

## Enoncé et corrigé du TD sur la puissance statistique et le nombre de sujets à inclure dans une étude clinique (séance n°6)

### Objectifs d'apprentissage visés du TD n°6

- (A) **Définir** l'inférence statistique ;
- (A) **Citez** dans une phrase les pourcentages qui vont être comparés puis testés, à partir d'effectifs fournis dans un tableau, dans l'objectif de tester l'association entre deux variables binaires ;
- (A) **Conclure** à partir des résultats d'un test statistique (incluant la valeur du degré de signification) comparant plusieurs (deux ou plus) moyennes (que les données soient appariées ou non), plusieurs médianes, et plusieurs pourcentages, en faisant entre autres de l'inférence statistique si elle est possible à faire ;
- (A) **Interpréter** les informations contenues dans un tableau issu d'une étude clinique comprenant des pourcentages, moyennes, médianes, et degrés de signification ;
- (A) **Choisir** le test statistique (parmi les tests de Fisher, du Chi<sup>2</sup>, de Student pour séries non appariées, de Student pour séries appariées, de Mann-Whitney, de l'ANOVA, de Kruskal-Wallis) en fonction de la question de recherche d'une étude, des indicateurs que l'on souhaite comparer puis tester, et de la validité des conditions de validité des tests si plusieurs tests statistiques sont possibles pour une même question de recherche ;
- (A) **Communiquer** de façon pertinente à l'issue d'un résultat du test statistique non significatif ;
- (A) **Identifier** la situation dans laquelle un message issu de la recherche clinique nécessite un groupe contrôle ;
- (A) **Citer** l'hypothèse nulle H<sub>0</sub> d'un test statistique, qu'il soit ou non présenté dans un tableau issu d'une étude clinique ;
- (B) **Définir** la puissance statistique d'une étude ;
- (B) **Interpréter** la valeur de la puissance statistique d'une étude à partir des paramètres qui ont été nécessaires à un calcul de nombre de sujets à inclure, à partir d'une copie d'écran de BiostaTGV ;
- (B) **Interpréter** la valeur du nombre de sujets à inclure dans une étude à partir des paramètres qui ont été nécessaires à ce calcul, à partir d'une copie d'écran de BiostaTGV.

### Descriptif général du TD

Ce TD va utiliser les résultats de l'article de Dye et coll. paru en 2013 dans le Journal of Veterinary Internal Medicine, intitulé « Randomized, Controlled Trial of Budesonide and Prednisone for the Treatment of Idiopathic Inflammatory Bowel Disease in Dogs », soit en français « essai clinique randomisé du Budesonide et de la Prednisone pour le traitement de la Maladie Inflammatoire Chronique de l'Intestin (MICI) chez le chien ». L'objectif principal de l'étude était de montrer que les deux traitements ont des effets différents sur le contrôle de la MICI évalués 6 semaines après le début du traitement, chez les chiens atteints de MICI. L'échantillon était constitué de 18 chiens ayant reçu du Budesonide pendant 6 semaines, et de 16 chiens ayant reçu de la Prednisone pendant 6 semaines (le traitement de référence). Tous les chiens de l'étude pouvaient être considérés comme indépendants. Le risque d'erreur de 1<sup>ère</sup> espèce est fixé, comme d'habitude, à 0,05 (5%).

L'un des critères de jugement pour comparer les effets du traitement 6 semaines après le début du traitement était la valeur du score clinique CIBDAI<sup>1</sup> (caractère quantitatif) à 6 semaines.

Le tableau ci-dessous présente un extrait des résultats obtenus dans l'essai clinique. Dans ce tableau, les auteurs ont présenté la moyenne (SD) des scores cliniques CIBDAI le jour de l'initiation du

---

<sup>1</sup> Pour « Canine Inflammatory Bowel Disease Activity Index ».

traitement (« Baseline ») et celle à 6 semaines après le début du traitement, dans chacun des deux groupes.

**Table 3.** Comparison of selected variables between baseline and 6 weeks for each study group and between study groups at 6 weeks (mean [SD] or number and percentage).

	Prednisone Group			1	Budesonide Group			Prednisone versus Budesonide at 6 weeks
	Baseline	6 Weeks	P-Value		Baseline	6 Weeks	P-Value	P-Value
CIBDAI score	8.8 (3.3)	2.5 (3.4)	<0.0001	7.5 (2.6)	1.3 (2.0)	<.0001	.22	

1) Peut-on inférer de la valeur du degré de signification pointé par la flèche n°1 que la Prednisone est un traitement efficace pour réduire le score clinique CIBDAI après 6 semaines de traitement chez des<sup>2</sup> chiens atteints de MICI ?

