

Quatre effets, phénomènes et paradoxes de la médecine

Leur signification et leurs racines historiques

Peter Kleist

PFC Pharma Focus AG, Volketswil



Quintessenz

En tant que lecteurs de littérature médicale, vous devriez éviter de vous laisser entraîner en terrain mouvant par les quatre phénomènes suivants:

- L'effet Hawthorne: la modification des comportements naturels de sujets d'étude en raison de leur participation à cette dernière peut entraîner une surévaluation des effets du traitement, en particulier dans le groupe contrôle.
- La régression à la moyenne: l'inclusion de patients avec des valeurs très élevées ou très basses à l'entrée dans une étude donne l'illusion que la variabilité statistique des mesures ultérieures est une amélioration sous traitement.
- Le paradoxe de Simpson: lorsque des facteurs déterminants (in)connus («confounders») influencent les données, le résultat global d'une étude (de cas témoins) peut être complètement modifié par des analyses de sous-groupes.
- Le phénomène de Will Rogers: l'amélioration des possibilités diagnostiques ou l'augmentation artificielle de la prévalence d'une maladie peut améliorer le pronostic d'un patient sans que ses paramètres de mesure n'aient subi un quelconque changement.

La réalisation d'études randomisées contrôlées, dont le protocole décrit clairement la procédure, est le meilleur moyen pour éliminer ces quatre phénomènes.

Summary

Four effects, phenomena and paradoxes in medicine. Their relevance and historical roots

The reader of medical literature must beware of the following four phenomena:

- *Hawthorne effect: involvement in a study may change a study participant's normal behaviour and result in overestimation of treatment effects, particularly in the control group.*
- *Regression to the mean: when patients are enrolled in a study with extremely high or low values, the statistical variability of repeat measurements may look like a true treatment effect.*
- *Simpson's paradox: if an (un)known confounder influences the result of a (case control) study, subgroup analyses may reverse the effects shown in the overall study results (comparison of cases and controls).*
- *Will Rogers phenomenon: a patient's mean prognosis may improve merely due to more sensitive diagnostic tools or an artificial increase in disease prevalence, without any change in the patient's measurement values.*

The conduct of randomised and controlled studies on the basis of a specified methodological protocol is the most successful means of eliminating or neutralising these four phenomena.

Introduction

Les données ne mentent jamais – vraiment? En tout cas, elles peuvent parfois masquer la vérité. Comme vous le savez certainement, différentes formes d'un biais peuvent entraîner des distorsions dans les résultats des études cliniques. Cet article a pour objectif de vous familiariser avec la signification clinique, agrémentée de quelques anecdotes historiques, de quatre «particularités» choisies et peut-être moins connues, qui peuvent nous induire en erreur lors de l'interprétation des données cliniques: l'effet Hawthorne, le phénomène de la régression à la moyenne, le paradoxe de Simpson et le phénomène de Will Rogers.

L'effet Hawthorne

L'effet Hawthorne dit ce qui suit: les sujets d'une étude modifient leur comportement naturel par le simple fait qu'ils savent qu'ils participent à une étude et qu'ils sont sous observation. L'effet Hawthorne a fait l'objet d'une étude spécifique ayant inclus deux groupes de patients à la veille d'une opération du genou: les patients qui ont été informés qu'ils allaient être inclus dans un essai scientifique sur leur état de santé après l'intervention, se sentaient d'une manière générale plus vite mieux et présentaient nettement moins de douleurs postopératoires que les patients à qui l'on n'avait rien dit [1].

Une étude comparable a été réalisée avec des patients souffrant de douleurs tumorales légères à modérées. Ceux-ci ont reçu, dans le cadre d'un design en cross-over, du naproxène ou un placebo. Si les patients savaient qu'ils étaient participants à une étude, le placebo avait des effets antalgiques subjectifs plus puissants que le naproxène chez les sujets qui pensaient recevoir l'antalgique actif dans le cadre de la procédure thérapeutique normale, ignorant donc qu'ils participaient à un essai clinique [2]. Dans d'autres études, la réaction individuelle aux hypnotiques, à la caféine ou aux chewing-gums contenant de la nicotine et aux placebos correspondants dépendait largement du fait que les patients étaient informés ou non de leur participation à un test clinique [3].

