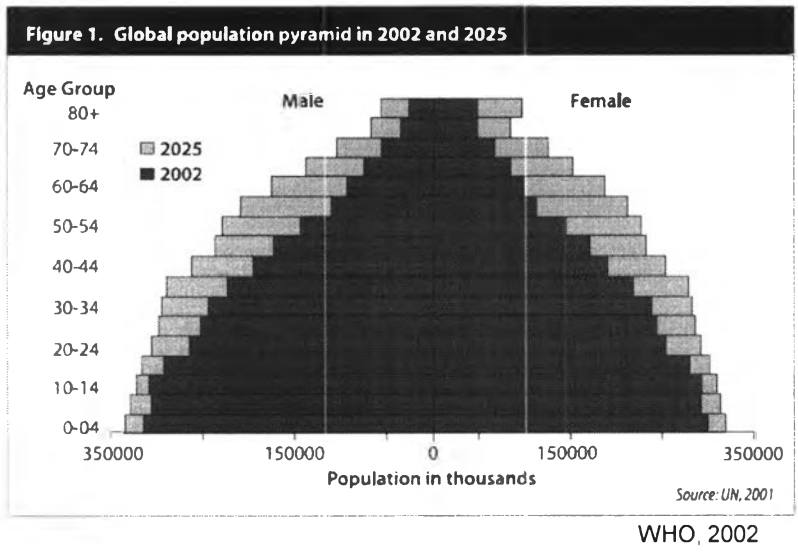
CHAPTER II

LITERATURE REVIEW

2.1 AGING

The world is experiencing demographic revolution, with the number of the elderly population increasing dramatically (WHO, 2002). China has also marched to an ageing society; at the end of 2006, there were about 104.2 million people over 65 years of age, which accounted for 7.9% of the total population (NBS China, 2006). Ageing is accompanied by physiological, psychological, social as well as economic changes, which determine that the elderly population may become vulnerable to inadequate nutrition (Chen et al., 2001; Brownie, 2006).



WHO, 2002

Figure 2 Global population pyramids in 2002 and 2025

2.2 EDENTULOUS

The edentulous patient has not disappeared. While the prevalence of edentulism is less than what it was 20 years ago, about 33 percent of Americans older than 65 years of age were completely edentulous as of 2000 (U.S. Surgeon General, 2000; Tomar, 1997) The prevalence of edentulism in elders is reported to range from 25% to 30% in North America (Beltran-Aguilar et al., 2005; Millar et al., 2005) and from 15% to 72% in Europe (Muller et al., 2007). Tooth loss is associated with increasing age and negative effects on OHRQoL, while increasing age alone is associated with fewer negative effects on OHRQoL (Steele et al., 2004). A 1992 national survey found that 10.5% of Americans aged 18 to 74 years are completely edentulous, 7.2% are edentulous in one arch, and 21.4% use removable complete or partial denture prosthesis. Among those 75 years or older, 43.9% are edentulous in both arches (Redford M et al, 1996).

Table 2 Epidemiology of Edendulousness

Epidemiology of Edentulousness

WHO region/Country	Edentulous (%)	Age group (year)
Europe		
Albania	69	65+
Austria	15	65–74
Bosnia and Herzegovina	78	65+
Bulgaria	53	65+
Denmark	27	65–74
Finland	41	65+
Hungary	27	65–74
Iceland	15	65–74
Italy	19	65–74
Lithuania	14	65–74
Poland	25	65–74
Romania	26	65–74
Slovakia	44	65–74
Slovenia	16	65+
UK	46	65+
Africa		
Gambia	6	65+
Madagascar	25	65–74
The Americas		
Canada	58	65+
USA	26	65–69
Eastern Mediterranean		
Egypt	7	65+
Lebanon	20	64–75
Saudi Arabia	31–46	65+
South-East Asia		
India	19	65–74
Indonesia	24	65+
Sri Lanka	37	65–74
Thailand	16	65+
Western Pacific		
Cambodia	13	65-74
China	11	65–74
Malaysia	57	65+
Singapore	21	65+

WHO, 2005

As above, the prevalence rate of complete edentulous senior Thai people above 65 years is 16 percent of total population (WHO, 2005).

Continuous alveolar bone resorption following tooth removal may eventually result in an impaired bearing area for full dentures (Tallgren, 1972). The effect will be a decrease of denture stability and retention. Improper prosthesis adaptation may cause increased discomfort, including pain and problems with basic functions such as speech and mastication. Resorption of the alveolar ridges may also compromise the patient's facial appearance. Alveolar bone resorption tends to affect the mandible more severely than the maxilla. This is probably related to a smaller denture-bearing area and a less favorable distribution of occlusal forces to the bone (Tallgren, 1972). Additional factors determining patients' satisfaction with their dentures include denture quality, the degree of positive dentist-patient interaction, previous denture experiences, and the patient's psychological well-being and personality (Vervoom et al., 1988).

