Developing a short oral health-related quality of life instrument (OHIP) for the edentulous population

Muneera Alshamrany

Dental Sciences Program

Faculty of Dentistry

McGill University,

Montreal

A thesis submitted to McGill University in partial fulfillment of requirements of the degree of Master.

TABLE OF CONTETS:

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude and appreciation to my dearest teacher and wonderful supervisor, Dr. Jocelyne Feine, for taking me through this great journey. At the beginning of this journey, I started as an unskilled student and at the end I became a qualified researcher full of self-confidence. I had good times and bad times during which I had great support and inspiration from Dr. Feine and my wonderful colleagues. As an international student, I would like to express my greatest admiration and thanks to McGill University, the institute, for continuing to provide a unique atmosphere, support, and services.

I also would like to express my thanks to my co-supervisor Dr. David Locker for his valuable comments and thoughtful guidance throughout the research period and beyond. Also, I would like to extend my thanks to Dr. Finbarr Allen, my co-supervisor, for his comments and for teaching me how to organize my thoughts.

My special thanks and indebtedness go to Dr. Manal Awad, for the assistance she provided especially with the statistical analysis; Stephanie Wollin our secretary for making everything easier; and each and every one who helped.

Finally, and most importantly, I would like to dedicate this achievement to my dearest husband, my lovely daughters, Hala and Salma, and my family and thank them all for their continuous support and understanding.

ABSTRACT

Objectives: The aim of this study is to modify the Oral Health Impact Profile (OHIP-49) to develop a shortened version appropriate for edentulous population and evaluate its measurement properties. **Methods:** Data were obtained from a previous randomized clinical trial. Three main phases were conducted: 1) development of a condition-specific form using internal reliability, regression analysis, and item-impact methods, 2) evaluation of its psychometric properties, and 3) further item-reduction with assessment of the psychometric properties. **Results:** The item-impact method yielded the most favorable results. A 22-item form resulted in α = 0.73-0.86 and comparable discriminate validity with the OHIP-49. Subscale change scores were significantly correlated with satisfaction change scores ($P < 0.01$), indicating good construct validity. Effect sizes were high (2.36-1.29). Further item-reduction resulted in lower construct validity and higher responsiveness. **Conclusions:** The 22-item short-form has excellent responsiveness. Further item-reduction would involve a trade-off between construct validity and responsiveness.

RÉSUMÉ ANALYTIQUE

Objectifs : L'objectif de cette étude consiste à modifier le questionnaire OHIP-49 (*Oral Health Impact Profile*) afin de concevoir une version abrégée appropriée pour la population édentée et pour évaluer les propriétés de mesure d'un tel questionnaire. **Méthodologie :** Les données proviennent d'un essai clinique aléatoire antérieur. Trois phases principales ont été menées : 1) l'élaboration d'un questionnaire portant sur des conditions précises en utilisant des méthodes de fidélité interne, d'analyse régressive et d'évaluation de l'incidence sur les unités-questions; 2) l'évaluation de ses propriétés psychométriques; et 3) une réduction plus poussée des unités-questions et une évaluation supplémentaire de ses propriétés psychométriques. **Résultats :** La méthode d'évaluation de l'incidence sur les unités-questions a donné les résultats les plus satisfaisants. Un questionnaire de 22 éléments a donné α = 0,73-0,86 ainsi qu'une validité discriminatoire comparable au questionnaire OHIP-49. Il y avait une corrélation significative entre les scores d'écart de sous-échelles d'évaluation et les scores d'écart de satisfaction ($P< 0,01$) ce qui indique une bonne validité conceptuelle. Les valeurs de l'effet étaient élevées (2,36-1,29). Une réduction plus poussée des unités-questions a donné une validité conceptuelle moins élevée et une réceptivité plus élevée. **Conclusions :** Le questionnaire abrégé comprenant 22 unités-questions a une excellente réceptivité. Une réduction plus poussée des unités-questions déboucherait sur un compromis entre la validité conceptuelle et la réceptivité.

INTRODUCTION

I. GENERAL QUALTY OF LIFE (QoL)

A. Definition of Quality of Life

"Quality of life" (QoL) is an ill-defined term. The first time the term "quality of life" was used was in 1920 by Pigou in his book about economics and welfare (1, 2). Since that time, the term has been applied in different contexts to describe happiness (3), satisfaction (4), living standard (5), climate (6), or environment (7). However, among these various concepts there was no general agreement on the definition of QoL. Most authors who have attempted to define QoL have agreed that QoL is a complex concept to define. The World Health Organization-Quality of Life (WHOQOL) group defined Quality of Life as, "Individuals' perception of their position in life in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards, and concerns. It is a broad ranging concept affected in a complex way by the persons' physical health, psychological state, level of independence, social relationships, personal beliefs, and their relationship to salient feature of the environments" (8). Numerous definitions of QoL were reviewed by Farquhar (9). These definitions range from simple or "focused" to more complex or "global". One example of a focused definition was provided by Mandola and Pelligrini, in which they described QoL as "the individual's achievement of a satisfactory social situation within the limits of perceived physical health" (10). The previously mentioned WHOQOL group definition of QoL is clearly more complex because it encompasses numerous dimensions in life. Farquhar concluded that all definitions fall into three different categories; " 'global' which express

QoL in general terms such as degree of satisfaction with life, 'component ' that break down QoL into specific parts or dimensions, such as health, life satisfaction and psychological well-being; and 'focused' that emphasize only one or two of the range of possible component parts of life."(9).

The absence of clear agreement on a precise definition of QoL is most likely due to the subjective nature of the topic (11). According to O'Connor, individuals may perceive QoL differently based on many concomitant factors such as life events and experiences, personal variables and general psychological mechanisms (12).

Another dilemma one encounters in understanding the concept of QoL is the interchangeable use of QoL with Health Related Quality of Life (HRQOL) (13, 14, 15). This issue will be discussed in the section of HRQOL.

B. Measurement of Quality of Life:

There is a general consensus that QoL encompasses all external and internal aspects of life (16). External aspects are the objective variables such as housing, leisure activities, work, and the environment. Internal aspects are the subjective matters in life such as values, satisfaction, and preferences. Developing a tool or an instrument, i.e. a method of measurement, which can measure all of these aspects in a reliable and valid fashion in different populations, is nearly impossible (16). The many obstacles in measuring the concept of QoL are primarily due to difficulty in moving from the conceptualization process toward the operationalization process through a process that can link QoL to one or more specific, concrete indicators or operational definitions. Particularly, it has been suggested that there is a lack of definitive criteria for quality of life and acceptable

weighting systems for the incorporation of subjective and objective indicators (17, 18). To address this, some investigators have conducted national opinion surveys asking respondents open questions to prioritize important elements of life (9, 19). Generally, aspects prioritized were relationships with families and relatives and health, followed by material elements and activities such as finances and social and leisure activities. Although there is no measure that can assess all of these elements, 'element-specific' measurements have been developed in huge numbers. For example, there are numerous measures in the literature that specifically assess the effect of the social aspect of life on QoL.