*Non, car comme nous l'avons vu dans le TD précédent sur les séries appariées, le test statistique pointé par la flèche n°1 compare la moyenne du score clinique au moment de l'initiation du traitement (« Baseline ») et à celle à 6 semaines après l'initiation du traitement, à l'aide du test de Student pour séries appariées. De cette significativité, on peut inférer que, sous l'hypothèse d'absence de biais d'association, il y a des chances pour le score clinique évolue réellement en 6 semaines de traitement avec de la Prednisone. Mais il manque un groupe contrôle (des chiens sans aucun traitement) pour pouvoir éventuellement inférer que le traitement est efficace (sans groupe contrôle, on ne peut absolument pas exclure que cette évolution soit liée à autre chose que cette prise de ce traitement).*

2) Les indicateurs qui sont comparés puis testés par le test statistique dont la valeur du degré de signification vaut 0,22 pointé par la flèche n°2 sont la moyenne du score CIBDAI à 6 semaines post traitement parmi les 16 chiens sous Prednisone (2,5) et la moyenne de ce score parmi les 18 chiens sous Budesonide (1,3). Quel est le nom de ce test statistique dont le degré de signification vaut 0,22 ?

*Le test statistique est le test de Student pour séries non appariées (car les deux groupes comparés sont le groupe de chiens traités par la Prednisone et celui des chiens traités par le Budesonide, deux groupes indépendants constitués de chiens différents).*

3) A partir de la valeur du degré de signification pointé par la flèche n°2, concluez tout d'abord au niveau de l'échantillon, puis ensuite en faisant de l'inférence statistique, si cette dernière est possible à faire.

*Le degré de signification vaut 0,22, donc supérieur à 0,05.  $H_0$  est donc acceptée et il n'est donc pas possible de faire de l'inférence. Ainsi, une seule phrase doit être citée, celle qui concerne la conclusion au niveau de l'échantillon : « dans l'échantillon, la moyenne du score CIBDAI après 6 semaines de traitement parmi les 16 chiens traités avec de la Prednisone (2,5) n'était pas significativement différente de la moyenne du score CIBDAI après 6 semaines de traitement parmi les 18 chiens traités avec du Budesonide (1,3 ;  $p = 0,22$ ).*

4) Vu que la valeur moyenne du score clinique CIBDAI à 6 semaines est plus faible dans le groupe Budesonide que dans le groupe Prednisone (1,3 versus 2,5, respectivement), et en supposant que cette différence puisse être jugée comme cliniquement importante, est-ce que les auteurs de l'étude peuvent dire lors d'une conférence : « à la vue de nos résultats, dans la population des chiens atteints de MICI, il semble que la valeur moyenne du score CIBDAI 6 semaines après traitement par du Budesonide soit moins importante que celle parmi les chiens traités par de la Prednisone » ?

*Absolument pas. La différence n'étant pas significative, on n'a pas le droit de faire comme si elle l'était, et donc de faire de l'inférence (ce que les auteurs souhaitent faire ici) !*

<sup>2</sup> Ou « les million de »

5) A partir du résultat obtenu dans l'étude, les auteurs peuvent-ils dire lors d'une conférence : « à la vue de nos résultats, dans la population des chiens atteints de MICI, il semble que la valeur moyenne du score CIBDAI 6 semaines après traitement parmi les chiens traités par du Budesonide soit voisine de celle parmi les chiens traités par de la Prednisone » ?

*Absolument pas (encore une fois 😊), c'est LA belle erreur à ne surtout pas commettre. Quand  $H_0$  ne peut pas être rejetée ( $p > 0,05$ ) il est interdit d'inférer en disant qu'il semble que  $H_0$  soit vraie. On reste au niveau de l'échantillon en disant ce que l'on a dit dans la question n°3, et c'est tout.*