The meticulous construction of a well-fitting new set of dentures frequently is sufficient to resolve the patient's problems. Sometimes minor corrections of the denture-supporting area are needed to optimalize the condition of the oral tissues for their supporting function. However, when this does not give sufficient relief, more extensive surgical and prosthetic procedures may be necessary to enlarge the denture-bearing area and allow construction of a proper fitting prosthesis to satisfy the patient's functional, esthetic, and comfort requirements. These procedures include preprosthetic surgery, such as vestibuloplasty and deepening of the floor of the mouth. In general, a vertical mandibular height of at least 15 mm is required for a successful vestibuloplasty (Stoelinga, 1984).

2.3 ORAL HEALTH-RELATED QUALITY OF LIFE (OHRQoL):

Oral health-related QOL is one dimension of a wider concept of QOL (Bowling 2005), and is defined in relation to optimum levels of mental, physical, role, and social function; it includes relationships, as well as perceptions of health, fitness, life satisfaction and wellbeing (Bowling, 2001). Oral health-related QOL is defined as an individual's assessment of how the following affect his or her wellbeing: functional factors, psychological factors, social factors, and experience of pain/discomfort in relation to orofacial concerns (Inglehart et al., 2002). OHRQoL characterizes an individual's perception of buccal health, and can be used as indicator of the advantages of the prosthetic rehabilitation strategies (John et al., 2004; Strassburger et al., 2006). The Oral Health Impact profile (OHIP) questionnaire is one of the most technically sophisticated instruments for assessment of OHRQoL (Locker, 1995).

The OHIP was developed in Australia by Slade and Spencer in 1994, and several versions of the tool have been developed, one of which was recently translated into Brazilian Portuguese (Pires et al., 2006). The tool comprises 49 questions distributed into seven sub-scales. Among the short version that have been developed, such as the OHIP-14, the OHIP-EDENT is seen as the most appropriate for edentulous patients, as it presents a set of specific questions (Allen and Locker, 2002). In 2007, Souza et al. translated the OHIP-EDENT into Portuguese, and this version was then back-translated into English. The OHIP-EDENT is a 19-question survey, grouped as seven subscales or domains: functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability, and handicap (Slade et al., 1994; Pires et al., 2006; Souza et al., 2007; Allen and Locker, 2002).

Table 3 The 21 items from the original OHIP [Slade and Spencer, 1994] that were selected for the OHIP-G 21 [John et al., 2006]

Selection of OHIP-items, original numbering	Questions on the OHIP-G21 (each question below ended with the phrase: "because of problems with your teeth, mouth, or dentures")
Functional limitation questions	
1	Have you had difficulty chewing any foods…..
2	Have you had trouble pronouncing any words…
3	Have you noticed a tooth which doesn't look right?
4	Have you felt your apperance has been affected...
Physical pain	
10	Have you had painful aching in your mouth?
11	Have you had a sore jaw?
13	Have you had sensitive?
14	Have you had toothache?
15	Have you had painful?
17	Have you had sore spots in your mouth?
Psychological discomfort	
19	Have you been worried by dental problems?
20	Have you felt uncomfortable about the appearance…
Psychological disability	
36	Have you felt depressed…
37	Have your concentration been affected…
38	Have you been a bit embrrassed….
Social disability	
39	Have you avoided going out…
40	Have you been less to lerant of your spouse or family….
42	Have you been a bit irritable with other people…
43	Have you had difficulty doing your usual jobs…
Handicap	
48	Have you been totally unable to function…
49	Have you been unable to work to your full capacity…

2.4 CONVENTIONAL DENTURES

Allen and Mc. Millian (2003) have shown that patients who requested implants but received conventional dentures reported little satisfaction with dentures and only modest improvements in the quality of life.

The disadvantages of the complete removable denture are:

-Extensive detail required for proper fabrication

-Lack of stability (especially in mandible)

-Lack of retention (especially in mandible)

-Continued loss of alveolar bone leading to further instability and lack of retention

-Patients using such dentures may be led to believe professional dental care no longer is needed

-Lack of chewing function when ill-fitting

-Social concerns (slippage, unnatural appearance) (James et al., 2003).

2.5 IMPLANT SUPPORTED OVERDENRURES (IOD)

The Advantages of the implant-supported overdenture are:

- As few as two to four implants may be used for support
- Good stability
- Good retention
- Improved function
- Improved esthetics
- Reduced residual ridge resorption
- Simplest implant-supported prosthesis

- Possible incorporation of existing denture into the new prosthesis (James et al., 2003).