Over the past three decades, the need to evaluate and improve services and intervention programs and to allocate resources properly has inspired the development of a large number of instruments to measure QoL. A variety of terms have been used to label these tools; such as scale, index, rating, assessment, and questionnaire but the goal is the same.

II. HEALTH RELATED QUALITY OF LIFE (HRQOL):

A. Concepts and Definitions:

It appears that defining Health Related Quality of Life (HRQOL) is no less confusing than defining QoL. Many authors have pointed to the overlapping meanings of the terms 'QoL,' 'health status,' and 'patients outcome' to explain the cause of this confusion (13, 20, 21). Fitzpatrick et al addressed this problem clearly by saying that using the term 'QoL' "...misleadingly suggests an abstract or philosophical set of judgments or issues relating to life in the broadest sense of factors outside the person,

such as living standard, political or physical environment". Because of this, the authors avoid using the term 'QoL' in health settings (14). On the other hand, some authors who acknowledge this problem prefer to use both terms, 'QoL' and 'health status', interchangeably and argue that, "...there are advantages and disadvantages in considering QoL as an outcome in clinical medicine and health care." (15).

Another fundamental source of confusion is the ambiguity of the 'health' concept. Many authors agree that the complex nature of health itself is the reason for the debate on how to best measure it (15, 22, 23, 24). Moreover, defining health is a subjective matter that is perceived differently by researchers, clinicians, health planners and patients. Nevertheless, defining health is crucial as the concept of HRQOL rests on the concept of health in addition to QoL. Some authors are in favor of the WHO model (25) of health (26, 27, 28), while others disapprove its use as a guide to health domains in HRQOL research (15). In fact, some authors consider that an "imprecise" health concept is a desirable approach because it allows more flexibility (29). In a general sense, despite the inconsistency of accepted health domains, four emerge more often in the literature (27). These domains are: physical function, psychological status, social interaction and somatic sensation.

Since it is critical to precisely define HRQOL in scientific research, QoL from a health prospective has been defined as the social, emotional, and physical well-being of patients following treatment (30) and the impact of disease and treatment on disability and daily functioning (31).

Currently, researchers are calling for more attention on "building a conceptual model or theory as a foundation for the construct of HRQOL" (1). This would contribute to a better understanding of its components (32, 33, 34, 35).

B. Measurement of HRQOL (Evolution of HRQOL instruments):

HRQOL research has rapidly expanded. The number of research articles found following a MEDLINE search for "quality of life" jumped from 40 articles between 1966 6and 1974 to over ten thousand between 1986 and 1994 (37). This rapid expansion of the literature reflects a huge growth in the number of new HRQOL measures. The Quality of Life Instrument Database QOLID, developed by MAPI Research Institute, lists over 1000 QoL instruments (37). It has been suggested that producing this large set of instruments is a remarkable success in respect to the vague concepts and terminology of QoL (14). Instruments vary in their objectives that range widely from assessing general versus specific areas and targeting general versus specific populations. Their mode of administration could be self-administered, interviewer-administered or telephone-administered.

One well-known QoL instrument is the Medical Outcomes Study (MOS) 36-Item Short Form (SF 36) which was designed to satisfy minimum psychometric standards necessary for group comparisons involving generic health concepts that is, concepts that are not specific to any age, disease, or treatment group (38). The Sickness Impact Profile (SIP) is another popular questionnaire that provides a descriptive profile of changes in a person's behavior due to sickness (39). As obvious from the objectives of the SF 36 and the SIP, these apply specifically to health more than any other domain, but there are many other questionnaires such as WHO Quality of Life Assessment (WHOQOL &

WHOQOL-BREF) that measure an individual' perceptions of the general quality of their lives (40). Most of the most popular questionnaires, such as the SF 36, the SIP and the WHOQOL & WHOQOL-BREF, have been translated into several languages.

Reasons behind the increase in HRQOL research:

Pharmaceutical and device manufacturers are facing substantial competition in the market. Therefore, they collect HRQOL data in studies of new drug applications (NDAs) to gain consumers and large purchasers and to win the approval of regulatory agencies (41). Consequently, regulatory agencies, such as the US Food and Drug Administration (FDA), have instituted guidelines for the design and validation of HRQOL instruments (42, 43). In addition, attention to patients' involvement in decision making has increased, especially when more than one therapy is available. Manufacturers strive to provide more precise information on how their treatment can positively impact quality of life (44). As a result, the interest in both utility measures and HRQOL research has increased (45, 46). Utility measures assess a patient's preferences and are particularly useful in evaluating alternative therapies, given the patient's current health status (47). At the level of clinical practice, various physicians wish to track changes in their own patients' health status and to understand the impact of illnesses and therapies on QoL. In response to these demands, researchers have developed a large number of HRQOL instruments and continue to improve and to modify them.

Evolution:

The evolution of HRQOL measures is closely linked with human nature. Humans themselves evolved from primitive to more complex beings, and so have their perceptions and responses to the surrounding environments. Over the last 150 years, a

human understanding of health has developed in phases (15). Health was originally perceived as survival, but then widened to include freedom from disease and then to the ability to perform daily activities. Today, health has an even broader meaning and includes, according to the World Health Organization (WHO), happiness, social and emotional wellbeing, and QoL. Thus, HRQOL measures have changed as the definition of health has broadened.

HRQOL studies have evolved from simple attempts to assess health into more sophisticated studies that incorporate advances in information technology within HRQOL research. The late 1960s and early 1970s marked the start of HRQOL research, while the 1980s and 1990s have witnessed the methodological and analytical development of HRQOL research (1, 41).

Earlier, researchers adopted theories from other sciences, such as psychology and economy, to form the theoretical basis for HRQOL measurement. Psychometrics is a science that scales subjective judgments in numbers. HRQOL instruments, which are based on this theory, measure individuals' understanding, feelings, attitudes, symptoms and capabilities through multiple choice questions (47). Advances in econometric theory have also benefited HRQOL measures through the field of decision theory (48) and utility instruments (14). In decision theory, utility tools measure the patients' willingness to take on risks in order to change their HRQOL (47). Approaches to utility measures, depending on the type of risk a patient is welling to take, are: 1) time tradeoff, in which a patient is asked how many years he or she would trade to regain health, and 2) willingness to pay, in which a patient is asked how much income he or she would pay to gain a given health outcome (15). Both psychometrics and econometrics provide

complementary understanding of HRQOL methodology (49). Finally, the field of clinicometrics, that has been recently developed (50, 51), contributes to HRQOL research through assessing the clinical challenges of health status scale development (52).