Cet effet concerne a priori tous les participants à une étude clinique, autrement dit aussi bien ceux qui sont inclus dans le groupe interventionnel que ceux du groupe contrôle. Si les résultats obtenus dans un groupe contrôle sont supérieurs à la moyenne suite à un effet Hawthorne, la mise en évidence d'un effet dans le groupe interventionnel peut être rendue plus difficile, voire impossible, car la marge de manœuvre pour démontrer l'existence de la supériorité d'une mesure thérapeutique devient extrêmement petite – c'est par exemple le cas pour un produit de soins dentaires, puisque tous les patients participant à l'étude s'astreignent à une hygiène bucco-dentaire meilleure que dans leur vie de tous les jours [4]. En termes d'information générale aux patients, cet effet ne constitue cependant pas uniquement un inconvénient, puisqu'après tout tous les patients en profitent – ne serait-ce qu'en participant à l'étude.

L'effet Hawthorne ne concerne en fait pas seulement les sujets des tests, mais aussi les responsables des études eux-mêmes: la participation aux études observationnelles a par exemple entraîné une restriction significative des indications à la prescription d'antibiotiques chez les enfants atteints d'infections des voies respiratoires [5] et une diminution significative du nombre des hospitalisations des enfants souffrant d'asthme [6], à la suite d'un renforcement du traitement de la maladie sur la base d'évidence par les médecins participants aux études. Dans une étude multicentrique randomisée sur des patients insuffisants rénaux, un protocole spécifique pour le traitement de l'anémie n'a montré aucun avantage par rapport au traitement local standard – et ceci en raison d'une constante amélioration du traitement des patients du groupe comparatif tout au long de l'essai [7]. Par rapport à des patients extérieurs à l'étude, les patients du groupe contrôle des essais sur la thrombolyse post-infarctus présentent même des taux de survie meilleurs, qui ne peuvent pas s'expliquer uniquement par l'inclusion de patients ayant un meilleur pronostic [8].

Conclusion

L'effet Hawthorne peut conduire à une surestimation de l'efficacité dans les groupes contrôle des essais cliniques et rendre plus difficile, voire impossible, la mise en évidence de l'efficacité d'un traitement. Des traitements standards non spécifiés dans les groupes de comparaison d'études contrôlées ou dans le cadre d'études d'observation ne sont souvent pas représentatifs d'un traitement standard dans des conditions de tous les jours.

Note historique

L'existence d'un effet Hawthorne a été décrite à la suite d'essais effectués dans les usines de Hawthorne de la Western Electric Company aux

Etats-Unis [9]. Les études réalisées entre 1924 et 1927 devaient montrer si des modifications simples des conditions de travail pouvaient avoir des effets sur la productivité des ouvriers. La productivité a augmenté, pour ne citer que cet exemple, en conditions d'éclairage aussi bien améliorées que péjorées – même si les ampoules étaient remplacées par de nouvelles, à la puissance lumineuse identique.

Ce n'étaient cependant pas les conditions extérieures qui étaient décisives pour les résultats, mais bien la participation à l'étude en soi, c'est-à-dire l'attention particulière des sujets et leur impression que quelque chose s'était passé.

Le phénomène de la régression à la moyenne

Le phénomène de la régression à la moyenne joue un rôle lorsque les paramètres mesurés ne donnent pas de valeurs constantes, mais indiquent une variabilité, par ex. en raison de variations intra-individuelles, d'adaptations circonstancielles ou d'imprécisions dans les mesures.

C'est par exemple le cas de mesures de la tension artérielle dispersées aux alentours d'une valeur moyenne. La variabilité des mesures – aussi bien chez un individu que lors de mesures successives au sein d'un groupe de patients – suit une distribution normale. Or, la sélection des participants à une étude se fait souvent sur la base de valeurs extrêmes, c'est-à-dire nettement augmentées ou diminuées au départ. Ces valeurs ne correspondent plus à une distribution normale, mais sont plutôt groupées à une extrémité de la courbe de distribution (variabilité) de la population étudiée. Si la «vraie» valeur avait été connue, on n'aurait très probablement pas inclus ces patients dans l'essai. Les mesures répétées effectuées au cours de l'étude chez les patients sélectionnés sur la base d'une valeur de tension artérielle une fois élevée a toutes les chances de donner des valeurs proches de la valeur moyenne individuelle (inconnue) ou de la moyenne du groupe, soit des valeurs plus basses qu'au moment de l'inclusion [10].

