Need of Generating Level 1 Evidence in the Therapy of Endocrine Glands

Anurag Srivastava

I just crossed my 65th birthday and retired from the All India Institute of Medical Sciences, New Delhi, India, serving as the Chairman of the Department of Surgical Disciplines.

In this long surgical career, I had the good fortune to interact and work in close association with surgical practitioners of several disciplines, viz. general surgery, plastic surgery, cardiac surgery neurosurgery, orthopedic surgery, ocular surgery, and pediatric surgery, just to name a few. A unique observation was made in this long career that two breeds of surgeons hardly ever conduct randomized trials: the plastic surgeons and the endocrine surgeons.

Randomized clinical trials are considered the gold standard for the practice of evidence-based medicine.

Why do We Need Randomized Trials?

The decision about the most effective therapy out of a choice of two or three options should be made in the most scientifically valid and reliable manner. The success of the therapy should be recorded in a truthful manner not vitiated by the effect of other factors. This effect of other factors vitiating /corrupting the results of a study is called "bias" The main problem with formulating treatment strategies based on observational studies is the fact that many biases creep in and distort the truth. Three types of biases may corrupt the results of a medical study: the selection bias, the measurement bias (also called information or ascertainment bias), and confounding bias.

Randomization allows study groups to be comparable with regard to both known and unknown risk factors and confounders and removes the investigator bias in the allocation of therapy (called selection bias). Moreover, it guarantees that the statistical tests will achieve a valid significance level.

Randomized controlled trials offer protection from selection bias and measurement bias (with blinded assessment of outcomes) and surely avoid confounding bias.

Let us appreciate as to how to remove biases in the assessment of outcomes of a particular operation on cancer?

We need to thoughtfully conceive and design a sufficiently empowered (with appropriate sample size), multicenter (preferably to recruit more cases and to make the study generalizable), randomized controlled trial recording important patient-reported outcomes (PRO) and quality of life parameters (QOL) along with oncologic outcomes, viz. 2, 5, and 10 years of disease-free and overall survival. The outcome parameters should be ascertained by a blinded investigator (who is unaware of the type of surgery performed) using valid and reliable tools and questionnaires. The statistician too should be blinded/masked about the intervention groups and perform cogent statistical analyses using "individual patient data meta-analysis."

Balco Medical Centre, Naya Raipur, Chhattisgarh, India

Corresponding Author: Anurag Srivastava, Balco Medical Center, Naya Raipur, Chhattisgarh, India, email: dr.anuragsrivastava@gmail.com

How to cite this article: Srivastava A. Need of Generating Level 1 Evidence in the Therapy of Endocrine Glands. Indian J Endoc Surg Res 2021;16(1):1–4.

Source of support: Nil

Conflict of interest: None

Let us ponder over this question to decide the appropriateness of a surgical technique. We may review some pertinent literature from the realm of research methodology.

Major Greenwood, a medical statistician, wrote in 1923 that "…I should like to shame [surgeons] out of the *comic opera* performances which they suppose are statistics of operations." Only when the quality of publications in the surgical literature has improved will surgeons reasonably be able to rebut the charge that as much as half of the research they undertake is "misconceived."[1]

Of all the medical disciplines, surgery contributes significantly to improve health and well-being all over the world. In order to do so efficiently and reliably, decisions about surgery are informed by high-quality outcome data.[1]

In order to bring the objectivity in research practices into the operating theater, there is a need to change the attitude of the surgeons and others. For surgeons, the culture of research and evaluation should be inculcated early in their training so that it becomes embedded in their professional pursuits. The ethical committees should be flexible to consider protocols on surgical innovations promptly. Institutions should provide support to surgical innovations and facilitate early registration of new surgical procedures. Methodologists must be sensitive to special issues and concerns of surgeons, such as difficulty in blinding the patient (we need the specific consent for a particular operation, incision of a big laparotomy compared to a laparoscopic small cut, cannot be hidden from a patient and treating team of doctors and nurses examining the patient in follow-up, and those assessing the outcomes of a trial).[1]

Horton also blamed the surgeons stating, "if almost half of surgical research truly resides in the case series and if this study design is the most vulnerable to criticism, logic insists that a large proportion of the surgical literature is of *questionable* value."[2]

Bunker et al. challenged surgeons to improve the evaluation of their surgical interventions and asked surgeons to carry out a rigorous assessment of the costs, benefits, and risks of operations.[3]

