The role of the randomised controlled trial in restorative dentistry and the correct purpose of observational data

S. Mickenautsch^{*1} and V. W. Berger²

Key points

Outlines the consequences of data comparison between non-randomised patient groups and its influence on clinical inference.

Shows that randomisation is essential for preventing invalid, potentially misleading results in restorative dentistry. Highlights the problems of inferring causality from correlation of observational data and shows its correct purpose.

It has been argued that the randomised controlled trial design is unsuitable for restorative dentistry and that cohort studies or the analysis of large observational datasets without randomisation is more suitable. This opinion article examines why randomisation in clinical trials is needed and why big observational data is not enough for clinical inference.

Introduction

In the last few months we saw in the scientific dental literature that tooth restoration survival data was compared between two patient groups, formed without the benefit of randomisation, and strong clinical conclusions were subsequently published in the *British Dental Journal*,¹ as well as in the *Journal of Dentistry*.² Also, correlations from big observational data were promoted³ and effectively applied for causal inferences^{1,4} and, lastly, for good measure, the utility of the randomised controlled trial (RCT) for restorative dentistry was publicly questioned altogether.³

In their 'Ultimate guide to restoration longevity in England and Wales', Lucarotti and Burke⁴ report in the *British Dental Journal* on observational survival data of over 13.8 million tooth restorations, and inform the reader that correlations between restoration survival data and other factors, such as the restoration type, can serve as the basis for causal relationships (ie restoration type causes length of restoration

Refereed Paper. Accepted 29 October 2018 Published online 18 January 2019 DOI: 10.1038/sj.bdj.2019.43

survival), simply because their analysed dataset is 'large'. They further state that comparisons of survival data between different restoration types are possible, because 'all restorations were treated similarly.' In their recent systematic review in the Journal of Dentistry, Ruengrungsom et al.² compute the median annual failure rates for various restoration types from a number of different clinical studies and present them, without further analysis, in a table. From that, the authors drew conclusions as to the superiority/inferiority of one restoration type above the other. In a similar sense, Burke and Lucarotti see no problem to conclude that one type of restoration was 'found to perform suboptimally, when compared with other restoration types.'1

With all this being the case, why is it wrong?

Why randomisation in clinical trials is needed

It may sound plausible, and indeed it may be tempting, to view RCTs in restorative dentistry as something that is too expensive, too slow and rather unhelpful when its results are inconclusive,³ that is, when they yield no statistically significant results, particularly in the absence of a sufficiently large sample size. In contrast, observational data collected in the 'real world' of dental practice networks seems more appealing, particularly when the results of different restorative treatment types can so easily be compared in simple graphs and tables.^{1,2} However, the expedience of observational studies carries over to other fields as well, and is most certainly not limited to restorative dentistry, and in these other fields, randomised trials are used anyway, despite the added expense. Why is that? The answer is that it is recognised that the results of comparisons of groups formed without randomisation will, in general, have nothing in common with what would reflect the 'real world' in terms of therapeutic truth. Patients who choose one treatment, or have it chosen for them on the basis of symptoms or clinical presentation, will differ systematically from patients who choose the other treatment (or have it similarly chosen for them). They will differ not only in the type of restorative treatment rendered but also in many other, mainly unknown (mostly even unknowable) factors that we call confounders.5 Any observed differences (or the lack of them) in restoration survival between different restoration types can as well be ascribed to such factors instead of to the clinical efficacy of the compared treatment options, thus generating invalid, potentially misleading results.6

When the results from single-arm trials investigating high-viscosity glass-ionomer (HVGICs) and amalgam restorations, identified through a systematic literature search,⁷ were matched for cavity type and follow-up period and statistically compared, a completely different effect estimate in terms of restorations failure rate was obtained than was established when HVGIC were compared

¹SYSTEM Initiative/Department of Community Dentistry, School of Oral Health Sciences, Faculty of Health Sciences, University of the Witwatersrand, 7 York Rd., Parktown/ Johannesburg 2193, South Africa; ²Biometry Research Group, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, * Correspondence to: Prof S. Mickenautsch Email: neem@global.co.za

OPINION

against amalgam restorations in RCTs: odds ratio (OR) 6.29, 95% CI: 1.34–19.27 of the former versus OR 1.00, 95% CI: 0.81–1.20 from RCTs.⁷ When point estimates from the two different study designs were statistically compared using the Mann-Whitney U test the difference was found to be highly significant (U = 25, $n_{non-RCT \text{ estimates}} = 26$, n_{RCT} $_{\text{estimates}} = 8$; p = 0.0013). In other words, the comparison of non-randomised data indicate a statistically higher failure rate of HVGIC restorations than amalgam, while RCT results indicated no such differences, beyond the play of chance.