Classification of existing HRQOL instruments:

There are varying classifications of HRQOL instruments throughout the literature. According to Fitpatrick et al, HRQOL instruments are classified into generic, disease-specific, site or region-specific, dimension-specific, summary items, individualized, and utility instruments (14).

a. Generic instruments :

Generic instruments are designed to measure several health domains and are suitable across different patient populations. The SF 36 and the SIP are two examples of generic instruments (47). The main advantage of the generic instruments is the wide applicability to all health problems and all subjects, which makes them suitable for comparative purposes (15). On the other hand, these instruments are lengthy and less sensitive for use in a disease-specific context (14).

b. Disease-specific instruments :

This type of HRQOL instruments focuses on specific health conditions and applies to certain patient populations. They have greater responsiveness (53) than the generic instruments. However, because disease-specific instruments are confined to a particular disease, they are not suitable for comparisons across outcomes in different diseases (12). There are a large

number of disease specific instruments available; one example is the Asthma Quality of Life Questionnaire (54).

c. Region- specific instruments:

These instruments are specific to one site or region of the body. The common use of the site-specific instruments is intervention outcomes assessment. Being highly relevant to one site of the body, they are very sensitive to any changes in that site but may fail to identify problems related to other sites of the body which could affect the particular site of concern (14). Titles of region-specific instruments usually indicate the region they apply to, for example the Oral Health Impact Profile (OHIP-49) (55) applies to the oral region.

d. Dimension- specific instruments:

Those instruments that measure one health dimension such as pain (56) or psychological well-being (57) are classified as dimension-specific. They provide detailed information about a specific health aspect but they could be a burden on patients already answering questionnaires regarding other health dimensions (14).

e. Summary / transition item instruments:

Also known as global single/ transition item, these are instruments that contain a single or very few questions. While summery item questionnaires summarize several health issues in brief answer(s), transition items questionnaires focus on comparing health status at two different points in time. Both summary and transition item instruments are very brief and simple,

but this could also account for not being too crude to pick up slight but important changes detected by more detailed measures (58). Further, they lack the ability to detect opposing trends in different aspects of health such as improvement in one health aspect while deteriorating in another (14).

f. Individualized instruments:

Individualized measures allow individual patients to select, weight, and rate the most important aspects of their own lives (59). Thus, high relevance and validity are the main advantages. However, individualized measures are less convenient than self-completed questionnaires, because they require more time from both the interviewer and the patient. The concept of individualized measures is recent but is attracting a considerable interest (14). One example of this type of measure is the McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) where a patient identifies up to five activities that have been affected negatively and then ranks them in term of priority (60).

g. Utility instruments:

As discussed before, these are preference-based measures that have been developed from economics. They measure preference by asking the patient to choose among a series of "alternative scenarios" of health status while accepting risks in order to achieve it (47). Therefore, they evaluate alternative therapies and express it as numerical values. However, most utility measures are less feasible due to intensive labor and time required for their application (14).

Guyatt described another useful classification based on the potential applications of health status measures (61). Discriminative measures are designed to distinguish between individuals or groups on a specific variable; predictive measures use a predetermined set of measurement categories to classify individuals into two or more groups and therefore, act as predictors of a condition. Intelligence tests are discriminative indexes because they discriminate between learning abilities. A diagnostic, or screening, measure such as Health Opinion Survey, a 20-item questionnaire, is an example of a predictor of mortality (62). Evaluative indexes are used to measure, in a longitudinal fashion, the magnitude of change in an individual or group on a specific variable. A classic recent example of those measures is disease-specific quality of life questionnaires for quantification of treatment benefit in clinical trials.

III. ORAL HEALTH RELATED QUALITY OF LIFE (HRQOL):

A. Historical background:

Oral Health Related Quality of Life (HRQOL) is relatively new but rapidly growing phenomenon. This phenomenon started and continued to materialize over the last two decades (63). Many authors explored the history of HRQOL and tried to explain the circumstances that have led to its prominence (63, 64, 65). Slade (64) and others (65, 66) have identified the shift in the perception of health from merely the absence of disease and infirmity to complete physical, mental, and social well-being, the definition by WHO, as the key issue in the conception of HRQOL and, subsequently, OHRQOL. This shift took place in the second half of the twentieth century and it was the product of a "silent revolution" in the values of highly industrialized societies from materialistic

values that concentrate on economic stability and security to self-determination and self-actualization values (67, 68). For example, maintaining physically healthy teeth and gingival (gum) would be the only dental care concerns of a patient with materialistic values, whereas a patient with as with post-materialistic concerns these values might broaden to include esthetic concerns and impact of appearance on self-esteem and interaction with others (66). It is self evident from literature that the notion of OHRQOL appeared only in the early 80's on the contrary to general HRQOL notion that started to emerge in the late 60's, almost two decades earlier than OHRQOL. One explanation for the hindrance in the HRQOL development could be the poor perception of oral diseases' impact on social life. Only forty years ago, researchers have rejected the fact that oral diseases could be related to general health (69, 70, 71). Davis has pointed out that, apart from pain and life threatening cancers, oral disease does not have any impact on social life and it is only linked with cosmetic issues (69). Likewise, others have argued that dental disease was considered as one of the frequent complaints such as headache, rash, and burns that were perceived as unimportant problems (70) and were rarely looked at as the classic "sick role" and further should not be an excuse for exemption from work (71). Later, since the late 70's, the foundation of OHRQOL concept started to shape up as more evidence grew of the impact of oral disease on social roles (72, 73, 74, 75), defying previous statements that have denied this impact.

Clearly, clinical indicators of oral diseases such as dental caries or periodontal diseases were not exclusively suitable to capture the new concept of health declared by the WHO, particularly the mental and social well-being. This has created a demand for new health status measures, in contrast to clinical measures of disease status. As a result,

researchers started to develop alternative measures that would evaluate physical, psychological, and social impact of oral conditions on an individual. These alternative measures are in form of standardized questionnaires.

B. Definition of Oral Health Related Quality of Life (OHRQOL):

Not surprisingly, the term "Oral Health-related Quality of Life" has no strict definition. However, there is a general agreement that it is a multidimensional concept (63). The available definitions varied from simple to more rigorous. An example of a simple definition is the one provided by the Surgeons General's Report on oral health which defines the OHRQOL as "a multidimensional construct that reflects (among other things) people's comfort when eating, sleeping, and engaging in social interaction; their self – esteem; and their satisfaction with respect to their oral health" (76). On the other hand, rigorous definitions are mostly the product of research designed to conceptualize oral health and OHRQOL by taking the OHRQOL construct and refining it by giving it a conceptual or theoretical definition. These types of definitions are more operational since it is feasible to link the definition to one or more specific, concrete indicators (77). Furthermore, these definitions are important as a first fundamental step towards developing OHRQOL measures.