Disadvantages are: Difficult to maintain without frequent adjustments and repair to the attachment mechanism between the implants and the denture. (Walton et al., 1993; Watson et al., 1997; Walton et al., 1997; Bergedal et al., 1998; Naert et al., 1999)

The risk factors for failure of dental implants are:

- Smoking
- Factors that affect healing of bone (such as diabetes, use of steroids)
- Untreated periodontal disease
- Anatomy (if bone in recipient site is inadequate, grafting may be necessary)
- Poor bone quality
- Inadequate practitioner training, experience or both
- Patient compliance concerns

In one literature indicated that implant-supported overdentures in the mandible provide predictable results with improved stability, retention, function and patient satisfaction compared with conventional dentures. When planning treatment for patients with edentulous mandibles, clinicians should consider the implant-supported prosthesis. (James et al., 2003)

The use of osseointegrated dental implants to support a dental prosthesis has become an accepted treatment modality because of the high implant success rates observed by clinicians and researchers. (Adell et al., 1990; Cox et al., 1987; Arvidson et al., 1992; Babbush et al., 1993)

A few retrospective studies have also indicated improvements in overall satisfaction of patients and their perception of chewing hard foods (Harle et al., 1993; Wismeijer et al., 1992). In one metaanalysis, they identified from 281 articles, out of those 18 peer-reviewed articles met prespecified criteria for inclusion. Specific out-comes of significance from these articles rated as level II are: (1) fixed implant-supported partial dentures do not provide significant improvement in masticatory performance compared to conventional removable partial dentures for Kennedy Class I and II partially edentulous mandibles; (2) the combination of a mandibular implant-supported or retained overdenture (IOD) and maxillary conventional complete denture (CD) provides significant improvement in masticatory performance compared to CDs in both the mandible and maxilla for a limited population having persistent functional problems with an existing mandibular CD due to severely resorbed mandible; and (3) the type of implant and attachment system for mandibular IODs has a limited impact. Specific outcomes of significance identified by articles rated as having a moderate level of evidence (level III) are: (1) mandibular fixed implant-supported complete dentures provide significant improvement in masticatory performance compared to mandibular CDs in subjects dissatisfied with their CDs; and (2) implant-supported mandibular resection dentures have an advantage over conventional dentures in masticatory performance on the defect side of the mouth (Fueki et al., 2007) Upon completion of treatment, both groups reported improvement ($p < 0.001$, Wilcoxon Ranks Sum test) in oral-health-related quality of life and denture satisfaction. There were no significant post-treatment differences between the groups, but a treatment effect may be masked by application of "intention to treat" analysis.

2.6 IOD in DIABETES MELLITUS

Abnormalities in chemotaxis, phagocytosis, and bacterial activity of polymorphoneuclocytes have been shown in poorly controlled diabetics (Mowat et al., 1971; Van Oss et al., 1971; Bagdade et al., 1974). Higher prevalence of osteoporosis in diabetics remains a controversial matter. Many studies have shown that periodontal disease is more prevalent and severe in patients with diabetes mellitus than in nondiabetics, but the promoting factors have not been fully understood (Seppala et al., 1993; Shlossman et al., 1990; Emrich et al., 1991; Thorestensson et al., 1993; Seppala et al., 1994; Tervenon et al., 1993).

2.7 PATIENT SATISFACTION

Population-based oral health studies have frequently defined satisfactory oral health as the presence of a minimum of 20 teeth or a particular number of contacting posterior pairs of teeth (Sheiham et al., 2001; Shimazaki et al., 2001). Recently, a randomized clinical trial in a selective population of dissatisfied denture wearers has shown the functional superiority of mandibular implant-retained overdentures in terms of patient satisfaction and their ability to comminute test food (Boerrigter et al., 1995; Geertman et al., 1994).

The results of one study demonstrated that satisfaction was correlated with age, gender, and past prosthetic history in the patients rehabilitated with the implant-supported overdentures. It shown that gender had significant correlation with comfort ($p < .0001$), years of being edentulous prior to implant/prosthetic treatment had direct effects on the general satisfaction and satisfaction of comfort ($p < .01$), one extra denture used before implant treatment resulted in less comfort ($p < .01$) and poorer function ($p < .001$). Elders

were more satisfied with aesthetic (OR = 0.96) and comfort (OR = 2.96). Number of adjustment appointments had a positive correlation with comfort ($p < .001$) (Siadat et al., 2008).

After receiving new dentures with mandibular implant supports, improved satisfaction "within subject" was prompt, durable, substantial, and statistically significant, regardless of the attachment mechanism, and with or without a reinforcing framework (MacEntee et al., 2005).

2.8 SEX DIFFERENCES IN DENTURE SATISFACTION

Elderly females are less satisfied with conventional dentures than elderly males with regards to aesthetics and ability to chew, but equally satisfied with implant overdentures. At 6 and 12 months after delivery, elderly edentulous males and females wearing mandibular implant overdentures were significantly more satisfied than those wearing conventional dentures (Pan et al., 2008). At baseline, females rated their ability to chew, aesthetics and ability to speak significantly lower than the males ($P < 0.05$). At 6 months post-denture delivery, females in the CD group gave lower ratings for ability to chew, aesthetics, general satisfaction, comfort and stability than the males. However, in the IOD group, males and females rated their general satisfaction and all six subcategories equally. At 12 months after denture delivery, females and males in the IOD group still rated their denture satisfaction equally. However females in the CD group still rated ability to chew and aesthetics significantly lower than the males.

Sex differences have been shown in many studies on oral health. In many countries, more females than males are edentulous. This male–female difference may

reflect a higher proportion of older females in a population (Millar et al., 2005). It was also reported that males are less concerned about their edentulism, less likely to operate for restorations and less likely to visit a dentist than females. It has been suggested that females report pain symptoms more willingly than males, and that they recall health problems to a greater extent than males do (Bingefors et al., 2004).