Which of the two results reflects the therapeutic truth? Which should we use to inform our patients? More help in answering these questions can be gleaned from an investigative trial simulation study in which the restoration failure rate between restoration type A and B was set to an absolute equal (Risk ratio (RR) = 1.00; p-value = 1.00). Two data scenarios, including confounders, were computed, one without (Scenario 1) and one with randomised patient allocation into treatment groups (Scenario 2).⁸

Scenario 1 was constructed by assuming the comparison of two interventions (A and B) with dichotomous outcomes (intervention failure: Y = 0; intervention success: Y = 1). A total of 22 simulated trials were formulated for each intervention. Half of the patients in each trial were assigned to an intervention A and half to an intervention B as simulated independent one-arm longitudinal studies. The treatment effect of both, intervention A and B, was set to be successful for all patients (Y = 1). Hence, when compared with each other, neither intervention would yield any result superior to that of the other (Risk ratio 1.00; p = 1.00). In addition, a percentage of all patients per intervention group were assumed to suffer from a confounding trait 'X' that would always result in intervention failure (Y = 0), regardless of the intervention type. The percentage of patients with 'X' in each intervention group per trial was called the trait frequency (TF). The TF (between 0 and 100%) was determined for each intervention group per trial using a random number generator. The random assignment of trait X represented a simulation of an unknown factor that may exist within a study sample and that may have a decisive influence on the trial results (for example, higher caries risk or specific oral hygiene behaviours that may impact on higher restoration failure rate etc).

In Scenario 2, a random sequence for allocation to group A or B, using fixed block randomisation (block-size 4) and in line with each sample size, was generated for each of the 22 trials generated in Scenario 1. All patients with their Y = 1/0 outcomes generated in Scenario 1 were combined for each trial and re-allocated to intervention group A or B along this random sequence. In this simulation, a complete lack of subversion of the randomisation and allocation process was assumed. It was further assumed that patients could not select their allocated intervention.

A random-effects meta-analysis was conducted for the 22 simulated trials per scenario and a pooled risk ratio (RR) with 95% confidence interval (CI) computed for each.8 The result of Scenario 2 was closer to the truth (RR 1.00; 95% CI: 0.96–1.04; p = 0.99). Computation of the non-randomised data generated, due to confounder influence, a misleading statistically significant result (RR 1.64; 95% CI: 1.22–2.19; p = 0.001), with a 64% lower restoration failure rate of restoration type B than A. As the effect difference between the compared interventions had been set at zero, the full confounding effect due to the randomly varying TF among patients per intervention group could be shown. The established 64% overestimation can mean that, for example, a true effect estimate of OR 0.30; CI 95%: 0.14-0.66 in favour of treatment is effectively distorted to an erroneous effect estimate of OR 2.21; CI 95%: 1.22-4.00 in favour of the control intervention the treatment is compared against. This means, while in truth, 17 out of 100 patients would have a better outcome with the treatment than with the control intervention, a 64% overestimation suggests not only that the control intervention is as successful as the treatment, but that an additional 18 out of 100 patients would be better than with the treatment.¹⁵ Nor is this discrepancy even surprising.

Shall the dental profession really risk informing patients about 'the potential for longevity of restorations' or 'for medico-legal reasoning' based on data comparisons from non-randomised patient groups only because RCTs are perceived as being too expensive, too inconclusive and too slow?^{3,4}

Non-randomised comparisons of patient data in clinical trials are based on the fallacious assumption that the compared patient cohorts are absolutely homogenous and are found to generate both artificially small standard errors and incorrect confidence intervals, and they also have poor agreement with actual RCT results.^{10,11} This implies that results from cohort studies or large observational datasets without randomisation, as suggested by Burke,³ (particularly to important questions of what placed restorative material survives better in the mouths of dental patients and what worse), may provide us with rather low verisimilitude.

Why big observational data is not enough

Even observations from big data, such as from 13.8 million tooth restorations,¹ are insufficient for valid causal inferences in regard to factors influencing their survival rate. Why? Because only associations between variables can be obtained that way. For sure, many million data points go a long way towards identifying highly statistically significant correlations, but even with the help of Cox regression it still will not be possible to discern whether any predictor variable actually caused the observed restoration survival or not. Correlations are just that: a relationship of variable A with that of B, and without any hint whether indeed A caused B, B caused A, or an unknown variable C caused both A and B. Such facts, commonly taught in undergraduate statistics courses, appear trivial, but seem rather counter intuitive in general human thinking.¹²

In fields in which RCTs are not possible (and restorative dentistry is not one of them) and in which reliance on sole observational data is the only choice, causal inference does not just rely on the mere statistically significant correlation between factors. Since epidemiological research has established the harm of active smoking based on exclusively observational data, there is consensus that, in order to allow at least some form of causal inference, all of Hill's criteria need to be fulfilled:¹³

- Strong association
- Counterfactual causality
- Repeated observation of the association
- A factor being specific for a particular outcome
- A factor has to precede the outcome it is assumed to affect
- A dose-response relationship exists
- There is a biological plausible explanation for the observed association
- The causal conclusion does not contradict present substantive knowledge in the field, particularly those derived from existing RCTs
- Analogous exposures and outcomes have already been established.