Un exemple hypothétique illustre très bien ce point: on mesure la tension artérielle chez 250 patients hypertendus dans le cadre d'un screening en vue de la participation à un essai clinique. La moyenne du groupe dont la distribution des valeurs est normale est de 90 mm Hg. Une nouvelle mesure de contrôle donnerait à nouveau une valeur moyenne proche des 90 mm Hg pour le groupe, car les chiffres individuels plus élevés chez certains sujets et plus bas chez d'autres par rapport à la première mesure tendraient à s'équilibrer sur l'ensemble du collectif.

Comme le protocole de l'étude planifiée a pour but de ne sélectionner que des patients dont la tension diastolique est supérieure à 95 mm Hg,

seulement 80 patients des 250 ayant participé au screening pourront finalement être inclus. La tension artérielle moyenne de ces mesures, qui sont celles d'un collectif choisi et qui ne relèvent donc plus d'une distribution normale, se situe maintenant à 100 mm Hg. Très probablement on a relevé dans une proportion assez importante de ces 80 patients des «valeurs extrêmement élevées».

En introduisant un traitement antihypertenseur chez ces 80 participants à l'étude et en mesurant une nouvelle fois la tension artérielle quatre semaines plus tard, la tension diastolique moyenne se situera peut-être cette fois vers 95 mm Hg, soit 5 mm Hg au-dessous de la valeur initiale. Mais: si on avait effectué chez ces patients la mesure de contrôle avant de commencer le traitement, on aurait également constaté une diminution de la tension artérielle et une partie d'entre eux n'auraient probablement plus répondu aux critères d'inclusion de >95 mm Hg.

La diminution de la tension artérielle observée n'est cependant ici que le fruit d'un phénomène statistique, c'est-à-dire la «régression à la moyenne», autrement dit «regression to the mean», qui donne l'apparence d'une amélioration vraie à une simple variabilité naturelle des mesures répétées [11]. Plus une valeur (extrême) s'écarte de la valeur moyenne, plus la probabilité que la valeur trouvée lors d'une mesure contrôle soit plus proche de la moyenne est élevée. Ceci explique aussi pourquoi l'action des antihypertenseurs paraît renforcée lorsque le degré de sévérité de l'hypertension augmente. Plus les valeurs initiales sont élevées, plus la baisse de tension artérielle sera marquée – un effet qui se retrouvera néanmoins dans les mêmes proportions sous placebo: c'est ainsi que l'on a observé dans l'«Australian Mild Hypertension Study» une diminution moyenne de la tension diastolique de plus de 9 mm Hg chez les patients du groupe placebo qui avaient les valeurs initiales les plus élevées [12].

La signification du phénomène de la régression à la moyenne est loin d'être négligeable et peut se rencontrer dans pratiquement tous les domaines de la médecine clinique. Ainsi, un (petit) groupe de patients perdant de la masse osseuse durant la première année d'observation malgré un traitement efficace a de grandes chances de voir augmenter sa masse osseuse l'année suivante, sans modifier le traitement et uniquement en raison du phénomène de la régression à la moyenne [13]. L'amélioration marquée des symptômes allergiques sévères dans une étude d'observation en protocole ouvert ne peut s'expliquer uniquement par l'action d'un antihistaminique. Elle est assurément aussi l'expression de la régression à la moyenne [14]. Et l'effet du vaccin contre la méningite C (réduction jusqu'à 90% des cas de méningite en Grande-Bretagne) est très certainement surestimé, dans la mesure

où la vaccination a été effectuée au cours d'une période où l'incidence de la maladie était particulièrement élevée [13, 15].

Conclusion

Les études observationnelles non contrôlées avec des patients sélectionnés ayant des valeurs initiales élevées ou basses surestiment l'effet thérapeutique réel – l'inclusion d'un groupe contrôle prend ici toute son importance. Les distorsions de type «regression to the mean» peuvent alors être neutralisées par l'effet de soustraction de l'action du traitement testé et du traitement comparatif. Les améliorations observées sous placebo sont – en partie du moins – aussi dues à une régression à la moyenne [16]. S'il existe un risque de phénomène de type «regression to the mean», l'inclusion de patients dans les études cliniques devrait être effectuée sur la base de mesures initiales multiples.