Professor Michael Baum, a famous teacher of breast surgery in London, wrote "results of a properly designed randomized trial will enable the intelligent clinician to weigh up the benefits against the harms and to keep the patient with sufficient knowledge to make rational decisions based on his or her own values, fears, and expectations."[4]

Pragmatic vs Explanatory Trials

Most of the earlier randomized controlled trials were designed and conducted to measure the "efficacy" of the intervention compared to the placebo or "standard of care." These trials were called explanatory trials and were designed to assess the efficacy of a therapy in the ideal world, the procedure being performed on a highly selected subgroup of patients, in a highly selective and controlled environment by expert and well-trained or skilled surgeons or physicians.[5]

In order to assess, as to how effective the same operation or same therapy would prove in the "real life" or in the "real world" where doctors or surgeons of average competence treat patients under ordinary hospital environment, pragmatic trials were evolved. Put simply, the pragmatic trials are meant to evaluate the effectiveness rather than efficacy in the real world. The degree of trial pragmatism can be evaluated following the acronym PRagmatic Explanatory Continuum Indicator Summary (PRECIS-2).[5]

In the case of thyroid cancer surgery, this is exactly what we need: to arrange a large multicenter pragmatic trial to be conducted in centers with surgeons of reasonable competence in performing thyroid surgery. This is of particular importance because the majority of thyroid surgery is conducted by surgeons of average competence in general surgical units. So far the claim about "wonderful" results of total thyroidectomy as the ideal operation for cancer is based on the outcome of operation performed by highly trained and skilled thyroid surgeons in very few selected centers, the "ivory towers of medicine." We want a safe and effective operation that is well within an affordable reach of the masses both in the low- and middle-income countries as well as in the affluent West.

Does it ever occur to us that the technique of thyroid surgery initiated and developed by Theodore Kocher goes on only with slight modifications till today, nobody ever questioned the rationale of a particular step of the operation or its justification (except Dr Ramakant asking the rationale of raising a skin flap in thyroid surgery and evolving "flapless thyroidectomy"). In fact, very few critical evaluations of the surgical techniques of parathyroidectomy, adrenalectomy, or pancreatic tumor removal have been undertaken through the rigor of science and the balance of randomized controlled trials.

Ever since Ronald Fisher proposed the concept and conduct of randomized trials in medicine and Sir Austin Bradford Hill conceived the first trial of streptomycin in pulmonary tuberculosis, way back in 1948, surgeons have embarked upon several some large and some not so large randomized trials to evaluate the usefulness, harms, and lately the cost-effectiveness of various surgical procedures.[6] As students in surgery, we read with great interest and enthusiasm The National Surgical Adjuvant Breast and Bowel Project (NSABP) trials on breast cancer care starting as early as 1971 (NSABP B04 trial comparing radical mastectomy with total mastectomy and total mastectomy + radiotherapy) in America and the famous Milan trial in Italy by Dr. Umberto Veronesi starting in 1973 (Milan I trial comparing QUART with Halsted Radical mastectomy). Even procedures, like highly selective vagotomy, midline vs paramedian incisions for laparotomy, mass vs layered closure techniques of laparotomy wound, prevention of surgical site infection with tetracycline peritoneal and wound lavage, stapled vs handsewn low anterior rectal excision, were all critically appraised by appropriately sized randomized controlled trials.

I find that most of the surgical maneuvers described to treat afflictions of the endocrine organs are based on personal choice and prospective cohort data/case series. The surgical technique is largely influenced by the teachers of a center where the surgeon received her/his initial formative training. The American Thyroid Association guidelines are not based on the evidence generated from randomized controlled trials.

I give some examples of the possible situations where the results of a well-conceived and conducted randomized controlled trial may modify the approach to patients with endocrine diseases. For example, I take a patient with benign nodular goiter where the predominant nodule is confined to one lobe with only very tiny nodules (not palpable only visible on ultrasound scan) located in the opposite thyroid lobe. Most of the endocrine surgeons would perform a total thyroidectomy in this patient. To be the devil's advocate, I may argue against this approach and perform a randomized controlled trial where such patients are randomized to either total thyroidectomy or hemithyroidectomy removing only the grossly enlarged thyroid lobe. The putative benefits would be far too many: the risk of damaging superior and recurrent laryngeal nerves, and the damage to parathyroid glands would all be reduced by half by a simple rule of arithmetic and common sense. I am aware of the long-term follow-up studies of patients harboring very small papillary cancer (<1 cm) for up to 30 years in Japan with very little chance of progression from this microscopic disease to overt malignancy. This being the case why can we not leave tiny microscopic or subcentric sonographically detected nodules in the seemingly normal thyroid lobe to Mother Nature. One can easily operate as and when these nodules become enlarged or cause compression on the airway or food passage or result in a cosmetic deformity.

