

# Power:

## How to Use it in Evidence-Based Practice

# Utilisation de la puissance dans la pratique fondée sur les données probantes

BY / PAR

SANDRA ENGBERG  
PHD RN CRNP  
FAAN

AND / ET

ELIZABETH A.  
SCHLENK PHD RN

### Introduction

**I**n this fictitious study, the investigators compared the time to healing of stage 3 pressure ulcers in subjects treated with the new dressing (XYZ dressing). Fifty patients were recruited into the study and randomly assigned to the new dressing or a usual care control group. The investigators reported that there was no significant difference in the time to wound healing in the XYZ and control groups ( $p=0.15$ ). They concluded that the new dressing was no more effective than usual care (the dressing currently used on the unit). Should you accept their conclusion?

In answering this question, that is, deciding whether there was really no difference in the effectiveness of the 2 dressings, you need to determine if the study was adequately powered. Within the context of research, power is the ability of a statistical test to detect a difference when one exists in the target population.<sup>1</sup> The target population is the entire group of persons in which you are interested in applying the findings of a study. The goal of research is to use data from a sample to make inferences about the target population, while minimizing the likelihood of making erroneous conclusions about that population.<sup>2</sup>

Statistical tests such as the *t* test, analysis of variance or regression are used to test a hypothesis in a sample of the target population. Typically, 2 hypotheses are considered. One is the null hypothesis, the hypothesis that there is no difference. Testing for statistical significance is based on the null hypothesis.<sup>3</sup> In a study com-

### Introduction

**P**renons pour exemple une étude fictive, au cours de laquelle les investigateurs ont déterminé le délai de cicatrisation des plaies de pression de stade 3 chez des sujets traités au moyen d'un nouveau pansement (pansement XYZ). Cinquante patients ont été inscrits à l'étude et répartis au hasard entre deux groupes : un groupe traité au moyen du nouveau pansement et un groupe témoin recevant les soins habituels. Les investigateurs n'ont pas observé de différence significative entre les deux groupes pour ce qui est du délai de cicatrisation des plaies ( $p = 0,15$ ). Ils ont conclu que le nouveau pansement n'était pas plus efficace que les soins habituels (le pansement actuellement utilisé au service). Devez-vous accepter leur conclusion?

Pour répondre à cette question, c'est-à-dire pour savoir s'il n'y avait vraiment pas de différence d'efficacité entre les deux pansements, il faut déterminer si l'étude avait la puissance voulue. Dans le contexte d'un projet de recherche, la puissance désigne la capacité d'un test statistique de détecter une différence quand il y en a une dans la population cible<sup>1</sup>. La population cible est l'ensemble des personnes auxquelles on désire appliquer les résultats d'une étude. Le but d'un projet de recherche est d'utiliser les données provenant d'un échantillon pour tirer des conclusions sur la population cible, tout en réduisant au minimum la possibilité que ces conclusions soient erronées<sup>2</sup>.

Les tests statistiques, comme le test *t* et l'analyse de la variance ou de la régression, sont utilisés pour véri-

**Sandra Engberg**  
and **Elizabeth A.  
Schlenk** are with  
the University of  
Pittsburgh School of  
Nursing in Pittsburgh,  
Pennsylvania.

**Sandra Engberg**  
et **Elizabeth A.  
Schlenk** sont  
de l'École des  
sciences infirmières  
de l'Université de  
Pittsburgh, à Pittsburgh,  
en Pennsylvanie.

This article was originally published in the *Journal of Wound Ostomy and Continence Nursing*. It is reprinted here because it is well-written, understandable, comprehensive and relevant to *Wound Care Canada* readers. It is a natural progression in the series of columns entitled "Understanding the Scientific Literature" as it explains power in terms of concepts that have been covered in columns published this year. This article indicates why clinicians need to understand the statistical power of a research project and why they need to see a power calculation or sample size determination when reading an article, particularly if no effect has been detected.