Note historique

Nous devons le terme de «regression to the mean» au généticien et statisticien anglais Sir Francis Galton (1822–1911), cousin de second degré de Charles Darwin et «inventeur» de l'identification par empreintes digitales. Il a en effet observé, lors de la mesure de la taille de plus de 900 enfants adultes et de leurs parents, que les parents de très grande taille avaient en règle générale des enfants de plus petite taille par rapport à la leur (qui est cependant tout de même au-dessus de la moyenne), alors que les parents de très petite taille ont habituellement des enfants qui sont plus grands qu'eux-mêmes (quoique restant plus petits que la moyenne) [17].

Le paradoxe de Simpson

Il s'agit d'un phénomène lié à différentes méthodes d'analyses statistiques. Selon que l'on considère le groupe global ou des sous-groupes du collectif de participants à une étude, on peut aboutir à des résultats contradictoires. Le paradoxe survient principalement dans les études d'observation non randomisées (notamment les études de cas témoins), mais aussi dans le cadre de calculs non valides du «number needed to treat» (NNT) dans les méta-analyses.

La fumée du tabac est bonne pour la santé. C'est la conclusion à laquelle on aurait pu parvenir en établissant la relation entre le tabagisme et le taux de survie à long terme dans un collectif de 1300 femmes britanniques. Ces femmes ont participé, entre 1972 et 1974, à une enquête établissant si elles fumaient ou non. Vingt ans plus tard, on a simplement considéré si elles étaient encore en vie – avec le résultat suivant: 31% des non fumeuses étaient décédées dans l'intervalle, contre 24% seulement des fumeuses.

Cette manière d'aborder le problème masque cependant l'effet de l'âge des participantes. Ce n'est que dans les années 1960 et 1970 que les femmes se sont mises à fumer et le phénomène touchait essentiellement des femmes jeunes. Dans ce groupe de fumeuses, proportionnellement très petit, la répartition des groupes d'âge était de plus très inhomogène. La répétition de cette même analyse en tenant compte de l'âge révèle une mortalité augmentée des fumeuses de tous les groupes d'âge, à une exception près [18].

Un autre exemple d'étude de cas témoins illustre encore mieux le phénomène [19]: une étude rétrospective a comparé deux techniques d'ablation de calculs rénaux: la chirurgie à ciel ouvert (données des années 1972–1980) et de la néphrolithotomie percutanée (données des années 1980–1985). La chirurgie à ciel ouvert a eu un taux de succès de 78% (273 patients sur 350) contre 83% (289 patients sur 350) pour la néphrolithotomie, qui semblait donc plus efficace. Lorsque ces données étaient cependant reconsidérées en tenant compte de la taille des calculs (< ou ≥ 2 cm), ce résultat prend cependant une toute autre allure: la chirurgie à ciel ouvert est plus efficace que la néphrolithotomie pour les lithiases de petite taille (93 vs. 83%, autrement dit 81/87 patients vs. 234/270 respectivement), ainsi que pour ceux de grande taille (73 vs. 69%, soit 192/263 patients vs. 55/80). Vous pouvez tranquillement refaire les calculs – il n'y a pas d'erreur. Cette inversion des résultats est appelée paradoxe de Simpson [20].

Que nous montrent ces exemples? Dans les évaluations globales des deux groupes, un facteur déterminant (confounder) a été complètement négligé (âge des femmes ou choix de la technique par le médecin traitant en fonction de la taille de la lithiase). S'ajoute encore le fait que la taille des sous-groupes analysés, ainsi que les prévalences (proportion de fumeuses) et les incidences (taux de succès de l'ablation des calculs rénaux) différaient nettement. Pour cette raison, il n'est pas non plus possible de calculer un NNT sur la base de la somme de tous les événements, dans le cadre d'une méta-analyse, si la taille des groupes et les résultats des études prises individuellement varient beaucoup entre eux.