6) Suite aux résultats de cette étude publiée, une autre équipe de cliniciens chercheurs veut réaliser un essai clinique randomisé sur des chiens avec une MICI cette fois-ci plus sévère, pour tenter de montrer que le score CIBDAI moyen à 6 semaines de traitement Budesonide est inférieur à celui à 6 semaines de traitement Prednisone. Pour calculer le nombre de chiens à inclure dans leur étude, dont elle veut que la puissance statistique soit de 80%, l'équipe fait tout d'abord l'hypothèse qu'avant attribution aléatoire de l'un des deux traitements, les chiens auront en moyenne un score clinique CIBDAI de 11. Grâce à ses connaissances du terrain, l'équipe s'attend à ce que la moyenne du score clinique CIBDAI des chiens qui reçoivent de la Prednisone (le traitement historique de référence chez des chiens atteints de MICI) pendant 6 semaines atteigne la valeur de 4,5. Ils font ensuite l'hypothèse que le score CIBDAI moyen après 6 semaines de Budesonide atteigne la valeur de 3. L'équipe fait enfin l'hypothèse que la SD du score CIBDAI parmi les chiens atteints de MICI est d'environ 2,5. Combien de chiens avec une MICI sévère l'équipe doit-elle recruter dans chaque groupe de traitement ? (Vous cliquerez sur « Calculer » en face de « Comparer deux moyennes », après avoir cliqué [ici](#) pour répondre à la question. Vous choisirez les groupes que vous voudrez pour les groupes 1 & 2 sur le site Internet, et les tests statistiques sont toujours bilatéraux<sup>3</sup>. Par ailleurs, sur le site, « Ecart type commun  $\sigma$  » signifie SD.)

Calcul du nombre de sujets nécessaires	
Comparer deux proportions binomiales	Calculer
Comparer deux moyennes	Calculer
Tester l'équivalence de deux proportions binomiales	En travaux
Etudes cas-témoins	Calculer
Etudes cas croisée	Calculer

<sup>3</sup> la différence entre un test statistique unilatéral et un test statistique bilatéral est hors programme, car tout d'abord, cette distinction fait appel à des éléments de statistique mathématique « dure », et ensuite, tous les articles de recherche clinique utilisent des tests statistiques dits « bilatéraux ».

Après avoir cliqué sur « Calculer » en face de « Comparer deux moyennes ». Nous obtenons cet écran ci-dessous, que j'ai rempli :

1 & 2 :  $\mu_1$  et  $\mu_2$  sont les moyennes attendues respectivement dans le groupe Prednisone (arbitrairement « groupe 1 ») et Budesonide (arbitrairement « groupe 2 ») dans la population des chiens atteints de MCI sévère (d'où la lettre grecque) (cf. énoncé de la question). On met ici les valeurs attendues de 4,5 et 3.

3 : si la moyenne attendue dans le 1<sup>er</sup> groupe est de 4,5 et la moyenne attendue dans le 2<sup>ème</sup> groupe est de 3, cela conduit de fait à une réelle différence de moyennes (noté  $\Delta$  dans le cours et dans le polycopié, mais notée ici « d »<sup>4</sup>) égale à 1,5. Je pourrai d'ailleurs vous montrer plus tard dans le TD que le nombre de sujets nécessaires sera identique si nous avons mis d'autres nombres pour  $\mu_1$  et  $\mu_2$  mais tels que  $|\mu_1 - \mu_2|$  reste égale à 1,5.

4 : la SD attendue est de 2,5 (cf. énoncé de la question), donc c'est ce que l'on met ici.

5 : le risque d'erreur est fixé à 0,05, donc c'est ce que l'on met ici.

6 : comme l'équipe a souhaité mettre en place un essai clinique avec une puissance de 80%, c'est ce qu'il faut indiquer ici.

7 : les tests statistiques sont toujours bilatéraux.

8 : on clique sur « Calculez » pour calculer le nombre de sujets à inclure dans l'essai clinique, et l'on obtient ceci :

<sup>4</sup> Ce qui n'est pas très heureux, car normalement, le site aurait dû mettre une lettre grecque, comme il l'a fait pour «  $\mu$  »...

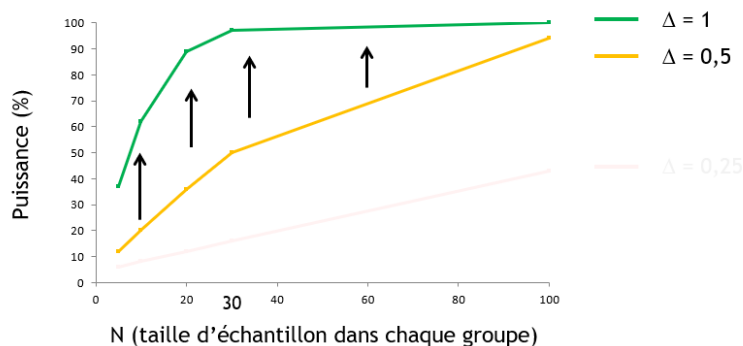
Le nombre de sujets à inclure dans un tel essai est donc de 44 chiens dans le groupe Prednisone et 44 chiens dans le groupe Budesonide, soit 88 chiens au total à recruter dans l'essai clinique.