In one study, the effect of oral prostheses on the quality of life of head and neck cancer patients was investigated. It was reported that in both a cancer group wearing maxillofacial prostheses, as well as in a control group wearing conventional dentures, females rated most variables lower than the males (Moroi et al., 1999). What may cause these differences in ratings by females and males? Those may be explained by either physical or psychological differences between the sexes. It has been suggested that variety of factors may contribute, including hormonal alterations, (Kuba et al., 2005) blood pressure (Fillingim et al., 1996) and psychological factors (Logan et al., 2004). Furthermore, sex role expectancies and anxiety may moderate sex differences (Logan et al., 2004).

After the extraction of teeth, residual ridge resorption in the alveolar portion of the jaw begins. Compared with males, elderly females are at a higher risk of severe resorption in the edentulous mandible than males (Xie et al., 1997). The bone mineral content in edentulous females' mandible also decreases with aging, while that of males' increases slightly. Osteoporosis may be regarded as a cofactor of residual ridge resorption in women (Solar et al., 1994).

2.9 MALNUTRITION

Malnutrition is defined as "the state of deficient energy or protein intake or absorption, which is characterized by weight loss and changes in body composition" (Lochs et al., 2006). Malnutrition leads to poor outcomes such as functional decline, frailty, the decline of quality of life and higher mortality (Persson et al., 2002; Ödlund Olin et al., 2002; Bartali et al., 2006).

A previous study, that using a common screening tool (i.e. serum albumin), has reported less than 1% of the healthy, community-dwelling elderly were at moderate risk of malnutrition (Sugiyama et al., 2000). Because one of the signs of malnutrition (e.g. the declines in muscle strength) can occur even within the normal range of serum albumin levels (Schalk et al., 2005). Although there are many factors associated with malnutrition, including demographic, physical and psychosocial factors such as age, living alone, eating and oral problems, low functional capacity, depression and more (Chen, 2005; Takahashi, 2006).

2.10 PAIN AND SWELLING AFTER IMPLANT PLACEMENT

Following dental implant placement, patients presented different degrees of pain and swelling as a direct consequence of surgery (Muller et al., 2001; Guarinos et al., 1998). In the series published by Muller et al (Muller et al., 2001), among 221 dental implants, 45% of the patients reported pain in the first 24 postoperative hours. Moreover, in another series (Peñarrocha et al., 2000), among 80 patients with 226 implants, maximum pain was recorded 6 hours after the operation, and was moderate intensity. Another study (Guarinos et al., 1998) comprising 70 patients and 163 implants reported peak pain after 12 hours (VAS intensity score = 3.5).

In one study examined postoperative pain and swelling in patients subjected to dental implant placement, and correlated to different clinical and surgical parameters. Analysis of the relation between pain 6 hours after the operation and the different study variables only showed a statistically significant correlation to the number of implants ($r=0.311$; $p=0.048$). Analysis of the relation between inflammation 48 hours after the operation and the different study variables showed statistically significant correlations to more advanced patient age ($r=0.386$; $p=0.013$), surgery in edentulous patients and free extremes ($F=7.293$; $p=0.002$), and to surgical approaches in the posterior sector of both jaws ($F=4.908$; $p=0.013$). A significant relation was also observed between swelling and the number of implants, when an increased number of implants were placed ($r=0.389$; $p=0.012$). In the series pain was found to be of maximum intensity after 6 hours, and was reported to be mild by 41.5% of the patients (González-Santana et al., 2005).

2.11 ECONOMIC ASSESSMENT IN DENTURES

COST

Direct and indirect costs of treatment and maintenance were calculated, in Canadian Dollars-CAN $, for each patient up to 1 yr after delivery of the prostheses in 1999 (Takanashi et al., 2004). Direct costs included labor, materials, drugs, laboratory work, and radiography. The time spent by the clinicians and the surgical assistant was measured, and opportunity costs for labor were estimated with the use of census data and other sources. A record of all drugs and disposable and re-usable materials was kept, and market prices were obtained. Indirect costs included the patients' transportation and cost of their time while receiving treatment. Overhead costs were calculated as a percentage of

the 'clinician time' cost. A detailed description of the techniques has been published (Takanashi et al., 2002, 2004).

Effectiveness

Subjects completed the OHIP-20 survey at the one-year follow-up appointment, and subscale and total sum scores were calculated without item-weighting (Awad et al., 2003; Heydecke et al., 2003). Lower scores represent better outcomes.

Cost-effectiveness Analysis

They assumed that no implants would be lost before death (Adell et al., 1990; Zarb and Schmitt, 1996), and that treatment would not alter longevity. From census data, we calculated that the average life expectancy of a 65-year-old Quebec-Canadian from the greater Montreal area would be 17.9 yrs. We also tested the effect of decreasing life expectancy by 5 yrs, and compared cost effectiveness separately in males and females, assuming remaining life expectancies of 15.7 and 19.8 yrs, respectively (Statistics Canada, 1996). It is recommended that all costs and outcomes be discounted, starting with the second year, to reflect loss of capital (Gold et al., 1996). There is some debate about the rate of this 'social discount' that should be applied (Gold et al., 1996); 3% is often used, but we also calculated the effect of using a discount rate of 5%.