Reprinted with permission from: Engberg S, Schlenk EA. Power: How to use it in evidence-based practice. *J Wound Ostomy Continence Nurs.* 2010;37(1):14-16. Published by Lippincott Williams & Wilkins.

paring characteristics in 2 groups, the null hypothesis would be that there is no difference in the characteristic(s) of interest in the groups. In a study examining an intervention, the null hypothesis would be that there is no difference in the outcome before and after the intervention (i.e. a 1-group study) or no difference in the outcome between the treatment and control groups. The second hypothesis is the alternative hypothesis, the hypothesis that there is a difference, for example, a difference in characteristic(s) in the groups or in the outcome following an intervention. In the fictitious study, the null hypothesis is that there is no difference in the time to wound healing between the group receiving XYZ dressing and the control group and the alternative hypothesis is that there is a difference in the time to wound healing between the 2 groups.

When researchers interpret the results of statistical tests in their studies, they can make 2 different types of errors: a type I or type II error (Table 1). A type I error occurs when investigators conclude, based on the  $p$  value, that there was a significant relationship or treatment effect between their groups when none exists in the target population (i.e. the null hypothesis is true). Investigators control the probability of making a type I error when they select the alpha ( $\alpha$ ) that they will utilize to determine whether the results of their statistical tests are significant. Typically,  $\alpha$  is set at 0.05, which means that the investigators are willing to accept a 5% probability (1 chance in 20) that they made a type I error.

The other type of error is a type II or beta ( $\beta$ ) error. A type II error is made when the investigators, based on a nonsignificant  $p$  value, accept the null hypothesis that there was no relationship or no treatment effect when one exists in the target population. Power ( $1 - \beta$ ) is the probability that if a difference exists in the target population it will be detected by the statistical test performed. If, for example,  $\beta$  is set at 0.10, the investigators are willing to accept a 10% chance of not detecting a significant effect if it exists. This  $\beta$  of 0.10 means that the power to detect a treatment effect is 90%.

Ideally, investigators would like the probability of making both type I ( $\alpha$ ) and type II ( $\beta$ ) errors to be 0,

fier une hypothèse dans un échantillon de la population cible. Règle générale, on envisage deux hypothèses. L'une d'entre elles est l'hypothèse nulle, selon laquelle il n'y a pas de différence. La vérification de la signification statistique est fondée sur l'hypothèse nulle<sup>3</sup>. Dans une étude qui compare les caractéristiques dans deux groupes, l'hypothèse nulle serait qu'il n'y a pas de différence entre les groupes pour ce qui est de la ou des caractéristiques d'intérêt. Dans une étude qui porte sur une intervention, l'hypothèse nulle serait que l'intervention ne modifie pas le devenir des patients (étude sur un seul groupe) ou qu'il n'y a pas de différence pour ce qui est du devenir des patients entre le groupe traité et le groupe témoin. La seconde hypothèse est l'hypothèse alternative, selon laquelle il y a une différence entre les groupes, par exemple quant à la ou aux caractéristiques ou quant au devenir des patients après une intervention. Dans l'étude fictive ci-dessus, l'hypothèse nulle est qu'il n'y a pas de différence quant au délai de cicatrisation des plaies entre le groupe traité au moyen du pansement XYZ et le groupe témoin, et l'hypothèse alternative est qu'il y a une différence quant au délai de cicatrisation des plaies entre les deux groupes.

Quand les investigateurs interprètent les résultats des tests statistiques sur leurs études, ils peuvent faire deux types d'erreur : une erreur de type I ou une erreur de type II (voir tableau 1). On parle d'erreur de type I quand les investigateurs concluent, sur la foi de la valeur  $p$ , que la comparaison entre les groupes a fait ressortir une relation ou un effet thérapeutique significatif quand il n'y en a en réalité pas dans la population cible (soit hypothèse nulle vraie). Les investigateurs limitent la probabilité d'erreur de type I quand ils choisissent le seuil alpha ( $\alpha$ ) qu'ils utiliseront pour déterminer si les résultats des tests statistiques sont significatifs. En général, le seuil  $\alpha$  est fixé à 0,05, ce qui signifie que les investigateurs sont prêts à accepter une probabilité de 5 % (1 chance sur 20) d'erreur de type I.

