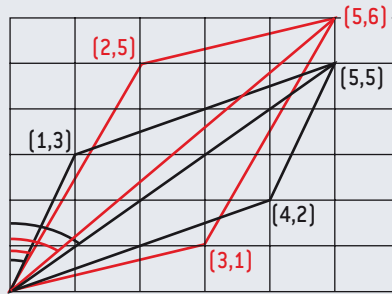
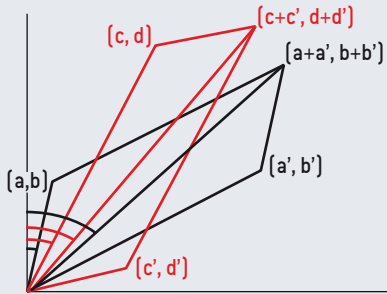


2. Interprétation graphique



Le paradoxe de Simpson apparaît lorsque $a/b < c/d$, $a'/b' < c'/d'$ et que néanmoins $(a + a')/(b + b') > (c + c')/(d + d')$. Les graphiques où les rapports a/b sont une fonction monotone des angles indiqués en fournissent la preuve graphique. Grâce à cette représentation, on trouve un exemple qui prouve que les trois inégalités, même dans des cas simples impliquant de petits nombres, peuvent être vraies simultanément.

Point de vue 1. Quand j'ignore si c'est un homme ou une femme (car par exemple je suis en train de traiter le cas d'un malade anonyme), je ne tiens compte que de la statistique générale qui est celle qui s'applique dans ce cas. Je donne donc le médicament, car je sais qu'il conduit à la guérison dans 50 % des cas d'après le tableau 1, alors que le placebo ne conduit à la guérison que dans 40 % des cas. En revanche, lorsque j'ai plus de précisions sur la personne à traiter et que je sais s'il s'agit d'un homme ou d'une femme, je regarde la statistique correspondante (donc le tableau 2 ou le tableau 3). Pour un homme, je donne le placebo, car la statistique du tableau 2 pour les hommes me conseille de donner le placebo. Pour une femme, je donne aussi le placebo, car la statistique du tableau 3 pour les femmes me dit que c'est préférable. L'information sur le sexe du patient me conduit à changer de prescription. Ce n'est pas absurde : l'apport de nouvelles informations justifie souvent de changer ses choix.

Mais à y regarder de près, ce point de vue met mal à l'aise et ne semble pas rationnel. C'est ce qu'exprime le second point de vue.

Point de vue 2. La conclusion du point de vue 1 est absurde. Un patient est soit un homme, soit une femme, et que ce soit l'un ou l'autre quand je connais le sexe du patient la consigne déduite est la même : donner le placebo. Ma décision ne dépend pas du résultat de la question : « S'agit-il d'un homme ou d'une femme ? » Il en résulte donc, de toute évidence, que même quand j'ignore le sexe du patient, je dois donner le placebo et ne pas

4

Yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	20	20	50 %
Placebo	16	24	40 %

5

Yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	20	20	50 %
Placebo	16	24	40 %

6

H.yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	19	13	59,375 %
Placebo	6	2	75 %

7

H.yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	17	11	60,71 %
Placebo	8	4	66,66 %

8

F.yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	1	7	12,5 %
Placebo	10	22	31,25 %

9

F.yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	3	9	25 %
Placebo	8	20	28,57 %

tenir compte de la statistique générale qui ne sert plus à rien dès l'instant où je dispose des deux statistiques particulières. On se trouve dans un cas assez usuel : la connaissance de nouvelles données (ici les statistiques particulières concernant séparément les hommes et les femmes) change la conclusion que je faisais avant d'en disposer. Ce changement est général et n'est pas dû à ma connaissance du sexe du patient, mais à ma connaissance des tableaux 2 et 3. Il n'y a rien d'irrationnel à changer d'avis quand on est mieux informé, certes, mais ici le changement doit conduire à oublier le tableau 1 qui ne sert plus à rien quand on a les deux autres.

Notons que, pour tirer une règle de données statistiques, on exige en général d'avoir des effectifs totaux plus élevés que ceux qui apparaissent dans nos tableaux. Mais cela est sans importance pour tout ce que nous venons de dire et pour tout ce que nous dirons plus loin, car on peut très bien imaginer que tous les nombres de nos tableaux sont multipliés par 100 ou même 10 000 et que les déductions que nous faisons des statistiques des tableaux sont de ce fait parfaitement sûres.

Le point de vue 2 est confirmé par un argument formel : de $(A \Rightarrow B)$ et $(\text{non-}A \Rightarrow B)$, on déduit B. C'est ce qu'on nomme parfois le « principe de la chose sûre » : si, lorsque A est vrai, j'en déduis B, et que, lorsque A est faux, j'en déduis aussi B, alors c'est que B est toujours vrai.

Le point de vue 2 correspond à l'analyse classique du paradoxe et semble nous en libérer. On l'exprime parfois en disant qu'il

faut prendre des précautions pour agréger des données. Je croyais satisfaisant ce point de vue, quand j'ai pris conscience d'une critique qui en montre l'absurdité.