To express the accumulation of costs and outcomes over the remaining life of the patient, they computed the present discounted value (PDV) of both costs PDVc and outcomes PDVo. The PDV is the weighted sum of a given variable discounted over time. The standard formula (Drummond et al., 1997) to compute the PDV is:

$$\sum_{i}^{LEXP} \frac{X_i}{(1+r)^{i-1}}$$

Where LEXP is the life expectancy in years, *Xi* is the value of the variable (cost or outcome) in year i, and r is the discount rate. They converted PDVc and PDVo into their respective constant annual flows, referred to as the equivalent annual value for cost (EAVc) and for outcome (EAVo), using the formula

$$E = \frac{K}{A(n,r)}$$

Where *E* is the equivalent annual cost/outcome, *K* the purchase price, and *A(n,r)* the annuity factor (n expected life years at interest rate r) (Drummond *et al.*, 1997). To calculate the cost-effectiveness ratio, we divided the between-group difference in EAVo by the difference in EAVc (Heydecke et al., 2005).

2.12 MEASUREMENT TOOLS

The Mini-Nutritional Assessment (MNA) was used to evaluate the risk of malnutrition. The MNA was developed in Europe to evaluate the nutritional status of the elderly (Guigoz et al., 1996) and translated into Japanese,(Kuzuya et al., 2005) including 18 items (range of scores, 0–30) consisting of four domains: (i) anthropometric; (ii) general; (iii) dietary; and (iv) subjective assessment. Participants were categorized by the MNA into three groups: (i) well-nourished (324 points); (ii) at risk of malnutrition (17–23.5 points); and (iii) malnutrition (<17 points). The validity of the MNA for malnutrition has been proven for the Japanese elderly; the sensitivity and specificity for

hypoalbuminemia was 0.81 and 0.86, respectively (Kuzuya et al., 2005). The MNA also has been shown to have good reliability according to internal consistency (Cronbach's alpha, 0.8) and test–retest reliability (intraclass correlation coefficient, 0.89) (Bleda et al., 2002). Anthropometric assessment of the MNA was conducted according to the MNA clinical practice user guide. The MNA (Guigoz et al., 1996) is a comprehensive tool that was originally developed for the assessment of nutritional status of older patients in clinics, nursing homes, hospitals, or among those who are otherwise frail. More recently, the MNA has also been used in the assessment of nutritional risk in community dwelling elderly populations (Delacorte et al., 2004).

Visual Analogue Scale (VAS)

The patients rated pain intensity based on a visual analog scale (VAS from 1 to 10) and verbal scale (1 = no pain, 2 = mild pain, 3 = moderate pain, 4 = intense pain), and inflammation as follows: 1 = none (absence of inflammation), 2 = mild (intraoral swelling in the surgical zone), 3 = moderate (extraoral swelling in the surgical zone), 4 = intense (extraoral swelling extending beyond the surgical zone). All scores were recorded 2, 4, 6, 12 and 24 hours after the operation, and on day 2, 3, 4, 5, 6 and 7.

Body Mass Index (BMI)

BMI was calculated as weight (kg)/[height (m)]2; BMI classification for adults of the Asia-Pacific Region was used(WHO, Obesity 2000).

Classification of BMI (Kg/meter^2)

< 18.5 (underweight)

18.5 – 22.99 (normal)

23.0 – 24.99 (overweight)

25.0 – 29.99 (level I obesity)

≥ 30.0 (level II obesity)

2.13.1 RANDOMIZED CLINICAL TRAIL (RCT)

The key distinguishing feature of the usual RCT is that study subjects, after assessment of eligibility and recruitment, but before the intervention to be studied begins, are randomly allocated to receive one or other of the alternative treatments under study. Random allocation in real trials is complex, but conceptually, the process is like tossing a coin. After randomization, the two (or more) groups of subjects are followed up in exactly the same way, and the only differences between the care they receive, for example, in terms of procedures, tests, outpatient visits, follow-up calls etc. should be those intrinsic to the treatments being compared. The most important advantage of proper randomization is that it minimises allocation bias, balancing both known and unknown prognostic factors, in the assignment of treatments." (Moher et al., 2010).

The terms "RCT" and randomized trial are often used synonymously, but some authors distinguish between "RCTs" which compare treatment groups with control groups not receiving treatment (as in a placebo-controlled study), and "randomized trials" which can compare multiple treatment groups with each other (Ranjith et al., 2005). RCTs are sometimes known as randomized control trials (Chalmers et al., 1981). RCTs are also called randomized clinical trials or randomized controlled clinical trials when they concern clinical research (Peto et al., 1976; Peto et al., 1977; and Wollert et al.,

2004). However, RCTs are also employed in other research areas such as criminology, education, social work and international development.