OPINION

Against this background, Burke and Lucarotti⁴ would have needed not only to show a strong correlation between one type of restorative material and lower survival data, but also that all the above criteria were fulfilled, including that their observation that one type of restoration was 'found to perform sub-optimally, when compared with other restoration types' did not contradict results from existing RCTs. In fact, existing RCTs do not concur.^{14,15}

Nevertheless, observational data from large datasets, such as that used by Lucarotti and Burke,¹ can provide strong and useful evidence for hypothesis generation that is invaluable for the justification of the planning, funding, and conduct of experimental investigations (in many case by use of well-designed RCTs), in order to establish what actually causes what. While a causal relationship between two factors cannot exist without them being correlated, correlations between factors can exist without one causing the other. Or in other words: causality is based on correlation but not all correlations include causalities. Hence, the conduct of causal investigations by use of RCTs between factors that don't correlate will be a waste of resources. By establishing where strong correlations between factors exist and where not, such observational research is the basis for selecting worthwhile hypotheses to pursue and this is the correct purpose of observational data in fields that are amenable to RCTs.

Randomisation in clinical controlled trials, particularly concerning tooth restoration survival, is an essential requirement in order to establish therapeutic truth in restorative dentistry. To abandon randomisation, because it may render clinical trials expensive and slow, is to abandon our hope that clinical advice to patients, information to healthcare funders, oral healthcare providers and managers correspond with reality and does not reflect unknown or even unknowable confounder effects. Equally, it is not possible to reliably obtain correct causal inferences from mere associations, and this remains true no matter how large the dataset that is analysed. However, observational data from large datasets are indeed helpful in hypothesis development and the justification for conducting, funding and managing RCTs.

- Burke F J T, Lucarotti P S K. The ultimate guide to restoration longevity in England and Wales. Part 3: Glass ionomer restorations – time to next intervention and to extraction of the restored tooth. *Br Dent J* 2018; 224: 865–874.
- Ruengrungsom C, Palamara J E A, Burrow M F. Comparison of ART and conventional techniques on clinical performance of glass-ionomer cement restorations in load bearing areas of permanent and primary dentitions: A systematic review. J Dent 2018; DOI: 10.1016/j. jdent.2018.07.008.

- Burke F J T. End of the road for the randomized controlled trial in restorative detistry? *Dent Update* 2017; 44: 806–808.
- Lucarotti P S K, Burke F J T. The ultimate guide to restoration longevity in England and Wales. Part 1: methodology. *Br Dent J* 2018; 224: 709–716.
- Berger V W. Trials: The worst possible design (Except for All the Rest). Int J Person Centered Med 2011; 1: 630–631.
- Berger V W. What do non-randomized trials offer above and beyond randomized trials? *Contemp Clin Trials* 2013; 35: 168–169.
- Mickenautsch S, Yengopal V. Direct contra naïve-indirect comparison of clinical failure rates between highviscosity GIC and conventional amalgam restorations: an empirical study. *PLoS One* 2013; 8: e78397.
- Mickenautsch S. SYSTEM Research Note on: How should competing clinical interventions be compared in dentistry? – A simulation-based investigation. J Minimum Interv Dent 2013; 6: 73–80.
- 9. Cates C. *Visual Rx Calculator*. Available at http://www. nntonline.net/visualrx/ (accessed August 2018).
- Glenny A M, Altman D G, Song F et al. International Stroke Trial Collaborative Group. Indirect comparisons of competing interventions. *Health Technol Assess* 2005; 9: 1–134.
- Odgaard-Jensen J, Vist G E, Timmer A et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011; 4: DOI: 10.1002/14651858.MR000012.pub3.
- Bleske-Rechek A, Morrison K M, Heidtke L D. Causal inference from descriptions of experimental and non-experimental research: public understanding of correlation-versus-causation. J Gen Psychol 2015; 142: 48–70.
- Hill A B. The environment and disease: Association or causation? *Proc R Soc Med* 1965; 58: 295–300.
- Mickenautsch S, Yengopal V. Failure rate of high-viscosity GIC based ART compared with that of conventional amalgam restorations-evidence from an update of a systematic review. SADJ 2012; 67: 329–331.
- Mickenautsch S, Yengopal V, Banerjee A. Atraumatic restorative treatment versus amalgam restoration longevity: a systematic review. *Clin Oral Investig* 2010; 14: 233–240.