L'autre type d'erreur est l'erreur de type II, ou erreur bêta ( $\beta$ ). On parle d'erreur de type II quand les investigateurs, sur la foi d'une valeur  $p$  non significative, acceptent l'hypothèse nulle, selon laquelle il n'y a pas de relation ou d'effet thérapeutique quand il y en a en réalité un dans la population cible. La puissance ( $1 - \beta$ )

Le présent article a d'abord été publié dans le *Journal of Wound Ostomy and Continence Nursing*. Il est reproduit ici parce qu'il est bien écrit, compréhensible et pertinent pour les lecteurs de *Soins des plaies Canada*. C'est en outre un prolongement logique de la série d'articles intitulée *Compréhension de la littérature scientifique*, car il explique la puissance à la lumière des concepts qui ont été exposés dans les articles publiés jusqu'ici cette année. Le présent article explique pourquoi les cliniciens doivent comprendre la puissance statistique d'un projet de recherche et pourquoi il est important qu'un compte rendu présente la méthode de calcul de la puissance ou de détermination de la taille de l'échantillon, surtout quand aucun effet n'a été observé.

Reproduit avec l'autorisation de : Engberg S, Schlenk EA. Power: How to use it in evidence-based practice. *J Wound Ostomy Continence Nurs*. 2010;37(1):14-16. Publication : Lippincott Williams & Wilkins.

allowing them to be confident that their sample findings reflect reality in the target population. Unfortunately, achieving this goal generally requires a larger sample than can be feasibly recruited.<sup>3</sup> Although there are exceptions, in most situations type II errors (e.g. erroneously concluding that an intervention is not effective when in reality it is) are considered less critical than type I errors (e.g. erroneously concluding that the intervention is effective when it is not). For this reason, a type II error rate of 0.20 is often considered acceptable.

### Power and sample size

Power is an important consideration when studies are designed to examine the relationships between variables (e.g. characteristics associated with wound healing) or the impact of an intervention. Power is a function of the level of significance ( $\alpha$ ), effect size and sample size.<sup>4</sup> Power is inversely related to the level of statistical significance used to avoid making a type I error. While a small  $\alpha$  level such as 0.01 or 0.001 decreases the likelihood of making a type I error, it also decreases the power to detect significant effects when they do exist.

Effect size refers to the strength of the relationship or treatment effect that the investigators want to be able to detect. The effect size can be estimated from the findings of published studies or from previous pilot work that the investigators have done, or it can be based on the minimum clinical difference that is thought to be worth detecting. The effect size that investigators want to detect is directly related to power. The power to detect large treatment effects or differences between groups is greater than the power to detect small differences.

The other factor that determines power is the sample size of a study. Sample size is directly related to the power to detect significant differences that exist in the target population. Sample size is the factor that investigators can most readily manipulate to ensure that a study is adequately powered to detect differences in the target population. If the sample size is too

est la probabilité qu'une différence soit décelée dans la population cible au moyen du test statistique quand une telle différence existe. Par exemple, un seuil  $\beta$  de 0,10 signifie que les investigateurs sont prêts à accepter une probabilité de 10 % de ne pas déceler d'effet significatif, si un tel effet existe. Un seuil  $\beta$  de 0,10 veut dire que la puissance de détection d'un effet thérapeutique est de 90 %.

Idéalement, les investigateurs aimeraient que la probabilité d'erreur tant de type I ( $\alpha$ ) que de type II ( $\beta$ ) soit de 0, ce qui leur donnerait la certitude que les résultats obtenus dans l'échantillon reflètent la réalité dans la population cible. Malheureusement, pour y parvenir, il faut en général un échantillon dont la taille n'est pas réaliste<sup>3</sup>. Bien qu'il y ait des exceptions, dans la plupart des cas, les erreurs de type II (p. ex. conclure à tort qu'une intervention n'est pas efficace quand elle l'est en réalité) sont considérées moins critiques que les erreurs de type I (p. ex. conclure à tort qu'une intervention est efficace quand elle ne l'est pas en réalité). C'est pourquoi un taux d'erreur de type II de 0,20 est souvent jugé acceptable.