Usually, existing conceptual models of health and HRQOL were used to construct new models specific to OHRQOL. In 1995, Gift and Atchison (78) developed a multidimensional concept of OHRQOL based on the structure of HRQOL model provided by Patrick and Erickson (79). According to that model, OHRQOL incorporates survival (absence of oral cancer, presence of teeth); absence of impairment, disease or

symptoms; appropriate physical functioning associated with chewing, swallowing and absence of discomfort and pain; emotional functioning associated with smiling; social functioning associated with normal roles; perceptions of excellent oral health; satisfaction of oral health; and absence of social or cultural disadvantage due to oral status (78). Similarly, Locker has developed a model of oral health earlier in 1988 in which he described consequences of disease (80). For example, disease can lead to impairment which may lead to functional limitation and/ or disability and finally handicap as a last consequence. Disability is more likely to occur when both discomfort and functional limitation have existed, and handicap is more probable if all three have happened (80).

Generally, all existing OHRQOL have a lot in common. As indicated by Gift *et al*, concepts of oral health and oral health – related behaviors reported in the literature were consistent form mid 60s through the early 90s (81). For instant, they said that different surveys in 1964, 1970s, 1980s, and 1990s have shown that the absence of perceived need was consistently the major cause of not going to the dentist.

C. Importance of Oral Health Related Quality of Life (OHRQOL):

The concept of OHRQOL is significant to mainly three areas; these are the clinical practice of dentistry, dental research, and dental education (65). Oral health-related quality of life has an obvious role in clinical dentistry which translates into the clinician recognition that they do not treat teeth and gums, but human beings. Besides, oral-related behaviors such as practicing good oral hygiene, having regular check up, and spending more money on esthetic dental care are motivated by OHRQOL concerns.

The notion of OHRQOL is tremendously important to all level of dental research. Successful research, whether basic sciences research, clinical studies, or community research makes a contribution to patient's QoL. On community research level, the concept of OHRQOL is especially vital to promote oral health care and access to care. For example, a clinical indicator such as decayed, missing, and filled teeth (DMFT) is not a suitable tool for advocacy on political level because it was designed mainly to quantify the magnitude of the disease (dental caries) but not the impact of that magnitude on individual's daily live and general health which can be better appreciated by dentist than politician. In contrast, politicians may appreciate the impact of dental caries when high DMFT scores are interpreted into impaired QoL due to inability to eat, sleep, or concentrate because of the associated pain, for instance. In this sense, OHRQOL is a better tool to communicate with policy-makers and negotiate access to care. Likewise, the same approach is more rational to educate individuals about their oral health. People are more likely to behave positively when they understand how oral diseases affect their general health and QoL opposed to the effect on their teeth or gums. Researchers have realized the importance of OHRQOL and started and continued to generate measurement instruments.

D. Measurement of Oral Health Related Quality of Life (OHRQOL):

Fundamentally, there are three categories of OHRQOL measure, as indicated by Slade (64). These are social indicators, global self ratings of OHRQOL, and multiple items questionnaires of OHRQOL. Briefly, social indicators are used to assess the effect of oral conditions on the communal level. Typically large population surveys are carried

out to express the burden of oral diseases on the whole population by means of social indicators such as days of restricted activities, work loss, and school absence due to oral conditions. While social indicators are meaningful to policy-makers, they have limitations in assessing OHRQOL. For example, using work days loss due to dental problem to measure the impact of oral diseases is not an appropriate indicator for those who are not working.

Global self-ratings of OHRQOL, also known as single-item ratings, refer to asking individuals a general question about their oral health. Response option to this global question can be in a categorical or Visual Analogue Scale (VAS) format. For example, a global question posing the question: "how do you rate your oral health today?" can have categorical responses ranging from "Excellent" to "Poor" or VAS responses on a 100 mm scale.

Multiple items questionnaire are the most widely used method to assess OHRQOL. Researchers have developed QOL instruments specific to oral health and the number continued to grow rapidly to comply with the demand of more specific measures. In addition, these measures can be classified into generic instruments that measure oral health overall versus specific instruments. The latter, can be specialized in measuring specific oral health dimension such as dental anxiety (82), to specific to a disease or a condition such as head and neck cancer (83) and dentofacial deformity (84), or to specific population such as children (85). Also, OHRQOL instruments vary widely in term of in number of questions (items), and format of questions and responses. Eleven OHRQOL instruments that have been well tested to assess their psychometric properties such as reliability, validity, and responsiveness were presented in the First International

Conference on Measuring Oral Health (86). One of these instruments is the Oral Health Impact Profile (OHIP-49).

E. Oral Health Impact Profile (OHIP-49):

In 1994, Slade developed and tested the Oral Health Impact Profile to assess the social impact of oral conditions (55). Particularly, it measures self-reported dysfunction, discomfort, and disability attributed to oral conditions. It contains seven conceptual subscales; each one includes seven questions, thus it is known as OHIP-49 as it has a total of 49 questions. The seven subscales are functional limitations, physical pain, psychological discomfort, physical disability, psychological disability, social disability, and handicap. During the development phase of the OHIP-49, items were generated by a patient group of 18 – 78 years old; therefore, is targets adult population. Questions, items, of the OHIP 49 ask if the problem had been experienced and responses were presented in a 5-point Likert scale format with 4 = fairly often to 0 = never. A subscale score can be obtained by multiplying the code of each question by the relevant weight and then summing it. The overall OHIP-49 score is the sum of the seven subscale scores. Therefore, the higher scores indicate poorer health status.

Amongst other OHRQOL instruments, the OHIP-49 is considered the most popular, well known measure by a large number of researchers globally. This fact is clearly evident from the latest update of the MAPI research institute, in August 2004, which confirms that the OHIP-49 has been translated into ten languages (87). It has been used in a verity of research purposes such as discrimination between different groups (88), evaluation of OHRQOL of individuals (89), and to predict satisfaction with treatment

(90). Furthermore, many randomized clinical trials employed the OHIP-49 to measure OHRQOL as a primary outcome (91, 92, 93).