7) Interprétez en une seule (et longue) phrase tous ces résultats présentés sur le site BiostaTGV (puissance, moyennes attendues du score clinique CIBDAI dans chacun des deux groupes, SD attendue, et taille d'échantillon dans chacun des deux groupes qui vient d'être calculée). Autrement dit, explicitez en une phrase l'information que le site BiostaTGV vous a fournie pour répondre à la question 6.

Si, dans la population des chiens avec une MICI sévère, la vraie moyenne du score CIBDAI après 6 semaines de traitement Prednisone est égale à 4,5 si la vraie moyenne du score CIBDAI après 6 semaines de traitement Budesonide est égale à 3, et si la SD du score CIBDAI vaut 2,5, alors il faudra inclure 44 chiens traités avec de la Prednisone et 44 chiens traités avec du Budesonide (soit au total 88 chiens) pour que l'essai clinique randomisé ait 80% de chances d'obtenir une différence significative de moyennes observées dans l'essai clinique qui sera mis en place.

8) Si, dans leurs hypothèses, l'équipe avait sous-estimé la différence réelle de score CIBDAI moyen entre les deux groupes de traitement au moment du calcul du nombre de sujets à inclure, avec un nombre de chiens recrutés dans chaque groupe égal à celui calculé dans la question 6, est-ce que l'essai clinique aura plus ou moins de 80% de chances d'obtenir une différence significative entre les deux scores CIBDAI moyens ?

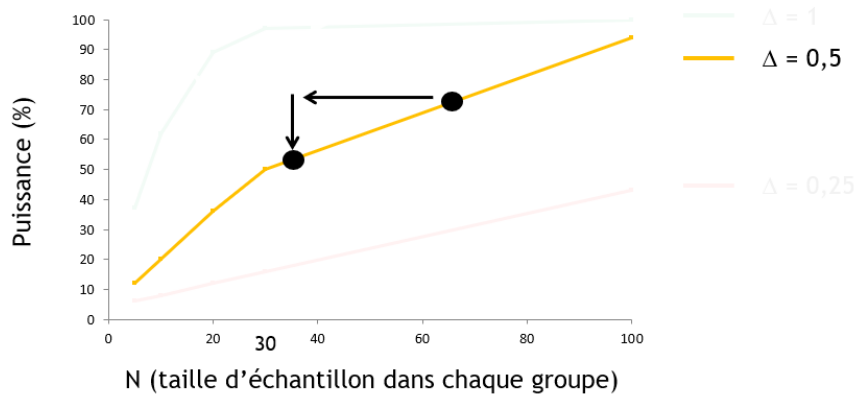
Nous avons vu dans les questions 6 et 7 que si la différence réelle de moyennes de score CIBDAI est égale à 1,5 (avec une SD de 2,5), alors avec 44 chiens dans chaque groupe, la puissance statistique de l'essai est de 80%. Si l'équipe avait sous-estimé la différence réelle de score CIBDAI moyen entre les deux groupes de traitement au moment du calcul du nombre de sujets à inclure, cela signifie qu'en fait, la différence réelle serait plus grande que celle envisagée par l'équipe (1,5 de différence). Dans cette situation-là, à taille d'échantillon égale, la puissance statistique serait plus élevée (cf. diapo du cours magistral) :



Donc l'essai clinique aurait plus de 80% de chances d'obtenir une différence significative entre les deux scores CIBDAI moyen si l'équipe avait sous-estimé la différence réelle de score CIBDAI moyen entre les deux groupes de traitement au moment du calcul du nombre de sujets à inclure.

9) Le nouvel essai clinique a été mis en place par la nouvelle équipe. Malheureusement, elle n'a pas réussi à recruter autant de chiens que le calcul le demandait (cf. question 6), mais 28 chiens « seulement » dans chaque groupe. L'essai clinique aura-t-il ainsi plus, ou moins, de chances que 80% d'obtenir une différence significative entre les deux scores CIBDAI moyens, avec les mêmes hypothèses énoncées dans la question 6 ?

