Randomiser. Anglicisme signifiant «tirer au sort»; de même, une randomisation est un tirage au sort.

SUMMARY

The cluster-randomized trial

Within the framework of a clinical trial, one is sometimes brought to randomize groups of subjects rather than the subjects themselves. This type of experimental design has methodological consequences of importance (in particular biostatistical consequences) and can limit the extrapolation of the results. A cluster randomized trial is thus justified only if the context or the question of interest imposes it and biostatistical issues have then to be handled in a rigorous way. ♦

RÉFÉRENCES

1. Donner A, Klar N. *Design and analysis of cluster randomization trials in health research*. London: Arnold, 2000: 178 p.
2. Murray DM. *Design and analysis of group-randomized trials*. New York: Oxford University Press, 1998: 468 p.
3. Campbell MJ, Donner A, Elbourne DR. Design and analysis of cluster randomized trials. *Stat Med* 2001; 20: 329-496.
4. Donner A, Piaggio G, Villar J, et al. Methodological considerations in the design of the WHO antenatal care randomised controlled trial. *Paediatr Perinat Epidemiol* 1998; 12 (suppl 2): 59-74.
5. Chosidow O, Chastang C, Brue C, et al. Controlled study of malathion and d-phenothrin lotions for *Pediculus humanus var capitis*-infested schoolchildren. *Lancet* 1994; 344: 1724-7.
6. Sommer A, Tarwotjo I, Djunaedi E, et al. Impact of vitamin A supplementation on childhood mortality. A randomised controlled community trial. *Lancet* 1986; 24: 1169-73.
7. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *Br Med J* 1998; 316: 1455.
8. Howard-Jones N. Human experimentation in historical and ethical perspectives. *Soc Sci Med* 1982; 16: 1429-48.
9. Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979; 300: 1242-5.