The outcome of such randomized controlled trials could be recorded as PRO, or QOL data of speech analysis, serum calcium and serum parathormone levels, the need to administer intravenous calcium and thyroxine for replacement (obviously, it would be needed in all cases undergoing total thyroidectomy, with only some patients in the hemithyroidectomy group depending on serum thyroid-stimulating hormone (TSH) levels).

A similar trial will be well in order for dealing with early differentiated thyroid cancers carrying a low probability of recurrence. Thus, patients with tumors <4 cm in largest diameter may be randomized to either undergo total thyroidectomy or only hemithyroidectomy.

One can envisage similar randomized controlled trials in the realm of treating hyperparathyroidism by single-side neck exploration vs bilateral neck exploration. One could plan an operation with intraoperative image guidance vs not; surgery done by "Mr ROBOT" or by human; direct neck incision vs endoscopic minimal access; and so on. The imagination of the human mind being the limit.

In the case of suprarenal glands, a trial of open vs minimal access approach may be well in order. Not only the access/approach to a gland may be the theme of a research question but we may critically evaluate the tools of surgical dissection, viz. diathermy, sharp dissection or more modern energy vessel sealing devices and clips, and unilateral vs bilateral removal.

In order to bring home this need for decision-making based on the high-quality evidence, we made a humble endeavor in the area of early thyroid cancers at the AIIMS, New Delhi.

The basis for the systematic approach to papillary carcinoma thyroid was laid by Dr EL Mazzaferri in his retrospective, multi-institutional, registry-based observational studies published in the 1970s. Ever since the world has been following the dictum of "total thyroidectomy with radioiodine ablation, follow-up body scan, and lifelong TSH suppression with thyroxine."

There is an urgent need to generate evidence in the following areas of thyroid cancer surgery:

- Initial workup: Ultrasound vs ultrasound and computed tomography scan and fine needle aspiration cytology vs needle biopsy
- The extent of thyroid removal: Total vs less
- The extent of lymph node clearance
- Adjuvant therapy
- The surveillance strategy for follow-up.

Let us ponder over various obstacles in the way of a randomized controlled trial in thyroid cancer:

- Rarity of disease: Thyroid carcinoma usually constitutes a small fraction of all the cancers in the body, and it would require a very large multicentric effort to accumulate a sufficient number of patients for a long time. Hence, most clinicians have been hesitant to start a randomized controlled trial on this disease.
- Rarity of the event of death: If we choose 5, 10, or even 20 years of mortality as the main outcome, we will have to wait for a very long time (many of us would retire by the end of the study!) as the death in thyroid cancer is extremely rare.
- Most of the follow-up data from cohort studies suggest nearly similar survival in patients operated by total or subtotal thyroidectomy. According to the principles of statistics, the sample size needed to demonstrate the significance is very large when the difference between the two or more groups is very small. In one estimate, about 8,000 patients would be needed to be randomized and followed up for 20 years to demonstrate a statistically significant difference in the survival between total thyroidectomy and hemithyroidectomy. Let us take an example of calculating the sample size for a two-arm parallel design randomized trial with superiority hypothesis (here the null hypothesis is that the two or more treatments have the same effect and alternate hypothesis is that one treatment is superior to other):

The sample size in a randomized controlled trial with Superiority Hypothesis $= 16 \times p \,(100 - p)/(d \times d)$ in each arm, where "p" is the average response and can be computed as the mean of the two groups $= (p_1 + p_2)/2$ and the "d" is the difference between response of two $= (p_1 - p_2)$.

Suppose survival in total thyroidectomy group $= 90\%$ (p_1) and in hemithyroidectomy $= 85\%$ (p_2), then the average survival $p = (90 + 85)/2 = 87.5$ and $d = 90 - 85 = 5$.

Substituting these values of "p" and "d" in the formula for sample size calculation, we get 700 patients to be recruited in one arm of the study. Usually, we add about 10% as lost to follow-up $= 70$ patients. Adding this number, we get a total of 770 patients to be randomized in one arm or 1,540 patients in total to be randomized in a trial to demonstrate a difference in survival of 5% between the two groups with 95% confidence (5% alfa error) and 80% statistical power.

With a smaller difference in survival between the two groups, the sample size would inflate further.