Plus rien ne va !

Critique du point de vue 2. Imaginons que je dispose aussi d'une statistique sur les patients aux yeux clairs et d'une statistique sur les patients aux yeux foncés, et qu'elles mènent chacune à la conclusion que le médicament est plus efficace que le placebo.

Une telle situation est tout à fait possible et les données que nous proposons pour les tableaux 4 et 5 sont compatibles avec les trois premiers tableaux. C'est ce que démontrent les tableaux suivants 6, 7, 8 et 9 ; pour chacune des quatre catégories « Hommes aux yeux clairs », « Hommes aux yeux foncés », « Femmes aux yeux clairs », « Femmes aux yeux foncés », ils indiquent les taux de réussite respectifs du médicament et du placebo.

On vérifiera sans peine que les tableaux 6 et 7 redonnent 2, que 8 et 9 redonnent 3, que 6 et 8 redonnent 4, et enfin que 7 et 9 redonnent 5. Une situation comme celle envisagée ici est parfaitement possible.

Avec ces tableaux, le point de vue 2 est insoutenable. En effet, si j'adopte ce point de vue et que je l'applique non pas à la distinction homme-femme, mais à la distinction yeux clairs-yeux foncés, je dois conclure qu'il faut prescrire le médicament. Le point de vue 2 conduit donc, selon qu'on s'attache au sexe des patients ou à la couleur de leurs yeux, à des prescriptions contradictoires. Dans le cas décrit par l'ensemble des tableaux présentés, tous parfaitement compatibles, le point de vue 2 conduit à deux conclusions opposées ; il serait irrationnel de le défendre. En résumé, le point de vue 1 est clairement absurde, et le point de vue 2 aussi car il conduit à des prescriptions contradictoires. Comment sortir du paradoxe ?

Il semble inévitable d'opter pour un point de vue prudent et d'admettre que sans données plus précises, aucune prescription ne sera sûre. Ce point de vue 3, expliqué ci-dessous, ne satisfera pas le médecin qui n'a pas la possibilité de différer

3. Le «Double-Simpson»

Ce jeu de données, que nous nommons le « Double-Simpson », a été découvert par Jean-François Colonna le 20 mars 2013 en menant des calculs intensifs. Il a les propriétés paradoxales suivantes.

– Dans le tableau 1, on considère une population de 183 personnes séparées en deux groupes complémentaires dont l'un prend le médicament et l'autre un placebo. Le taux de guérison est meilleur pour le groupe prenant le médicament qui semble donc devoir être prescrit aux malades.

– Quand on sépare les hommes et les femmes (tableaux 2 et 3), pour chaque catégorie le placebo est meilleur que le médicament. C'est le paradoxe de Simpson simple.

– Quand on considère séparément les personnes aux yeux clairs et les personnes aux yeux foncés (tableaux 4 et 5), le médicament est à nouveau meilleur.

– De plus, et ce fut le point le plus délicat à satisfaire, les données des tableaux 1-5 résultent de deux séries de tableaux, soit A6-A7-A8-A9, soit B6-B7-B8-B9, aux propriétés opposées. Pour chacune des quatre catégories (« Hommes aux yeux

clairs », « Hommes aux yeux foncés », « Femmes aux yeux clairs », « Femmes aux yeux foncés »), selon qu'on lit les tableaux A ou B dans la même catégorie, on trouve que le médicament surpasse le placebo ou l'inverse, mais on n'a jamais le même résultat pour A et B.

Cette divergence systématique entre les données A et B montre que quelqu'un disposant des tableaux 1-5 ne peut rien conclure pour aucune personne quelle que soit sa catégorie (Homme, Femme, Yeux clairs, Yeux foncés).

A6

H. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	21	13	61,76 %
Placebo	8	5	61,53 %

A7

H. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	19	13	59,37 %
Placebo	8	4	66,66 %

A8

F. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	3	6	33,33 %
Placebo	13	23	36,11 %

A9

F. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	4	10	28,57 %
Placebo	9	24	27,27 %

B6

H. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	21	14	60 %
Placebo	9	4	69,23 %

B7

H. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	19	12	61,29 %
Placebo	7	5	58,33 %

B8

F. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	3	5	37,50 %
Placebo	12	24	33,33 %

B9

F. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	4	11	26,66 %
Placebo	10	23	30,30 %

1

Général	Guéri	Non guéri	Taux de guérison
Médicament	47	42	52,80 %
Placebo	38	56	40,42 %

2

Hommes	Guéri	Non guéri	Taux de guérison
Médicament	40	26	60,60 %
Placebo	16	9	64 %

3

Femmes	Guéri	Non guéri	Taux de guérison
Médicament	7	16	30,43 %
Placebo	22	47	31,88 %

4

Yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	24	19	55,81 %
Placebo	21	28	42,85 %

5

Yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	23	23	50 %
Placebo	17	28	37,77 %