Un autre exemple va encore préciser ce point: si les études incluses dans une méta-analyse ont pour la plupart montré un avantage en faveur du sevrage du tabac en milieu institutionnel par rapport aux programmes sans soutien personnalisé, le taux global de patients ayant arrêté de fumer était de 14,3% (546/3820 patients) dans les groupes interventionnels contre 15,5% (356/2301 patients) dans les groupes contrôles [21].

Conclusion

Des facteurs déterminants inconnus peuvent fausser les résultats des études d'observation, en

particulier des études de cas témoins. Un nombre élevé de cas et des niveaux de signification statistiques n'offrent par ailleurs aucune protection contre les conclusions erronées [22]. En présence d'un paradoxe de Simpson, nous ne savons en réalité pas la vérité: en effet, les analyses des sous-groupes peuvent elles-mêmes être influencées par une autre variable inconnue, qui, si l'on en tient compte, pourrait à nouveau inverser le résultat. C'est pourquoi il convient de choisir un protocole d'étude randomisé et contrôlé chaque fois que cela est possible. Dans les méta-analyses publiées, on veillera à la méthode utilisée pour calculer le NNT.

Contexte historique

Le nom est attribué au statisticien américain Edward Hugh Simpson, qui a décrit ce paradoxe en 1951 [23]. A tort, car c'est bien auparavant que le mathématicien britannique Karl Pearson (1899) et le statisticien écossais George Udny Yule (1903) avaient attiré l'attention sur ce problème statistique [24].

La reconnaissance de l'existence de ce paradoxe a été largement renforcée après que l'Université de Berkeley ait été accusée, au cours des années 1970, de désavantager les femmes dans les admissions aux études. Effectivement, alors que 44% des candidats masculins étaient admis, 34% seulement des femmes étaient acceptés. Une évaluation détaillée au niveau des facultés a néanmoins montré qu'il y avait proportionnellement davantage de femmes admises que d'hommes dans pratiquement tous les domaines. Alors, où se trouve le «confounder»? Les femmes se portaient de préférence candidates dans des branches où le taux de refus était élevé, alors que les hommes s'intéressaient principalement à des domaines où un grand nombre d'étudiants étaient admis [25].

Le phénomène de Will Rogers

Le pronostic d'un patient peut-il s'améliorer alors que rien n'a changé dans son état de santé, ni dans ses paramètres de mesure? Oui – et cela en raison du phénomène de Will Rogers, autrement dit de glissement de classification ou «stage migration effect». Les statisticiens parlent parfois aussi d'«association criminelle (des données)». Au niveau médical, ce phénomène a été décrit pour la première fois en oncologie: une étude américaine a comparé deux cohortes de patients atteints d'un cancer du poumon. La cohorte la plus jeune, née dans les années 1970 et 80, avait un meilleur taux de survie à six mois que la plus âgée, née dans les années 1950 et 60. La distribution des stades du cancer (classification TNM) était identique dans les deux cohortes. L'amélioration apparente du pronostic n'était cependant pas à mettre sur le compte d'une op-

timisation du traitement, mais à de nouvelles techniques d'imagerie apparues dans l'inter-
valle. La meilleure sensibilité des méthodes de
dépistage des métastases a eu pour conséquence
qu'une partie de la cohorte la plus jeune s'est
vue attribuer un stade plus avancé de la maladie,
par rapport à la cohorte des patients plus âgés.
En d'autres termes, le pronostic s'est amélioré,
aussi bien dans les stades plus précoces du can-
cer (puisque les patients dont le pronostic était
plus réservé migrerait vers le stade suivant) que
dans le stade avancé (parce que les patients clas-
sés au stade avancé présentaient en moyenne
un meilleur pronostic que ceux qui pouvaient
être plus facilement attribués à ce stade) [26].
Des études de population publiées récemment
ont également mis en évidence ce phénomène
de l'amélioration du pronostic – dû à une amé-
lioration des méthodes diagnostiques et à une
migration des stades de maladie – dans le can-
cer de la prostate [27] et dans le cancer colorec-
tal [28].

Le phénomène de Will Rogers joue aussi un rôle
lorsque la prévalence d'une maladie change à
la suite d'un élargissement de sa définition.
L'abaissement de valeurs seuils considérées
comme pathologiques de la glycémie, de la ten-
sion artérielle et du cholestérol plasmatique aug-
mente la population diabétique, hypertendue
et hyperlipidémique. L'inclusion de patients avec
un meilleur pronostic suffit à améliorer le pro-
nostic global moyen de l'ensemble des patients
de ces groupes de population [29].

