Survivorship bias and its implications in dental research and literature

LM Sykes[1], GP Babiolakis[2] and LV Mkhonza[3]

ABSTRACT

What does the Texan sharpshooter, the World War 2 American air force, The Taj Mahal, cherry farming, and dentistry have in common? They have all been the subject of Survivorship bias. This paper will discuss the concept of survivorship bias in research by using examples of flaws in study designs or interpretation of data. It aims to serve as a caution to practitioners of the need to be aware and critical when reading literature and manufacturers' reports before making decisions regarding materials choices and treatment protocols to follow.

BACKGROUND

New developments in the field of dental materials and techniques have always, and continue to advance at an incredibly fast rate. In keeping with this, both manufacturers and clinicians need to play a part in monitoring the latest products to assess their laboratory and clinical performance, durability, and longevity. These characteristics then need to be compared to those currently in use before one can be replaced with the other. The literature is replete with studies investigating success rates and survival rates of different materials, components, and procedures. However, a recurring problem with comparative studies is the lack of standardization in experimental design, measurements, and scoring criteria. Numerous authors have published papers on standardized guides to utilise when assessing clinical success and survival, most notably in the field of dental implantology. These were developed to help formulate a way of making quantitative and qualitative comparisons that could be substantiated with statistical evidence. This paper will present a brief description and definition of common terms and calculations used in success and survival studies. It will then illustrate examples of inherent flaws in study designs or interpretation of data. This aims to caution practitioners of the need to be particularly critical when reading literature and manufacturers' reports.

Author affiliations:
1. Leanne M Sykes: *BSc, BDS, MDent, IRENSA, Dip Forensic Path, Dip ESMEA,* Head of Department of Prosthodontics, University of Pretoria. Orchid: 0000-0002-2002-6238
2. Dr George P Babiolakis: *BDS, MscDent, PGDipDent,* Registrar, Department of Prosthodontics, University of Pretoria, South Africa
3. Londiwe V Mkhonza: *BChD, Dip-Implantology,* Developmental post, Dentist, Department of Prosthodontics, University of Pretoria

Corresponding author: Leanne Sykes
Head of Department of Prosthodontics, University of Pretoria.
Email: Leanne.sykes@up.ac.za

Author contributions:
1. Leanne M Sykes: Primary author: 60%;
2. George Babiolakis: 25%;
3. Londiwe Mkhonza: 15%

Literature review
1. Success, Survival, and Failure

In restorative material studies different criteria, definitions and types of statistics may be used for reporting on performance. Some papers refer to percent success and percent survival.[1] In these Success is taken as the "number of restorations that survive without any adverse effects", while Survival relates to "the number of restorations that survive even though they may have exhibited some form of damage which has not adversely affected their function or aesthetics".[1,2] Others report on success and survival in terms of probabilities over set periods. They generally consider restorations and prostheses as a success if they have performed as expected for 5 or more years. Success and survival have at times been combined into one definition as "the intact survival of a prosthesis with acceptable surface quality, anatomical contour and function, and where applicable, with acceptable aesthetics".[2]

Restorations would be considered **Failures** if they "had to be replaced during this time for reasons such as secondary caries, irreversible pulpitis, excessive wear of opposing tooth surfaces, excessive erosion and roughening of their surface, ditching of the cement margin, unacceptable aesthetics, cracking, chipping, fracture or bulk fracture".[2] However, there is no standardization in the criteria used to measure these indicators or the parameters reported across different studies or with differing materials. For example are repairable fractures considered failures or technical complications? And should data collection and reporting encompass clinical prosthesis failures, technical complications, and biological side effects?[1] In addition, there needs to be a clear distinction between in vitro studies and data and those from clinical performance studies.