In order to circumvent this problem, we considered that we may use regional local recurrence in the neck as the outcome of main (primary) interest.

In general, the randomized controlled trial with a noninferiority hypothesis requires a smaller sample as we envisage a one-tailed hypothesis. In a superiority hypothesis trial, we consider a two-tailed hypothesis.

A noninferiority trial is a clinical trial wherein the objective is to establish that the experimental treatment is not clinically worse than the active comparison treatment by more than a small, predetermined margin. The basic idea behind such trials is that if treatment A is similar enough in efficacy to treatment B that the difference between the clinical outcomes is negligible, then a patient can use the alternative with fewer side effects/morbidity, reduced cost, simpler logistics, or improved quality of life. The noninferiority margin is the degree of efficacy that the investigator is willing to give up in exchange for the potential benefits offered by the new treatment modality. Choosing such a margin is difficult, as there are no explicit rules. Usually, findings from earlier studies and estimates of clinically relevant differences are combined for choosing a noninferiority margin.[7]

The sample size for a noninferiority trial is usually calculated under the assumption that the experimental and control treatment have equal effects. Under the assumption that the new treatment is noninferior, as is often the case for a new treatment modality, the required sample size decreases considerably. The sample size required for a superiority trial to demonstrate the small benefit is nearly 10 times larger than that required for the noninferiority trial and around four times as large for the larger effect.

Many landmark trials in surgery and other specialties have used noninferiority trials, such as COlon cancer Laparoscopic or Open Resection (COLOR) trial, Targeted intraoperative radiotherapy vs whole-breast radiotherapy for breast cancer (TARGIT-A trial), Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (AMAROS) trial, and The Continuous Infusion vs Double-Bolus Administration of Alteplase (COBALT) trial.

We combined these ideas and embarked upon a noninferiority trial with recurrence in the neck as the outcome of interest.

One of the main objectives of surgery and adjuvant treatment in thyroid cancer is to maintain sufficient QOL along with long-term survival. The assessment of QOL in thyroid cancer is important because it provides detailed information about patients' perceptions of their health. The quality of life is a multidimensional concept with different meanings according to the diversity of life contexts, which includes the maintenance of functional capacity, general satisfaction, personal fulfillment, and social interaction.

PRO has become important endpoints in comparative effectiveness research and in patient-centered health care.

It is observed that developing countries are neglecting numerous opportunities for improving health and better allocation of scarce resources that can achieve better health outcomes. The cost-effectiveness analysis acts as an important tool for identifying these neglected opportunities by highlighting interventions that are relatively inexpensive yet have the potential to reduce the disease burden substantially. This rule can also be applied in the management of thyroid cancer patients.

Hence, we decided to take forward the task of formulating level 1 evidence in the form of a randomized controlled trial involving the quality of life and cost-effectiveness to find out the most appropriate extent of thyroid surgery in the case of differentiated carcinoma of thyroid gland confined to one lobe without any high-risk features without compromising oncological safety.

Research Question

Is near-total thyroidectomy more cost-effective and associated with better quality of life postoperatively than total thyroidectomy without compromising significantly on oncological outcome in patients with differentiated thyroid carcinoma less than 4 cm confined to one lobe?

Objective

To ascertain and compare quality of life and cost-effectiveness along with postoperative complications and local recurrence rates in thyroid cancer patients treated by near-total thyroidectomy plus radio-ablation of the residuum vs total thyroidectomy.

Aims

- **Primary Outcome Variables:**
  - Quality of life at the baseline and 6 months postoperatively
  - Cost of care in the perioperative period
  - Hypocalcemia in the postoperative period requiring intravenous calcium therapy.
  - Voice Change assessment by speech analysis to evaluate the damage to recurrent laryngeal nerve and superior laryngeal nerves.
- **Secondary Outcome:**
  - Local recurrence rate in the neck at 2 and 5 years postsurgery.

Study Design

Noninferiority randomized controlled trial

- Arm A: Total thyroidectomy
- Arm B: Near-total thyroidectomy + radioiodine ablation of the residuum

Sample size calculation formula for a noninferiority study: In the example below, we have computed the sample size on local neck recurrence at 5 years (a secondary outcome in our study). However, most studies compute the sample size on the primary outcome not on the secondary outcome.

A noninferiority design has the following null hypothesis (H0).

H0: $\varepsilon \leq \delta$ and alternative hypothesis HA: $\varepsilon > \delta$, where ε is the difference in the proportion of success between two therapies. $\varepsilon = p_1 - p_2$; where

p_1 = success rate with near-total thyroidectomy = 50%, assuming 50% recurrence at 5 years. p_2 = success with standard total thyroidectomy = 60%, assuming 40% recurrence in 5 years. Assuming: $p_1 = 50\%$ or 0.5.