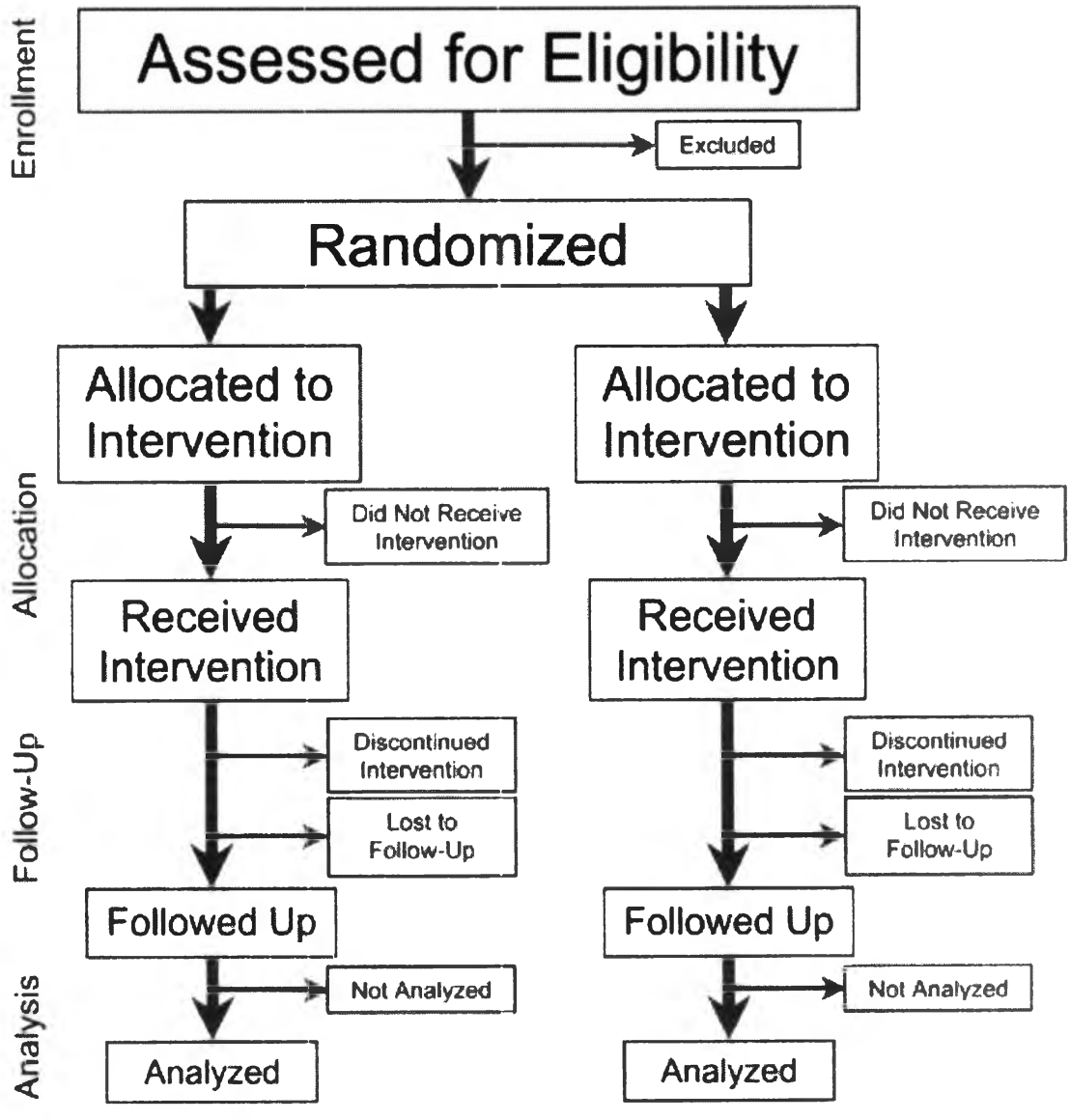


Figure 3 Flowchart of four phases (enrollment, intervention allocation, follow-up, and data analysis) of a parallel randomized trial of two groups, modified from the CONSORT (Consolidated Standards of Reporting Trials) 2010 Statement (Schulz et al., 2010).

2.13.2 CONsolidated Standards Of Reporting Trials (CONSORT)

The CONSORT statement was revised recommendations for improving the quality of reports of parallel-group randomized trials (Moher et al., 2001). There are as follows:

2.13.2.1 Title and Abstract

It included how participants were allocated to interventions.

(e.g., "random allocation", "randomized" or "randomly assigned").

Examples,

- In title: "Smoking reduction with oral nicotine inhalers: double blind, randomized clinical trials of efficacy and safety ..."
- In abstract: "Design: Randomized, double-blind, placebo-controlled trial"

They encouraged the use of structured abstracts when a summary of the report was required. Structured abstracts provide readers with a series of headings pertaining to the design, conduct, and analysis of a trial; standardized information under each heading. Some studies have found that structured abstracts are of higher quality than the more traditional descriptive abstracts) and that they allow readers to find information more easily.