In spite of broad applicability of the OHIP-49, there are some limitations to it is use. The OHIP-49 is generic instrument that measures the impact of oral conditions in general (87). As any generic instrument, it allows quantifying a dysfunction for individuals experiencing several disease conditions (55, 94). Thus, the items of the OHIP-49 have been structured intentionally to be widely appropriate (55); which can be viewed as an advantage especially if no disease or condition-specific instrument exists in a particular field. On the opposite side, generic instruments can be rather lengthy and more burden on patients completing a battery of measures in the same time. Another important issue is the potential loss of relevance of the questionnaire when applied in a disease specific context (95). For example, if the research question is specific to the dentate population, then edentate individuals would find questions about teeth such as ""have you noticed a tooth which does not look right?" irrelevant, and therefore would not respond to these questions. Consequently, difficulties can be raised during data analysis due to missing answers (96). Besides its length, a particular important limitation of the OHIP-49 as a generic instrument is in clinical trials as it would have fewer relevant items to the particular disease and intervention. Thus, sensitivity to change that could take place because of treatment might be less (14). These limitations have led to the development of three short forms with different purposes.

Existing Short- form OHIPs:

These limitations have led to the development of few short-form OHIPs. Currently, there are three short-form OHIPs, each developed for different purposes.

- *The OHIP 14:*

This short version was developed and tested, by Slade, to measure people's perceptions of the impact of oral conditions on their well-being (95). The OHIP14 provides an alternative shorter form that is more appropriate for research settings which do not permit use of the longer OHIP 49. As the name indicates, the OHIP 14 has 14 items with good psychometric properties.

- *The impact short-form OHIP:*

Locker and Allen have derived a subset of 14 items using a different shortening method than the one Slade has developed (94). Their aim was to develop and evaluate a short-form to measure changes in patient well-being.

- *The OHIP-EDENT:*

Again, Locker and Allen have developed and tested another short version, the OHIP-EDENT, of the OHIP-49 (97). This short-form is rather a condition-specific that intends to assess the social impact of edentulism on individuals and to be used in clinical studies of prosthodontic procedures.

Edentulism is a scientific term for tooth loss; it can be total or partial depending on the number of teeth lost. Despite improvements in oral health prevention, the prevalence of this condition in the older population is still substantial (98, 99, 100). The WHO estimated the percentage of people aged 65 years and older who lost all their natural teeth (totally edentulous) in Canada as 58% in 1993 (100). In spite of the declining prevalence of edentulism in the older population, the need for dental prostheses will still be great due

to the expanding elderly population. In Canada, it is estimated that elderly population will grow 36.5% by the year 2015 (101).

Individuals who lost their teeth have difficulties chewing food and talking, and are unhappy about their appearance. A strong association is found between edentulsim and a decrease in physical, psychological, and social functioning (102, 103, 104) and therefore, edentulsim is considered to be a handicap that may negatively affect individuals' quality of life (80).

Because individuals wearing complete dentures have problems with stability, great attention has been given to an improved intervention in form of implant supported dentures. In 2002, the research team at McGill University in Montreal, Canada issued a consensus statement supporting this treatment modality (105). The impact of implant supported dentures on OHRQOL is being studied, and these investigations would benefit significantly from edentulsim-specific measures (97).

IV. METHODS OF SHORTENING EXISTING HRQOL QUESTIONNAIRES:

The shortening process lacks a standard methodology. Accordingly, a large number of shortening approaches and strategies differ from one study to another. Coste *et al* have investigated a methodology of shortening Composite Measurement Scales (CMS) and concluded that inadequate conceptualization of the shortening process and inappropriate use and excess credit given to statistical techniques used for item selection have led to neglecting important methodological and statistical considerations in the shortening procedures (106). Those approaches were identified as statistical-based, expert-based and

combined approaches. Briefly, the statistical-based approach was defined as selecting or removing items based only on their statistical performance (e.g factor analysis, correlations, Cronbach's alpha, and others). The expert-based approach only was defined as judgment of redundancy or insufficient face validity of items based on authors' or other experts' opinion. The combined approach uses both previous methods to select or remove items. Moreover, Coste *et al* found that methods of evaluating the measurement properties of the short-forms were often inappropriate. Therefore, they recommended guidelines for shortening existing scales and evaluating the short-forms (106).

Other authors have disagreed with Coste's classification of shortening methods stating that "...to characterize one approach as relying on intuitive judgment and the other as empirically based is an oversimplification" (107). Furthermore, they argued that when using factor analysis, identified as a statistical-based approach by Coste *et al*, "a number of subjective decisions" must be made by investigators throughout the shortening process, such as deciding which final factor structure is satisfactory. Ultimately, however, it was agreed that combining both statistical, or "empirical", and expert, or "intuitive", methodologies is superior to either individually.

There are many methods used or proposed to shorten a questionnaire such as internal reliability, regression, factor analysis, item-impact, and others (106). Choosing among these methods should be made very cautiously because each method, when applied to shorten a questionnaire, can produce a quite different subset of items with different measurement properties (94, 107). For instance, The OHIP 14 was produced using the least square regression analysis and it was found to have a good discriminatory ability but with low responsiveness (108) and a floor effect, meaning that more items

with the most probability of "never" were selected (94). This indicates that the distribution of those items' responses is skewed. In response, the impact short-form OHIP was developed using another shortening method believed to include items to enhance a low floor effect. The latter form, also, had a better ability to measure change, i.e. greater responsiveness than the former, but it had inferior discriminatory ability (94). The same method, item-impact, was used to develop the OHIP-EDENT and was found to have good responsiveness (97). Subsequently, it is the only suitable instrument available to evaluate outcomes of interventions such as implant-retained prostheses in edentulous patients (109). However, developing and testing other alternative short forms for this purpose has not yet been explored. Given that applying different shortening methods could produce different subset of questions, other methods can be used, beside the item-impact, to develop different short forms that can be further compared with one another. For example, Slade has used internal reliability, regression analysis, and factor analysis methods to derive the OHIP 14 and concluded that regression short form had the most desirable psychometric properties (95). Similar concept can be applied to derive alternative short forms appropriate to evaluate intervention outcomes in an edentulous population. Nevertheless, many authors do not prefer factor analysis in the construction or shortening of HRQOL instrument because it was found to exclude items that are important to patients (110, 111). For that reason, internal reliability, regression analysis, and item-impact methods were used in this study to develop short forms appropriate to an edentulous population and that has the ability to measure change.

OBJECTIVES:

The objectives of this study are:

1. To use internal reliability, regression analysis, and item-impact methods to develop a short form of the OHIP-49 questionnaire appropriate for an edentulous population that has the ability to measure change.
2. To carry out a preliminary assessment of the psychometric properties (internal consistency, discriminative validity, construct validity and responsiveness) of this short questionnaire using existing data.