$p_2 = 60\%$ or 0.6.

The values of other factors were taken as follows:

- $Z\alpha = 1.645$ for $\alpha = 0.05$ for a one-tailed hypothesis
- $Z_\beta = 0.84$ for power = 80% for one- or two-tailed hypotheses

With all these values, the formula used in the calculation of sample size for one group was as follows:

$$N = (Z_\alpha + Z_\beta)\, 2[p_1(1 - p_1) + p_2(1 - p_2)]/[(p_1 - p_2) - \delta]2$$

Keeping a margin of noninferiority δ (δ pronounced as "delta") = 10%; sample required = 76 cases per group.

One may compute the sample size on the desired cost-effectiveness ratio also, as is done in health economics studies.[8,9]

The total cost can be calculated as the sum of the direct cost and indirect costs.

Cost-effectiveness (CE) can be calculated for each procedure as:

CE = Total cost incurred in one group/Number of successful procedures done in that group.

A successful procedure can be defined as one without hypocalcemia or voice changes in the immediate postoperative period. Success in a cancer trial may also include no recurrence for 2 or 5 years.

The advantage of near-total thyroidectomy over total thyroidectomy and lobectomy should be considered with a broad societal perspective keeping in mind the fact that most low-risk thyroid carcinomas are operated by general surgeons in tier II and III cities. These surgeons are different from the high-volume center surgeons in the Departments of Endocrine surgery in centers of excellence.

Summary

We may embark upon a multicenter randomized controlled trial with noninferiority hypothesis and randomize patients in three arms: (1) total thyroidectomy; (2) Hartley Dunhill procedure with radio-ablation of the residuum; and (3) Hartley Dunhill procedure alone.
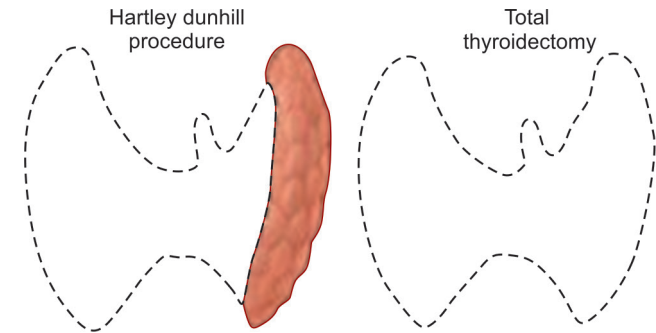
The main outcome of analysis should be PRO and QOL and cost-effectiveness must be computed to find the affordability of a particular regimen adopted.

References

1. Surgical research: the reality and the IDEAL. Lancet 2009;374(9695):1037. DOI: 10.1016/S0140-6736(09)61678-0. PMID: 19782850.
2. Horton R. Surgical research or comic opera: questions, but few answers. Lancet 1996;347(9007):984–985. DOI: 10.1016/s0140-6736(96)90137-3. PMID: 8606606.
3. Bunker JP, Barnes BA, Mosteller F. Costs, risks and benefits of surgery. Oxford: Oxford University Press; 1977.
4. Baum M. Reflections on randomised controlled trials in surgery. Lancet 1999;353(Suppl. 1):SI6-SI8. DOI: 10.1016/s0140-6736(99)90220-9. PMID: 10319923.
5. Ford I, Norrie J. Pragmatic Trials. N Engl J Med 2016;375(5):454–463. DOI: 10.1056/NEJMra1510059. PMID: 27518663.
6. Hill AB. Memories of the British streptomycin trial in tuberculosis. The first randomised clinical trial. Control Clinical Trials 1990;11(2):77–79. DOI: 10.1016/0197-2456(90)90001-I.
7. Hahn S. Understanding noninferiority trials. Korean J Pediatr 2012;55(11):403–407. DOI: 10.3345/kjp.2012.55.11.403.
8. Ramsey SD, McIntosh M, Sullivan SD. Design issues for conducting cost-effectiveness analyses alongside clinical trials. Annu Rev Public Health 2001;22(1):129–141. DOI: 10.1146/annurev.publhealth.22.1.129.
9. Noor Haslinda I, Juni MH, Rosliza AM, et al. Designing and conducting cost-effectiveness analysis studies in health care. Int J Public Health Clin Sci 2017;4(5):62–76.