In implant dentistry, a whole gamut of different benchmarks has been postulated for determining implant as well as restoration success and survival. Not only do these indices differ between researchers, specialties, and implant companies, but the measures are constantly being revised to account for extraneous factors such as patient profiles, surgical techniques used, occlusal considerations, implant design, and newer consensus opinions. This makes the comparisons of results on implant success rates far more complex and adds to the difficulty many clinicians have when deciding on which is the best to use for each specific patient.[3]

To further illustrate these difficulties consider how the parameters used for assessing implants changed within the first ten years of reporting. The first set of implant success criteria were developed by a group of experts in 1979.[4] The standards included that "the implant should display mobility less than 1mm in any direction; there should be no bone loss

larger than one third of the vertical height of the bone; there should be no symptoms, infection, nor damage to adjacent teeth; there should be no paraesthesia and no damage to the mandibular canal, maxillary sinus, as well as the nasal floor; and that the implant should function in the patient's mouth for five years in seventy-five percent of all patients".[4] In 1982, the criteria was expanded upon to include that implants "must have been present in the oral cavity for sixty months or more; that there should be a definitive lack of mobility; there should be no evidence of radiolucency on the radiograph in the cervical region of implants; the implant should be free of haemorrhage; the patient should display no pain or percussion sensitivity from the implant; and intraorally, there should be no pericervical granulomatosis, gingival hyperplasia, or widening of the peri-implant space on the radiograph".[5] In 1984, additional success criteria were then added as "subjective and objective indices".[6] "The subjective measures included proper function, absence of discomfort and/or pain, and the patients' belief that the implant had resulted in improvement of their aesthetics, as well as their emotional and psychological attitude". The objective criteria "included good occlusal balance and vertical dimension, maintenance of healthy collagenous tissue, and no polymorphonuclear infiltration".[7]

## 2. Accuracy, Precision, Validity, and Reliability

- **Accuracy** is an indicator of how close measurements are to the true value
- **Precision** refers to the degree to which repeated measurements will produce the same result. It also relates to how close repeated measurements are to each other
- **Validity** shows how appropriate and adequate a test is for that specific purpose, as well as how correctly the results have been interpreted and subsequently used.
- **Reliability** is a measure of consistency and shows the degree to which a method produces the same results when used at different times, under different circumstances, by the same or other observers.

Both accuracy and precision will affect validity and reliability (Figure 1).[8]

3. Survivorship bias

The Texan Sharpshooter Fallacy is a fictional story that was used to illustrate how misconceptions and false conclusions can arise when differences in data are ignored, while similarities are overemphasized. It relates to a human cognitive phenomenon called the "clustering illusion" where the observer interprets patterns when none exist. Legend has it a Texan fired a series of gunshots at the side of a barn. He then looked for the area where the most holes were clus-

Figure 2. Illustration of The Texan Sharpshooter Fallacy
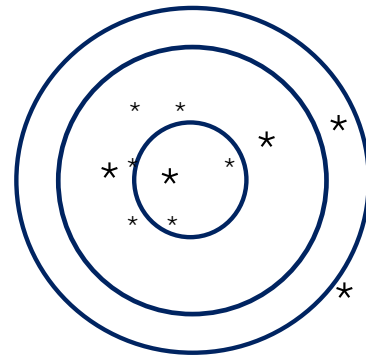
tered and drew a target around these, thus creating the false impression that he was an accurate sharpshooter.

Another famous example of survivorship bias dates back to World War 2. At the time, the American military asked mathematician Abraham Wald to study how best to protect airplanes from being shot down. The military knew that extra armour would help, but it couldn't be used to protect the whole plane as this would make the aircraft far too heavy to fly. Initially, they planned to examine the planes returning from combat, and see where they were worst hit, and then reinforce these areas. They identified the wings, around the tail gunner, and down the center of the body as the worst-hit zones and proposed to reinforce these areas (Figure 3).

But Wald realized they had fallen prey to survivorship bias because their analysis was missing a valuable part of the picture: the planes that were hit but that hadn't made it back. He reasoned that the returning planes surviving the bullet holes they were looking at actually indicated areas that could be hit and still allow the planes to keep flying, whereas those that didn't return must have been hit in "critical" areas. As a result, the military was planning to armour the wrong parts of the planes, the exact areas that didn't need reinforcing.[9]

A type of survivorship bias that can be seen in civil and urban architecture and construction is called "selection bias". This involves the selection of individuals or data in such a way that proper randomisation is not achieved, and the sample is not a true representation of the entire population. For example, commercial development entails that buildings are constantly being torn down and replaced with new structures that may have more modern designs, make better use of the space available, use stronger, more durable, or cost-effective materials, and have improved ergonomics and functionality. Only the most popular, or presumed beautiful, historic buildings