Conclusion

Ne faites pas aveuglément confiance aux résul-
tats d'études utilisant des groupes contrôles
historiques. Les progrès réalisés dans le diagnos-
tic ou une augmentation artificielle de la préva-
lence d'une maladie peuvent simuler une amé-
lioration du pronostic, qui ne doit pas être
attribuée à une modification des standards de
traitement ou à un nouveau médicament apparu
sur le marché [30].

Bibliographie recommandée

- Bland JM, Altman DG. Regression towards the mean. *BMJ*. 1994;308:1499.
- Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ*. 1994;309:1480–1.
- Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenom-
enon. Stage migration and new diagnostic techniques as a
source of misleading statistics for survival in cancer. *N Engl
J Med*. 1985;312:1604–8.

Contexte historique

C'est l'humoriste et animateur américain Will
Rogers qui a donné son nom au phénomène
décrit ci-dessus. Durant la crise des années
1930, de nombreux émigrants de l'état d'Okla-
homa (dit «okies») se sont établis en Californie.
Rogers, qui était lui-même né en Oklahoma en
1879, ne tenait pas en grande estime les émi-
grants et encore moins les californiens. On lui at-
tribue ainsi ces mots [31]: «When the Okies left
Oklahoma and moved to California, they raised
the average intellectual level in both states.»
(Lorsque les okies ont quitté l'Oklahoma pour
s'établir en Californie, ils ont augmenté le niveau
intellectuel des deux Etats.).

Conclusion

L'interprétation des données cliniques n'est
parfois pas simple et ce qui semble manifeste ne
reflète pas toujours la vérité. Comme lecteurs de
littérature médicale, vous devez connaître un
certain nombre de pièges potentiels, considérer
les résultats des études avec un œil critique et en
tirer les conclusions qui en découlent. Les études
d'observation non contrôlées, les études de cas
témoins, les comparaisons historiques et les
groupes témoins sans procédures standardisées
posent plus particulièrement problème et peu-
vent présenter des points faibles – un motif
essentiel pour nous de fonder l'évidence médi-
cale, dans toute la mesure du possible et dans
les limites éthiques défendables, sur des études
randomisées, contrôlées et ayant un protocole
clairement défini.

Dans la mesure où tout ce qui est publié ne ré-
pond pas toujours à l'ensemble des critères
de qualité requis, il est nécessaire d'augmenter
le «NNR», c'est-à-dire le «number (of studies)
needed to read», pour pouvoir répondre de façon
valable à une question clinique.

Cette formation continue est également agréée
par l'Association Suisse de Médecine Pharma-
ceutique (SGPM-ASMP).

- Fey MF, Bühner A. Lecture critique de la littérature clinique
professionnelle. Partie I: Chasse-trapes dans les rapports de
cas et les séries de cas. *Forum Med Suisse*. 2001;1(7):161–5.

Vous trouverez la bibliographie complète [1–31] dans la
version en ligne de l'article sous [www.medicalforum.ch/pdf/
pdf_f/2006/2006-46/2006-46-194.pdf](http://www.medicalforum.ch/pdf/pdf_f/2006/2006-46/2006-46-194.pdf).

Correspondance:
Dr Peter Kleist
PFC Pharma Focus AG
Chriesbaumstrasse 2
CH-8604 Volketswil
peter.kleist@pfc.ch