2.13.2.2 Introduction

Scientific background and explanation of rationale: In the first part of the Introduction, authors should describe the problem that necessitated the work. The nature, scope, and severity of the problem should provide the background and a compelling

rationale for the study. This information is often missing from reports. Authors should then describe briefly the broad approach taken to studying the problem. It may also be appropriate to include here the objectives of the trial.

2.13.2.3 Methods

The study participants

(a) Eligibility criteria for participants

(b) The settings and locations where the data were collected

Settings and locations affect the external validity of a trial. Health care institutions vary greatly in their organization, experience, and resources and the baseline risk of the medical condition under investigation. Climate and other physical factors, economics, geography, and the social and cultural milieu can all affect a study's external validity. Authors should report the number and type of settings and care providers involved so that readers can assess external validity. They should describe the settings and locations in which the study was carried out, including the country, city, and immediate environment (for example, community, office practice, hospital clinic, or inpatient unit). In particular, it should be clear whether the trial was carried out in one or several centers ("multicenter trials"). This description should provide enough information that readers can judge whether the results of the trial are relevant to their own setting. Authors should also report any other information about the settings and locations that could influence the observed results, such as problems with transportation that might have affected patient participation.

The study intervention

Precise details of the interventions intended for each group and how and when they were actually administered. Authors should describe each intervention thoroughly, including control interventions. The characteristics of a placebo and the way in which it was disguised should also be reported. In some cases, description of who administered treatments is critical because it may form part of the intervention. For example, with surgical interventions, it may be necessary to describe the number, training, and experience of surgeons in addition to the surgical procedure itself. When relevant, authors should report details of the timing and duration of interventions, especially if multiple-component interventions were given.

The study objectives

Need to specific objectives and hypotheses. Hypotheses are more specific than objectives and are amenable to explicit statistical evaluation. In practice, objectives and hypotheses are not always easily differentiated.

The study outcomes

(a) Clearly need to define primary and secondary outcome measures.

The primary outcome measure is the prespecified outcome of greatest importance and is usually the one used in the sample size calculation. Some trials may have more than one primary outcome. Having more than one or two outcomes, however, incurs the problems of interpretation associated with multiplicity of analyses. Other outcomes of interest are secondary outcomes. There may be several secondary outcomes, which often include unanticipated or unintended effects of the intervention. All outcome

measures, whether primary or secondary, should be identified and completely defined. When outcomes are assessed at several time points after randomization, authors should indicate the prespecified time point of primary interest. It is sometimes helpful to specify who assessed outcomes (for example, if special skills are required to do so) and how many assessors there were. Where available and appropriate, previously developed and validated scales or consensus guidelines should be used, both to enhance quality of measurement and to assist in comparison with similar studies.

(b) When applicable, any methods to enhance the quality of measurements (e.g., multiple observations, training of assessors). Authors should give full details of how the primary and secondary outcomes were measured and whether any particular steps were taken to increase the reliability of the measurements. Death (from any cause) is usually easy to assess, whereas blood pressure, depression, or quality of life are more difficult.

Sample size

(a) How sample size was determined: For scientific and ethical reasons, the sample size for a trial needs to be planned carefully, with a balance between clinical and statistical considerations. Large samples are necessary to detect small differences. Elements of the sample size calculation are: the estimated outcomes in each group (which implies the clinically important target difference between the intervention groups), the alpha (Type I) error level, the statistical power (or the beta [Type II] error level), and for continuous outcomes, the standard deviation of the measurements. Authors should indicate how the sample size was determined. If a formal power calculation was used, the authors should identify the primary outcome on which the calculation was based, all the

quantities used in the calculation, and the resulting target sample size per comparison group.

(b) When applicable, explanation of any interim analyses and stopping rules should be mentioned. Many trials recruit participants over a long period. If an intervention is working particularly well or badly, the study may need to be ended early for ethical reasons. This concern can be addressed by examining results as the data accumulate. However, performing multiple statistical examinations of accumulating data without appropriate correction can lead to erroneous results and interpretations. If the accumulating data from a trial are examined at five interim analyses, the overall false-positive rate is nearer to 19% than the nominal 5%.

Randomization: Generation

(a) It is the method used to generate the random allocation sequence: Ideally, participants should be assigned to comparison groups in the trial on the basis of a chance (random) process characterized by unpredictability. Authors should provide sufficient information that the reader can assess the methods used to generate the random allocation sequence and the likelihood of bias in group assignment. However, readers cannot judge adequacy from such terms as "random allocation", "randomization," or "random" without further elaboration. Authors should specify the method of sequence generation, such as a random-number table or a computerized random-number generator. Only 32% of reports published in specialty journals and 48% of reports published in general medical journals specified an adequate method for generating random numbers. In almost all of these

cases researchers used a random-number generator on a computer or a random-number table.