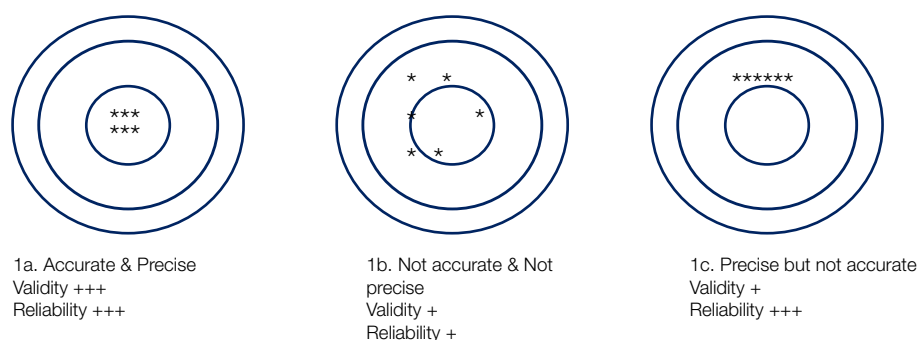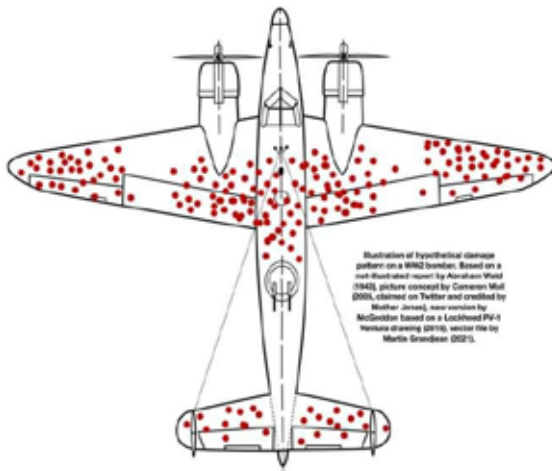Figure 1. Accuracy versus Precision (Reproduced from Sykes et al.[8])

1a. Accurate & Precise
Validity +++
Reliability +++

1b. Not accurate & Not precise
Validity +
Reliability +

1c. Precise but not accurate
Validity +
Reliability +++

Figure 3. American World War 2 aircraft (Reproduced from Mangel and Samaniego al.[9])

tend to be preserved and as such will survive from one generation to the next. Generally, the ugliest and weakest erections tend to have been demolished creating the erroneous impression that all old buildings were more beautiful and better built than modern structures.

Agriculture too has been the victim of "the fallacy of incomplete evidence" or "suppression of evidence". It occurs where only data that conforms to a certain standard or position is considered, while that which does not adhere to these criteria is ignored. For example, one presumes that cherry pickers will only select the ripest and healthiest fruits to send to market. An observer will see their load and erroneously conclude that most of the tree's fruit is likewise as good. This gives them false impressions of the overall quality of all the fruit as they see a mere "snapshot" of the best. (NB do not confuse this with cherry-picking which refers to selecting the most easily accessible fruit/data and ignoring that higher up / more difficult to reach)

Discussion of potential survivorship bias in dentistry
Bias is a systematic error in the design (selection bias) or methodology (information bias) of the study, leading to inacuracies in data or outcomes.[10] Many clinical decisions are based on studies that use success and survival parameters similar to those defined in this paper for dental materials and implantology. However, following the illustrations of survivorship bias, it may be interesting to re-look at both our current clinical practices as well as the literature on which these are based, and scrutinise the data for evidence of these types of errors. Treatment decisions based on clinical experience may be flawed in that happy patients tend to return for follow-up visits, while those who have had complications or treatment failures will often go to a different practitioner. Thus you may never be aware of what went wrong or seek answers as to why, and adjust your practices accordingly in the future. Erroneous conclusions can also be drawn from studies that do not compare "like-with-like" components, procedures, patient profiles, periods, surgical experience, occlusion, or any number of other dissimilar influences. Studies that do not take into account the outliers, or the "lost to follow-up" cases may miss out on valuable evidence, which could contradict

the data collected and analysed. Similarly, those that only look at successes and survival rates, or complications associated with the survivors, may never be aware of the number and types of failures. The use of Kaplan–Meier mathematics in statistical analyses is another factor that may have added to survivorship bias. The Kaplan-Meier statistics measures the estimated cumulative survival of a prostheses over time. It looks at a specific time interval and measures the number of failed and at risk prosthesis during this period. The problem with this tool is that the failure number is expressed as a numerator in the equation, but if a failure is not reported (for example if a patient does not return), it will not be registered as a loss and the estimated cumulative survival stays high.[11] Furthermore, in vitro and animal studies have limited use if they cannot mimic the exact human oral conditions where the entity under investigation will need to perform.