METHODS:

Subjects and data:

Data for this study were obtained from a previous randomized controlled clinical trial (93, 112). The trial was conducted at the Montreal General Hospital and Hôpital Notre Dame, in Montreal, Canada. In the trial, 102 participants were randomly assigned to two groups, one group (N=54) received a mandibular overdenture supported by two implants and a bar attachment, and the other group (N=48) received conventional dentures. Participants were adults, ages 35 to 65 years, who were edentulous for a minimum of 10 years, wearing complete dentures regularly, and wishing to replace their existing dentures. Fifty one percent of all participants were males and 49 % were females with an over all mean age of 50.3 years. The aims of the trial were to compare the effect of both treatments on OHRQOL, using the French version of OHIP 49, and the relative efficacy of mandibular implant-supported overdenture, measured by functional assessment (comfort, stability, ability to chew, speech, esthetic, and cleaning ability) and satisfaction ratings on 100-mm Visual Analog Scales (VAS). Responses of each OHIP question, within the previously described seven subscales, were presented on a 6 – point Likert scale ranging from "never" to "all of the time" with the following codes: 1= never, 2= rarely, 3= occasionally, 4= often, 5= very often, 6= all of the time. Therefore, a higher score indicated a poorer OHRQL. Participants completed the OHIP-49, functional assessment, and the general satisfaction ratings before and at 2 months after treatment. Both groups reported less than 50 % satisfaction and had difficulties with chewing prior to receiving treatment. Other information on socidemographic characteristics (age, sex, marital status, education, and income) were also obtained at baseline.

This study consisted of three main phases; 1) development of more than one short form of the OHIP-49 using different approaches, 2) evaluation of the psychometric properties of each short form, and 3) further item reduction and assessment of the psychometric properties. Phase one was carried out in two steps: initial item reduction followed by applying shortening techniques individually. In phase two, internal consistency reliability, discriminant validity, construct validity, and responsiveness were evaluated and compared between each short form. Phase three involved more item reduction, then selected psychometric properties (construct validity and responsiveness) were evaluated. Pre-treatment OHIP scores were used for the development of the short-form OHIPs (phase I) and the evaluation of both internal consistency reliability and discriminant validity. Both pre- and post-treatment scores and satisfaction scores were used to assess the construct validity and the responsiveness of the short-form OHIPs. Statistical analyses were carried out using the Statistical Analysis System software program (SAS institute) (113).

PHASE I. DEVELOPMENT OF THE SHORT FORMS:

Step1. Initial item reduction:

This step included intuitive and empirical approaches. First, items that are related to teeth only were eliminated. This step is necessary to ensure the removal of the irrelevant items using an intuitive approach. Then, item-item correlations were calculated and examined. If two items have a high inter-item correlation, ($r = 0.7$), these items are highly correlated and may measure the same construct. Consequently, one of them is considered to be redundant and can be deleted at this stage. The decision of which item should be

removed was based on experts' judgment. In this situation, two researchers (Manal Awad and Muneera AlShamrany) in the field of OHRQOL were consulted.

Step2. Applying the shortening techniques:

Three shortening methods were individually used to further reduce the number of items retained in step 1. These methods were internal reliability, regression analysis, and item-impact.

- **The Internal Reliability Method:**

Reliability is an important measurement property. It is defined as the extent to which a measure provides the same results on different trials, presuming the characteristics being measured do not change (114). It refers to reproducibility or consistency of scoring. There are different types of reliability; one of them is the internal reliability or internal consistency. A scale with good internal reliability means that it has homogeneous items that can assess different aspects of the same attribute (115). For example, a scale of physical function reflects physical performance but not emotional function. Internal reliability can be assessed by means of different statistical tests; the most commonly applied test is Cronbech's alpha coefficient (α) that evaluates correlations between individual items and a subscale or total score (item-total correlation). Typically, Cronbach's alpha is used in two situations (115): First, for item reduction, whether during the development of a new scale or when shortening an existing scale. The second situation is when testing the internal consistency of a newly developed scale.

In this study, item – subscale correlations were calculated for the OHIP 49 scores using Cronbach's coefficient alpha to find out what happens to the alpha of a subscale when an individual item is removed. If item removal decreased the alpha of the subscale

markedly (e.g., from 0.9 to 0.6), then that item was considered to be important and was kept in the short form. In contrast, if item removal resulted in a higher alpha for the subscale then that item was deleted from the short from.

- **Least Squares Regressing Methods:**

The common practice in the literature is to select items that make the greatest contribution to the R^2 statistic, which measures the fraction of variance in subscale score that is explained by least-squares regression on the total scale (116). Different regression analyses such as stepwise, forward, or backward regression can be used to shorten a scale. Moreover, a more controlled approach can be used with the regression analysis where only a predetermined number of items are permitted to enter the model according to their contribution to R^2 (95).

In the present study, different regression analyses (stepwise, backward, forward and adjusted R square) were carried out with the subscale score as the dependent variable and each item in that subscale as the independent variable.

- **Item-Impact Method:**

This method emerged a decade ago (107). Nevertheless, it is consistent with clinical sensibility theory described by Feinstein in 1987 as " a theory that uses a sensible method to assess symptoms based on their prevalence in those with a disorder (clinical coherence) and the importance of those symptoms for clinicians to define severity (weighting of symptoms)" (117). Many authors prefer this method to shorten a scale over the statistical methods because it selects items important to patients (118). Typically, item-impact scores are derived by multiplying item prevalence by their severity. Item prevalence is simply the proportion of patients experiencing the item. To obtain item severity (or mean

importance rating), patients are asked to rate the importance of each item using a Likert-type scale, usually a 5-point scale, ranging from "not important" to "very important", then the mean importance rating is calculated for each item.

Since this study was carried out on an existing data base and importance ratings were not collected, a modified method was used to derive item-impact scores. This modified method has been described previously (94, 97). First, a threshold of code 3 =occasionally, code 4 = some of the time, code 5 = most of the time, or code 6 = all of the time was used to dichotomize responses in order to identify those patients who had experienced at least some impact. Therefore, those patients who responded to code 1 = Never, or 2 = Rarely were combined in one group identified as "no impact" and those who responded to code 3, 4, 5, or 6 were combined in another group identified as "with impact". Second, item prevalence was determined for each item by calculating the proportion of patients in the "with impact" group. Third, the mean of the importance rating of each item was calculated by summing the response codes for subjects responded "occasionally" through "all of the time" and dividing this sum by the number of patients with those responses. Finally, the item-impact score was derived by multiplying the above two values then multiplying by the item weight (item's prevalence x item's mean importance rating x item's weight). Allison *et al* have developed Canadian weights for OHIP-49 from English and French speaking samples in Ontario and Québec, respectively (119). Given that the trials participants completed the French version of OHIP-49 and were from Montreal; Québec, the item weights generated for Québec were used in our study. To select items for the short-form, items were ranked within their corresponding

subscale according to the highest impact score, and the top scoring items in each scale were selected to maximize content validity.

How many items to be selected per subscale?