### Puissance et taille de l'échantillon

La puissance est une importante considération quand les études ont pour but d'examiner les relations entre des variables (p. ex. les caractéristiques associées à la cicatrisation des plaies) ou l'impact d'une intervention. La puissance dépend du seuil de signification ( $\alpha$ ), de l'ampleur de l'effet et de la taille de l'échantillon<sup>4</sup>. La puissance est inversement proportionnelle au seuil de signification statistique utilisé pour éviter de faire une erreur de type I. Un seuil  $\alpha$  faible, tel que 0,01 ou 0,001, réduit la probabilité d'erreur de type I, mais il réduit aussi la puissance de détection d'effets significatifs quand ils sont présents.

L'ampleur de l'effet désigne l'importance de la relation ou de l'effet thérapeutique que les investigateurs veulent pouvoir détecter. On peut estimer l'ampleur de l'effet à partir des résultats d'études publiées ou d'études pilotes menées antérieurement par les investigateurs, ou en se fondant sur la différence clinique minimale dont la détection est considérée comme valable. L'ampleur de l'effet que les investigateurs désirent détecter est directement liée à la puissance. La puissance de détection d'importants effets thérapeutiques ou de grandes différences entre des groupes est plus grande que la puissance de détection de petites différences.

L'autre facteur qui détermine la puissance est la taille de l'échantillon d'une étude. La taille de l'échantillon est directement liée à la puissance de détection de différences significatives qui existent dans la population cible. La taille de l'échantillon est le facteur que les investigateurs peuvent le plus facilement modifier

TABLE 1

### Correct and incorrect decisions about the target population

In the sample	In the target population	
	Alternate hypothesis is true	Null hypothesis is true
Alternate hypothesis is true	Accurate results	Type I error
Null hypothesis is true	Type II error	Accurate results

# Prevention through innovation

Preventing complications after surgery is one of your highest priorities.

It's one of ours too. PICO can help.

- PICO has all the effectiveness of negative pressure wound therapy (NPWT),<sup>1,2</sup> canister-free
- Significantly reduces cost of therapy<sup>3</sup>
- May help reduce surgical complications (eg, dehiscence) or complications from edema
- PICO is available for a wide range of acute and chronic wounds, and closed incisions



To experience PICO firsthand, visit [PossibleWithPICO.com](http://PossibleWithPICO.com) or call 1-800-876-1261 (US), 1-800-463-7439 (Canada)

## Breast reconstruction



Post surgical wound following breast reconstruction. The wound measured 15cm in length and 0.5cm in width. Note the closed incision is raised above the level of surrounding skin.



Patient with dressing in place on first assessment (day 1).



Post surgical wound after PICO removal day 11. Note some indentation from the PICO dressing but otherwise healthy surrounding skin. The surgical incision is closed, there is no exudate and the incision is flat. Compare with the tunneled IV site wound of same age which remains raised.

1. CSR/CT09/02 A prospective, open, non-comparative, multi-centre study to evaluate the functionality and dressing performance of a new negative pressure enhanced dressing in acute wounds.  
2. Data on File 1104011 Assessment of simplified NPWT device in pre-clinical blood flow studies.  
3. Data on File. Economic Analysis, Andriy Moshyk.

©2012 Smith & Nephew, Inc. All rights reserved. \*Trademark of Smith & Nephew. Certain trademarks registered in the US Patent & Trademark Office. RA

small, the differences in the target population might not be detected, whereas if the sample size is too big, the investigators may spend more time and money than needed to detect differences in the target population. Thus, a research plan should include a power calculation to determine how many subjects are needed to detect the desired effect size in the target population. When the findings of a research study are published, the results of the power calculation should be reported.