A slightly different, but equally relevant issue to consider when measuring success, is the objective versus the subjective judgment. What the clinician (and literature) deem to be successful in terms of objective outcome measures, may not be the same as the patients' opinions. If a seemingly perfect restoration does not meet the patient's expectations in terms of function, aesthetics, or comfort, should it be considered a success or a failure? And who decides?

CONCLUSIONS
In dentistry, we may make crucial errors in judgment if we base our clinical practice on research or observations of the survivors, the procedures that outperform others, or those things we deem to be successful. Conclusions based only on the successful attributes, without looking more broadly at the whole dataset, including the failures, will constitute survivorship bias. This may blind the dentist to their faults leading to oblivious repetition of the error. While it may be a hard and humbling exercise, clinicians and researchers need to be conscious and courageous enough to scrutinise their failures. They must ask themselves why something didn't work, discuss possible reasons with colleagues or manufacturers, learn from their mistakes, and most importantly share this knowledge with others. This is how real progress is made, and how we avoid survivorship bias.

References
1. Anusavice, K., Rawls, R., Esquivel-Upshaw, J. and Shen, C., 2012. Phillips' science of dental materials. 12th ed. Elseiver, pp.1-592.
2. Anusavice KJ. Standardizing failure, success, and survival decisions in clinical studies of ceramic and metal-ceramic fixed dental prostheses. Dent Mater.2012 Jan; 28(1):102-111
3. Chércoles-Ruiz, A., Sánchez-Torres, A. and Gay-Escoda, C., 2017. Endodontics, Endodontic Retreatment, and Apical Surgery Versus Tooth Extraction and Implant Placement: A Systematic Review. Journal of Endodontics, 43(5), pp.679-686.
4. Schnitman PA, Shulman LB. Recommendations of the consensus 8. development conference on dental implants. J Am Dent Assoc 9. 1979;98:373 7.
5. Cranin AN, Silverbrand H, Sher J, Salter N. The requirements and 10. clinical performance of dental implants. In: Smith DC, Williams DF, 11. editors. Biocompatibility of Dental Materials. Vol. 4. Boca

Raton, Fla: 12. CRC Press; 1982. pp. 198. 5.
6. McKinney R, Koth DL, St and k DE. Clinical standards for dental 13. implants. In: Clark JW, editor. Clinical Dentistry. Harperstown: Harper 14. and Row; 1984. pp. 1 11.
7. Batth RK. What is the difference between implant success and survival and how will it change the future use of implants as a permanent solution to tooth loss? 2014. A Boston University Theses & Dissertations. Accessed at: https://hdl.handle.net/2144/15043; Accessed on: 18-10-2021
8. LM Sykes, F Gani, Z Vally. Statistical terms Part 1: The meaning of the MEAN, and other statistical terms commonly used in medical research. SADJ, July 2016, Vol 71 (6) 274-278.
9. Mangel, M. and Samaniego, F., 1984. Abraham Wald's Work on Aircraft Survivability: Rejoinder. Journal of the American Statistical Association, 79(386):270.
10. Pandi, N., 2014. Bias in observational studies. Am J Orthod Dentofacial Orthop 145:542-543.
11. Layton, D., 2013. Understanding Kaplan-Meier and Survival Statistics. The International Journal of Prosthodontics, 26(3), pp.218-226.

Online CPD in 6 Easy Steps

The Continuous Professional Development (CPD) section provides for twenty general questions and five ethics questions. The section provides members with a valuable source of CPD points whilst also achieving the objective of CPD, to assure continuing education. The importance of continuing professional development should not be underestimated, it is a career-long obligation for practicing professionals.

Online CPD in 6 Easy Steps

1. Go to the SADA website www.sada.co.za.

2. Log into the 'member only' section with your unique SADA username and password.

3. Select the CPD navigation tab.

4. Select the questionnaire that you wish to complete.

5. Enter your multiple choice answers. Please note that you have two attempts to obtain at least 70%.

6. View and print your CPD certificate.