It has been recommended that a minimum of three items per subscale should be included (120) with a similar proportion of items in each subscale (121). Furthermore, others have suggested that the deletion of a significant proportion of items, more than 50 %, of the original scale may have a strong effect on the measurement properties of the short-form (106). Therefore, the top three items with the highest values in all of the methods (reliability, regression, and item – impact) were to be selected. Subsequently, every short form had 21 items, three items per subscale, to improve content validity (108). However, the number of items in the final short – form OHIPs was considered after phase III, when the effect of further shortening of each form on the quality of the measurement properties (construct validity, and responsiveness) was studied and compared with those properties of the 21 items short-form. The intention was to keep the items that will yield better psychometric properties.

PHASE II. EVALUATION OF THE PSYCHOMETRIC PROPERTIES OF THE SHORT FORMS:

Making modifications in a scale, in the form of item deletion, could affect the measurement properties. In some cases, the short forms possess rather different psychometric properties than those of the original scale. For that reason, an evaluation of measurement properties of the short forms should be performed after the item reduction phase (106). Principally, these measurement properties are reliability (internal

consistency and test-retest), validity (content, construct, and criterion), and responsiveness. At this stage, we evaluated internal consistency reliability, discriminant (criterion) validity, construct validity and responsiveness of the short-form OHIPs created in phase I.

1. **Internal reliability:**

Item-subscale correlations were calculated using Cronbach's alpha coefficient to test internal reliability of the short-form OHIP. An alpha of 0.9 is considered high; however, a value of 0.7 and higher also indicates good reliability.

2. **Discriminant validity:**

Discriminative validity denotes the ability of a measure to discriminate between different groups in term of some variables. In the Awad *et al* trial, pre-treatment scores were significantly (p= 0.0084) associated with gender, indicating that OHIP-49 had the ability to discriminate between males and females. Using the same data, the significance of the association between gender and each subscale score and the total score, as well, were used to assess discriminative validity of each short-form OHIP.

3. **Construct validity:**

Construct validity is the most "abstract" type of validity. It relates to the theory about the measure; i.e. test assumptions that a scale measures a specific construct and that the scale performs according to theoretical expectations (117). Typically, it is not possible to test construct validity directly. Instead, it is assessed by means of observing its relationships with other measures that it is hypothetically related to, in order to confirm that the direction and magnitude of the relationships are as hypothesized (114).

There is evidence in the literature that those patients who are satisfied with treatment they have received would report fewer oral health related problems (109, 112, 122, 123). Therefore, we used pre- and post-treatment OHIP scores and satisfaction ratings from the Awad *et al* trial to test the following hypothesis: there is a significant negative association between the change score of each subscale (of the short-form OHIPs) and the change score of the satisfaction ratings. Construct validity of the short-forms is confirmed if the magnitude and direction of the correlation is as predicted. Pearson's correlation was used to measure this correlation.

4. Responsiveness:

Responsiveness, or sensitivity, is defined as the ability of a scale to detect a change as a result of a treatment (114). This was assessed using the satisfaction ratings, as the anchor measure, to calculate effect sizes from pre- and post-treatment OHIP scores in the Awad *et al* trial. Effect sizes were only calculated for patients who reported a minimum of 20mm increase in their post-treatment satisfaction ratings. In order to obtain effect sizes, change scores (post-treatment scores subtracted form pre-treatment scores) were divided by the standard deviation of the pre-treatment scores. Cohen has described guidelines for effect size interpretation as the following: 0.2 is considered small, 0.5 is moderate, and 0.8 is large (124). For example, 0.2 means that there is a change of 20 % (1 SD) of the baseline score.

PHASE III: FURTHER ITEM REDUCTION AND ITS IMPACT ON SELECTED MEASUREMENT PROPERTIES:

Further item reduction was performed in 2 steps. First, the lowest ranking items within each subscale were eliminated to produce a 14 item short-form with two items per subscale, and then selected measurement properties (construct validity, and responsiveness) were evaluated. Subsequently, the 14 item short-form was further shortened by the removal of the lower ranking items, out of the remaining two items, within each subscale. This last step generated a 7 item short-form OHIP with one item per subscale. Finally, the same measurement properties were tested. The aim of this phase is to assess the impact of further shortening on the measurement properties.

RESULTS:

Table 1 shows the baseline characteristics of study participants in Awad *et al* according to the treatment assignment.

Phase I: developing the short forms:

Step1: initial item reduction:

Rates of each code response from "All the time" through "Never" for every item are displayed in table 2. Items 3, 13, 14, and 27 were related to teeth only. Item 3 was the question: have you noticed a tooth which does not look right?, item 13 was the question: have you had sensitive teeth, for example due to hot or cold food or drinks?, item 14 was the question: have you had toothache?, and item 27 was the question: have you been unable to brush your teeth because of problems with you teeth, mouth or dentures?. Therefore, these items were removed. Furthermore, item-item correlations (inter-item correlation) were calculated for the rest of the items and showed that items 1 and 16, 10 and 17, 16 and 28, and 21 and 23 were highly correlated, above 0.7, with each other. The expert panel viewed these items for redundancy and decided on removing item 10, 16, and 21. Therefore, the numbers of items removed per subscale, starting from the first to the seventh subscale, were: 1, 4, 1, 1, 0, 0, and 0 items. As a result, 42 items remained after completing the initial reduction step.

Step 2: Applying the shortening techniques:

1. Internal reliability method:

Generally, internal reliability (item-subscale correlation) was very high among items within the seven subscales and, in most of the time the values of alpha were close to the second decimal place by the deletion of individual items. For example, alphas for

items in the psychological disability subscale ranged from 0.8997 (item 35) to 0.9162 (item 33) with subscale alpha of 0.992. Ultimately, this method retained most of the 42 items. Therefore, we did not consider it as an appropriate method to develop a short-form OHIP.

2. least square regression analyses:

All random selection regression models (stepwise, forward, and backward) we examined retained approximately all of the 42 items. Therefore, the amount of contribution of each item to the correlation coefficient, R^2 was almost equal. When default stepwise procedures were conducted the sequential R^2 values of items were close to the third, and some times forth, decimal place. The shortest possible model had 39 items. So, this method was not considered to be appropriate in the development of short-form OHIP.

3. Item-impact method:

Table 3 displays item prevalence, severity, item weight, and impact score for each of the 42 items. Item-impact scores ranged from 5.54 (item1) to 0.10 (item 45) indicating a wide range of item impacts. Most of the highest impact scores came for the first three subscales. Conversely, the last two scales exhibited the lowest item impact scores. Therefore, to maximize content validity, the top three scoring items from each subscale were selected to produce a 21 item short-form OHIP. Table 4 lists the question contents of the impact short-form OHIP.