### Power in evidence-based practice

When you are considering research studies as potential evidence for clinical practice, it is important to determine whether a power calculation was reported. Power calculations are particularly important when the findings of interest were “not significant.” Without knowing if the study was adequately powered, you cannot be sure if nonsignificant findings reflect reality in the target population or if they are the result of a type II error. If a power calculation was performed to determine the sample size needed, it will be reported in the “Methods” section of the article.

If the findings of a study were not significant, it is also important to confirm that the study met the stated sample size goal. The following example of a power analysis is from a study comparing the impact of medical-grade honey to usual care on wound-healing rates in wounds healing by secondary intention.<sup>5</sup>

*It was estimated that 50% of patients treated with conventional dressings would achieve healing by 24 weeks. To detect a 20% greater healing rate in the honey treated group with 80% power and at the 5% statistical significance level, each group required 93 patients. A target recruitment of 200 patients was therefore needed to allow for potential dropouts.*

Based on the reported power analysis, the study needed to recruit 186 subjects (93 per group) in order to have an 80% probability of avoiding a type II error. The study was powered to detect a 20% difference in healing rates between the 2 groups (effect size). When data collected in this study were analyzed, there was not a significant difference in the healing rates in the 2 groups. Before you conclude that the honey is no more effective than conventional care in shortening the time to wound healing, you need to compare the final sample size to the number of subjects who, the power calculation indicated, were needed to detect the desired effect size. A study with nonsignificant findings can only be considered adequately powered if it meets

pour s’assurer que leur étude a la puissance voulue pour détecter des différences dans la population cible. Si l’échantillon est de trop petite taille, les différences dans la population cible pourraient ne pas être détectées, tandis que si l’échantillon est de trop grande taille, les investigateurs pourraient consacrer trop de temps et d’argent à la détection de différences dans la population cible. Par conséquent, le plan de recherche doit prévoir un calcul de la puissance qui permet de déterminer le nombre de sujets nécessaire à la détection de l’ampleur de l’effet désirée dans la population cible. Les résultats du calcul de la puissance doivent figurer dans le compte rendu d’une étude de recherche.

### Puissance et pratique fondée sur les données probantes

Quand on considère une étude de recherche pour déterminer si elle pourrait être une source de données probantes pour la pratique clinique, il est important de déterminer si le compte rendu présente le calcul de la puissance. Le calcul de la puissance est particulièrement important quand les résultats d’intérêt ne sont pas significatifs. En effet, quand on ignore si l’étude avait la puissance voulue, on ne peut déterminer avec certitude si des résultats non significatifs sont le reflet de la réalité dans la population cible ou s’ils sont le fruit d’une erreur de type II.

Si les résultats d’une étude ne sont pas significatifs, il est aussi important de confirmer si l’étude a atteint l’objectif fixé en ce qui a trait à la taille de l’échantillon. L’exemple ci-dessous d’analyse de la puissance est tiré d’une étude comparant le miel de qualité médicale aux soins habituels du point de vue de l’effet sur les taux de cicatrisation des plaies dans les cas de cicatrisation des plaies par deuxième intention<sup>5</sup>.

*On a estimé que chez 50 % des patients traités au moyen de pansements classiques, le délai de cicatrisation des plaies serait de 24 semaines. Pour détecter un taux de cicatrisation 20 % plus élevé dans le groupe traité par le miel avec une puissance de 80 % et un seuil de signification statistique de 5 %, il fallait 93 patients dans chaque groupe. On a donc fixé un objectif de recrutement de 200 patients, au cas où des patients se retireraient de l’étude. (p. 568)*

Selon l’analyse de la puissance présentée, 186 patients (93 par groupe) devaient participer à l’étude pour que la probabilité qu’il n’y ait pas d’erreur de type II soit de 80 %. L’étude avait la puissance voulue pour détecter une différence de 20 % entre les deux groupes quant aux taux de cicatrisation (ampleur de l’effet). L’analyse

its sample size goal. In this study, they were able to recruit only 105 subjects instead of the projected 186, which must be considered when the findings are interpreted. One possible explanation for the nonsignificant findings is that the study was not adequately powered to detect a significant treatment effect.