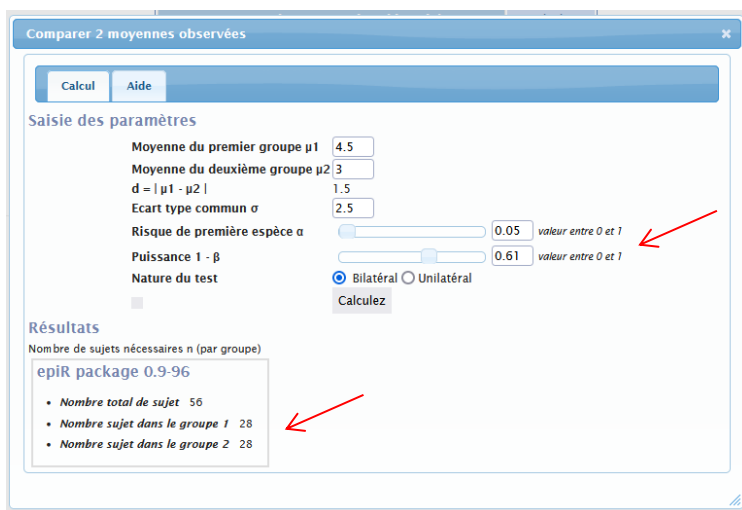
Si la taille de l'échantillon d'une étude diminue, à différence réelle égale (et SD égale), alors la puissance statistique de l'étude diminue (cf. diapo du cours magistral) :



Donc, l'essai clinique aura moins de chances que 80% d'obtenir une différence significative de scores CIBDAI moyens entre les deux traitements, si la différence réelle reste égale à 1,5 (et une SD égale à 2,5), avec 28 chiens dans chaque groupe plutôt que 44.

10) Utilisez le site Internet BiostaTGV pour connaître ce % de chances d'obtenir une différence significative entre les deux scores CIBDAI moyens, dans la situation décrite à la question précédente (utilisez la même page Internet BiostaTGV, et en tâtonnant).

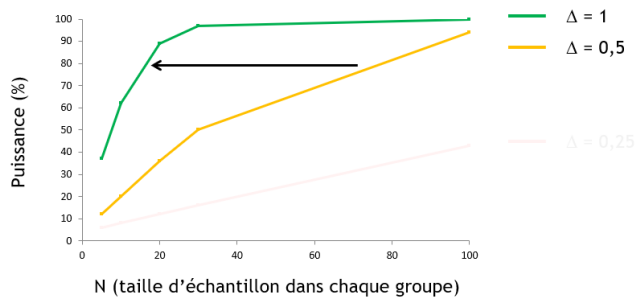
Pour connaître ce % de chances d'obtenir une différence significative entre les deux scores CIBDAI moyens, on réalise une simulation sur le site BiostaTGV, en tâtonnant sur la valeur de la puissance statistique afin d'obtenir 28 chiens dans chaque groupe :



Avec « seulement » 28 chiens dans chaque groupe, avec une différence réelle attendue de 1,5, et avec une SD de 2,5, alors la puissance de cet essai ne sera que de 61%. Autrement dit, avec « seulement » 28 chiens dans chaque groupe, avec une différence réelle attendue de 1,5, et avec une SD de 2,5, alors cet essai n'aura que 61% de chances d'obtenir une différence significative entre les deux scores CIBDAI moyens.

11) Pour que l'essai clinique ait malgré tout une puissance statistique de 80%, avec « seulement » 28 chiens dans chaque groupe, la valeur de la différence réelle entre les deux scores CIBDAI devrait-elle être plus importante ou moins importante que celle à laquelle les auteurs s'attendaient initialement (cf. question 6) ? En utilisant la même page Internet BiostaTGV, et en tâtonnant, trouvez quelle devrait être cette valeur de la différence réelle entre les deux scores CIBDAI.

Pour maintenir une puissance statistique de 80% avec moins d'animaux par groupe, il faut espérer que la différence réelle attendue soit plus élevée que celle initialement prévue (flèche noire ci-dessous à partir de l'une des diapos du cours magistral).



Pour connaître ce que devrait être cette valeur de la différence réelle avec 28 chiens par groupe pour avoir une puissance statistique de 80%, on réalise une simulation sur le site BiostaTGV, en tâtonnant sur la différence  $|\mu_1 - \mu_2|$  : soit en maintenant  $\mu_1$  égale à 4,5 et en modifiant la valeur de  $\mu_2$ , soit en maintenant  $\mu_2$  égale à 3 et en modifiant la valeur de  $\mu_1$  (tout en maintenant une SD de 2,5 et une puissance statistique de 80%), afin de maintenir 28 chiens dans chaque groupe :

Il faudrait que la différence réelle de moyennes de score CIBDAI soit égale à 1,9, et non plus 1,5 comme initialement attendue).