Quatre effets, phénomènes et paradoxes de la médecine

Leur signification et leurs racines historiques

Peter Kleist

PFC Pharma Focus AG, Volketswil

Références

- De Amici D, Klersy C, Ramajloji F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials*. 2000;21:103–14.
- Bergmann JF, Chassany O, Gandiol J, Deblois P, Kanis JA, Segrestaa JM, et al. A randomized clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. *Clin Trials Meta-Anal*. 1994; 29: 41–7.
- Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial. Gold standard or golden calf? *J Clin Epidemiol*. 2001; 54: 541–9.
- Winston JL, Bartizek RD, McClanahan SF, Mau MS, Beiswanger BB. A clinical methods study of the effects of triclosan dentifrices on gingivitis over six months. *J Clin Dent*. 2002;13:240–8.
- Mangione-Smith R, Elliott MN, McDonald L, McGlynn EA. An observational study of antibiotic prescribing behavior and the Hawthorne effect. *Health Serv Res*. 2002;37:1603–23.
- Zinman R, Bethune P, Camfield C, Fitzpatrick E, Gordon K. An observational asthma study alters emergency department use: the Hawthorne effect. *Pediatr Emerg Care*. 1996;12:78–80.
- Brimble KS, Rabbat CG, McKenna P, Lambert K, Carlisle EJ. Protocolized anemia management with erythropoietin in hemodialysis patients: a randomised controlled trial. *J Am Soc Nephrol*. 2003;14:2654–61.
- Jha P, Deboer D, Sykora K, Naylor CD. Characteristics and mortality outcomes of thrombolysis trial participants and nonparticipants: a population-based comparison. *J Am Coll Cardiol*. 1996;27:1335–42.
- Roethlisberger FH, Dickson WJ. *Management and the worker*. Cambridge, MA: Harvard University Press; 1939.
- Bland JM, Altman DG. Regression towards the mean. *BMJ*. 1994;308:1499.
- Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. 2005;34:215–20.
- The Australian therapeutic trial in mild hypertension. Report by the management committee. *Lancet*. 1980;i:1261–7.
- Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ*. 2003;326:1083–4.
- Bachert C, Vovolis V, Margari P, Murrieta-Aguttes M, Santoni JP. Mizolastine in the treatment of seasonal allergic rhinoconjunctivitis: a European clinical experience with 5408 patients managed in daily practice. *Allergy*. 2001;56:653–9.
- Wise J. News roundup: meningitis C rates show steep fall. *BMJ*. 2001;322:70.
- Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med*. 2001;344:1594–602.
- Galton F. Regression towards mediocrity in hereditary stature. *J R Anthropol Inst*. 1886;15:246–63.
- Appleton DR, French JM, Vanderpump MPJ. Ignoring a covariate: an example of Simpson's paradox. *Am Stat*. 1996;50:340.
- Charig CR, Webb DR, Payne SR, Wickham OE. Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. *BMJ*. 1986;292:879–82.
- Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ*. 1994;309:1480–1.
- Cates CJ. Simpson's paradox and calculation of number needed to treat from meta-analysis. *BMC Med Res Methodol*. 2002;2:1. Epub 2002 Jan 25.
- Heydtmann M. The nature of truth: Simpson's paradox and the limits of statistical data. *QJM*. 2002;95:247–9.
- Simpson EH. The interpretation of interaction in contingency tables. *J R Statist Soc B*. 1951;2:238–41.
- Tu Y-K, West R, Ellison GTH, Gilthorpe MS. Why evidence for the fetal origins of adult disease might be a statistical artefact: The "reversal paradox" for the relation between birth weight and blood pressure in later life. *Am J Epidemiol*. 2005;161:27–32.
- Bickel PJ, Hjammel EA, O'Connell JW. Sex bias in graduate admissions: data from Berkeley. *Science*. 1975; 187: 398–404.
- Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med*. 1985;312:1604–8.
- Albertsen PC, Hanley JA, Barrows GH, Penson DF, Kowalczyk PD, Sanders MM, Fine J. Prostate cancer and the Will Rogers phenomenon. *J Nat Cancer Inst*. 2005;97:1236–7.
- Shahrier M, Ahnen DJ. Colorectal cancer and survival in Europe: the Will Rogers phenomenon revisited. *Gut*. 2000;47:463–4.
- Schwartz LM, Woloshin S. Changing disease definitions: Implications for disease prevalence. Analysis of the third national health and nutrition examination survey, 1988–1994. *Eff Clin Pract*. 1999;2:76–85.
- Fey MF, Bühler A. Lecture critique de la littérature clinique professionnelle. *Forum Med Suisse*. 2001;1(7):161–71.
- Vineis P, McMichael AJ. Bias and confounding in molecular epidemiological studies: special considerations. *Carcinogenesis*. 1998;19:2063–7.

Correspondance:

Dr Peter Kleist

PFC Pharma Focus AG

Chriesbaumstrasse 2

CH-8604 Volketswil

peter.kleist@pfc.ch