### Conclusions

Power, the ability of statistical testing to detect significant differences that exist in the target population, is an important concept for investigators designing studies and for clinicians utilizing research findings in evidence-based practice. Investigators need to do a power calculation to make sure that their studies are adequately powered to detect clinically meaningful differences in the target population. When clinicians consider research studies as potential evidence for clinical practice and study findings are not significant, it is important to determine whether a power calculation was reported and to confirm that the study met the stated sample size goal. Unless a study was adequately powered, clinicians cannot be sure if nonsignificant findings reflect reality in the target population or if they are the result of a type II error. ☹

### References

1. Breaux RH, Carnat TA, Gaboury I. Inadequate statistical power of negative clinical trials in urological literature. *J Urol.* 2006;176:263-266.
2. Livingston EH, Cassidy L. Statistical power and estimation of the number of required subjects for a study based on the t-test: a surgeon's primer. *J Surg Res.* 2005;126:149-159.
3. Browner WS, Newman TB, Hulley SB. Getting ready to estimate sample sizes: hypotheses and underlying principles. In: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB (eds). *Designing Clinical Research*, 2nd ed. New York: Lippincott Williams & Wilkins; 2007:51-63.
4. Fox N, Mathers N. Empowering research: statistical power in general practice research. *Fam Pract.* 1997;14:324-329.
5. Robson V, Dodd S, Thomas T. Standardized antibacterial honey (Medihoney™) with standard therapy in wound care: randomized clinical trial. *J Adv Nurs.* 2009;65:565-575.

TABEAU 1

### Bonne et mauvaise conclusions sur la population cible

Dans l'échantillon	Dans la population cible	
	L'hypothèse alternative est vraie	L'hypothèse nulle est vraie
L'hypothèse alternative est vraie	Résultats exacts	Erreur de type I
L'hypothèse nulle est vraie	Erreur de type II	Résultats exacts

des données recueillies au cours de l'étude a révélé qu'il n'y avait pas de différence significative entre les deux groupes quant au taux de cicatrisation. Toutefois, avant de conclure que le miel n'est pas plus efficace que les soins classiques pour réduire le délai de cicatrisation des plaies, il faut comparer la taille de l'échantillon final au nombre de sujets qui, selon le calcul de la puissance, était nécessaire pour détecter l'ampleur de l'effet désirée. En effet, on ne peut considérer qu'une étude dont les résultats ne sont pas significatifs à la puissance voulue que si l'objectif en ce qui a trait à la taille de l'échantillon a été atteint. Les investigateurs de l'étude qui nous intéresse n'ont pu recruter que 105 sujets au lieu des 186 prévus, ce dont il faut tenir compte pour interpréter les résultats, car les résultats pourraient avoir été non significatifs entre autres parce que l'étude n'avait pas la puissance voulue pour détecter un effet thérapeutique significatif.

### Conclusions

La puissance, soit la capacité d'un test statistique de détecter les différences significatives qui existent dans la population cible, est un important concept pour les investigateurs qui conçoivent des études et pour les cliniciens qui utilisent les résultats des études dans la pratique fondée sur les données probantes. Les investigateurs doivent calculer la puissance pour s'assurer que leurs études ont la puissance voulue pour détecter des différences cliniquement significatives dans la population cible. Quand les cliniciens considèrent une étude de recherche dont les résultats ne sont pas significatifs pour déterminer si elle pourrait être une source de données probantes pour la pratique clinique, ils doivent déterminer si le compte rendu présente le calcul de la puissance et si l'objectif fixé en ce qui a trait à la taille de l'échantillon a été atteint. À moins qu'une étude ait la puissance voulue, les cliniciens ne peuvent déterminer avec certitude si des résultats non significatifs sont le reflet de la réalité dans la population cible ou s'ils sont le fruit d'une erreur de type II. ☹

Références (voir page 11)



**Publish in  
Wound Care  
Canada!**

Contact  
wcceditor@cawc.net  
for further information.