(b) Details of any restriction of randomization should be mentioned. (e.g. blocking, stratification). For example, "Women had an equal probability of assignment to the groups. The randomization code was developed using a computer random number generator to select random permuted blocks. The block lengths were 4, 8, and 12 varied randomly . . .". Restricted randomization describes procedures used to control the randomization to achieve balance between groups in size or characteristics. Blocking is used to ensure that comparison groups will be of approximately the same size; stratification is used to ensure good balance of participant characteristics in each group. Stratification ensures that the numbers of participants receiving each intervention are closely balanced within each stratum. Stratified randomization is achieved by performing a separate randomization procedure within each of two or more subsets of participants (for example, those defining age, smoking, or disease severity).

Allocation concealment

Method used to implement the random allocation sequence: (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned. Allocation concealment seeks to prevent selection bias, protects the assignment sequence before and until allocation, and can always be successfully implemented.

Randomization implementation

It means, who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.

Blinding or Masking

(a) Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment: so that they will not be influenced by that knowledge. Blinding is important to prevent bias at several stages of a controlled trial. Blinding of patients is important because knowledge of group assignment may influence responses to treatment. Patients who know that they have been assigned to receive the new treatment may have favorable expectations or increased anxiety. Patients assigned to standard treatment may feel discriminated against or reassured. Use of placebo controls coupled with blinding of patients is intended to prevent bias resulting from non-specific effects associated with receiving the intervention (placebo effects). Blinding of patients and health care providers prevents performance bias. Blinding of patients, health care providers, and other persons (for example, radiologists) involved in evaluating outcomes minimizes the risk for detection bias, also called observer, ascertainment, or assessment bias. Finally, blinding of the data analyst can also prevent bias. Knowledge of the interventions received may influence the choice of analytical strategies and methods.

(b) If done, how the success of blinding was evaluated. Authors should report any failure of the blinding procedure, such as placebo and active preparations that were not identical in appearance.

Statistical methods

(a) Statistical methods used to compare groups for primary outcome(s). Example,

"All data analysis was carried out according to a pre-established analysis plan. Proportions were compared by using Chi-squared tests with continuity correction or Fisher's exact test when appropriate. Multivariate analyses were conducted with logistic regression. The durations of episodes and signs of disease were compared by using proportional hazards regression. Mean serum retinol concentrations were compared by t test and analysis of covariance. "Two sided significance tests were used throughout." Actual P values (for example, P = 0.003) are preferred to imprecise threshold reports.

(b) Methods for additional analyses, such as subgroup analyses and adjusted analyses. Examples,

"Proportions of patients responding were compared between treatment groups with the Mantel-Haenszel Chi-squared test, adjusted for the stratification variable. Authors should clarify the choice of variables that were adjusted for, indicate how continuous variables were handled, and specify whether the analysis was planned or suggested by the data.

Results

(a) Flow of participants through each stage

A diagram is strongly recommended. Attrition as a result of loss of follow up, which is often unavoidable. A recent review of RCTs published in five leading general and internal medicine journals in 1998 found that reporting of the flow of participants was often incomplete, particularly with regard to the number of participants receiving the allocated intervention and the number lost to follow up. Reporting was

considerably more thorough in articles that included a diagram of the flow of participants through a trial, as recommended by CONSORT.

(b) Protocol deviations:

Describe protocol deviations from study as planned, together with reasons. For examples, "There was only one protocol deviation, in a woman in the study group. She had an abnormal pelvic measurement and was scheduled for elective caesarean section. Authors should report all departures from the protocol, including unplanned changes to interventions, examinations, data collection, and methods of analysis. Some of these protocol deviations may be reported in the flow diagram. The nature of the protocol deviation and the exact reason for excluding participants after randomization should always be reported.

(c) Dates of recruitment and follow-up:

If the trial was stopped owing to results of interim analysis of the data, this should be reported. Early stopping will lead to a discrepancy between the planned and actual sample sizes. Also, trials that stop early are likely to overestimate the treatment effect.

(d) Baseline demographics and clinical characteristics of each group: This information allows readers, especially clinicians, to judge how relevant the results of a trial might be to a particular patient. Although proper random assignment prevents selection bias, it does not guarantee that the groups are equivalent at baseline. Any differences in baseline characteristics are, however, the result of chance rather than bias. Baseline information should be efficiently presented in a table. For continuous variables,

such as weight or blood pressure, the variability of the data should be reported along with the average values, mean and standard deviation.

(e) Number of participants: Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat". State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%). The number of participants in each group is an essential element of the results. Failure to include all participants may bias trial results.

(f) Summary of results: For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (e.g. 95% confidence interval). For each outcome, study results should be reported as a summary of the outcome in each group (for example, the proportion of participants with or without the event, or the mean and standard deviation of measurements) together with the contrast between the groups, known as the effect size. For binary outcomes, the measure of effect could be the risk ratio (relative risk), odds ratio, or risk difference. For survival time data, the measure could be the hazard ratio or difference in median survival time; for continuous data, it is usually the difference in means. For all outcome measures. authors should provide a confidence interval to indicate the precision (uncertainty) of the estimate. They are especially valuable in relation to nonsignificant differences, for which they often indicate that the result does not rule out an important clinical difference. Results should be reported for all planned primary and secondary end points, not just for

analyses that were statistically significant. For more information, visit www.consort-statement.org.