Phase II. Psychometric properties:

Internal consistency reliabilities ranged from 0.73 (social disability and handicap) to 0.86 (psychological disability). Table 5 shows the discriminant validity of the original

and short-form OHIP. Total score difference mean of the impact short-form was significantly associated with gender, also scores of the functional limitation, physical disability and psychological disability subscale had a significant association with gender ($p< 0.05$). All subscale change scores (first column, table 6) of the short-form OHIP were significantly and positively correlated with the satisfaction change scores ($P< 0.01$) with correlation coefficients ranging from 0.82 (functional limitation subscale) to 0.50 (psychological disability subscale). Effect sizes of the subscales containing 3 items (first column, table 7) ranged from 2.36 (functional limitation) to 1.29 (handicap).

Phase III. Further item reduction and its impact on psychometric properties:

Tables 6 and 7 illustrate the results of phase III. Generally, there was a decrease in the correlation between the change score of each subscale (of the short-form OHIPs) and the change score of the satisfaction ratings as we continued the shortening process, indicating a decrease in the construct validity. On the contrary, the effect sizes showed an increasing trend except for the functional limitation subscale where effect sizes declined as the shortening process continued.

Table1[†]: Sociodemographic Characteristics of Study Subjects According to Treatment Allocation

Variable	Implant Group (N = 54) (%)	Conventional group (N = 48) (%)
Gender		
Males	26 (48)	26 (54)*
Females	28 (52)	22 (46)
Age		
≥ 50 years	33 (61	29 (60)*
< 50 years	21 (39)	19 (40)
Marital status		
Single	10 (19)	6 (13)*
Married	28 (52)	26 (54)
Divorced	11 (20)	12 (23)
Widowed	5 (9)	4 (10)
College degree		
Completed	32 (59)	34 (70)*
Incomplete	22 (41)	14 (30)
Income		
< 20,000	5 (9)	4 (8)*
20,000 – 40,000	21 (39)	19 (40)
41,000 – 60,000	15 (28)	19 (40)
> 60,000	13 (24)	6 (12)
Preferences		
Neutral	15 (28)	9 (19)*
Conventional	10 (19)	6 (14)
Implant	29 (53)	33 (68)

[†] Source: Awad et al (92)
* Not significant.

Table 2: Distribution of each response code of each item in the OHIP 49

Conceptual domain and item	All the time (N) (%)	Most of the time (N) (%)	Some of the time (N) (%)	Occasio-nally (N) (%)	Rarely (N) (%)	Never (N) (%)
Functional limitation						
Q1 Difficulty chewing	8 (8)	28 (27)	23 (22)	18 (18)	16 (16)	9 (9)
Q2 Trouble pronouncing words	0 (0)	7 (7)	13 (13)	28 (27)	21 (20)	33 (32)
Q3 Noticed tooth that doesn't look right	0 (0)	2 (2)	0 (0)	0 (0)	0 (0)	100 (98)
Q4 Appearance affected	2 (2)	12 (12)	9 (9)	34 (33)	20 (20)	25 (24)
Q5 Breath stale	1 (1)	5 (5)	10 (10)	15 (15)	33 (32)	38 (37)
Q6 Taste worse	2 (2)	10 (10)	9 (9)	27 (26)	28 (27)	26 (26)
Q7 Food catching	14 (14)	31 (30)	22 (22)	25 (25)	8 (8)	2 (2)
Q8 Digestion worse	2 (2)	7 (7)	10 (10)	20 (20)	23 (22)	40 (39)
Q9 Dentures not fitting	14 (14)	35 (34)	18 (17)	12 (12)	17 (17)	6 (6)
Physical pain						
Q10 Painful aching	2 (2)	16 (16)	12 (12)	21 (21)	23 (22)	28 (27)
Q11 Sore jaw	2 (2)	4 (4)	8 (8)	12 (12)	25 (24)	51 (50)
Q12 Headaches	1 (1)	3 (3)	2 (2)	8 (8)	12 (12)	76 (74)
Q13 Sensitive teeth	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	101 (99)
Q14 Toothache	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	101 (99)
Q15 Painful gums	5 (5)	22 (22)	6 (6)	30 (29)	24 (24)	15 (14)
Q16 Uncomfortable to eat	18 (18)	24 (23)	18 (18)	24 (23)	14 (14)	4 (4)
Q17 Sore spots	5 (5)	17 (17)	12 (12)	29 (28)	26 (25)	13 (13)
Q18 Discomfort (dentures)	10 (10)	22 (21)	27 (26)	16 (16)	19 (19)	8 (8)
Psychological discomfort						
Q19 Worried	0 (0)	6 (6)	6 (6)	25 (25)	26 (25)	39 (38)
Q20 Self - conscious	2 (2)	12 (12)	14 (14)	24 (23)	22 (22)	28 (27)
Q21 Miserable	2 (2)	9 (9)	14 (14)	20 (20)	24 (23)	33 (32)
Q22 Appearance	3 (3)	11 (11)	10 (10)	17 (17)	28 (27)	33 (32)
Q23 Taste	2 (2)	4 (4)	10 (10)	19 (19)	21 (20)	46 (45)
Physical disability						
Q24 Speech unclear	1 (2)	7 (7)	11 (11)	23 (22)	23 (22)	37 (36)
Q25 Others misunderstood	0 (0)	4 (4)	6 (6)	20 (20)	26 (25)	46 (45)
Q26 Less flavor in food	0 (0)	7 (7)	10 (10)	25 (25)	33 (32)	27 (26)
Q27 Unable to brush teeth	0 (0)	1 (1)	2 (2)	1 (1)	1 (1)	97 (95)
Q28 Avoid eating	9 (9)	20 (20)	17 (17)	17 (17)	26 (25)	13 (12)
Q29 Diet unsatisfactory	2 (2)	9 (9)	7 (7)	18 (18)	26 (25)	40 (39)

Table 2: Cont'd

Q30 Unable to eat (denture)	1 (1)	6 (6)	8 (8)	18 (18)	21 (20)	48 (47)
Q31 Avoid smiling	1 (1)	3 (3)	1 (1)	11 (11)	30 (29)	56 (55)
Q32 Interrupted meals	1 (1)	10 (10)	6 (6)	23 (23)	27 (26)	35 (34)
Psychological disability						
Q33 Sleep interrupted	0 (0)	2 (2)	0 (0)	4 (4)	19 (19)	77 (75)
Q34 Upset	1 (1)	2 (2)	2 (2)	16 (16)	26 (25)	55 (54)
Q35 Difficult to relax	0 (0)	4 (4)	3 (3)	11 (11)	18 (18)	66 (64)
Q36 Depressed	1 (1)	4 (4)	6 (6)	11 (11)	15 (15)	65 (63)
Q37 Concentration	1 (1)	1 (1)	3 (4)	10 (10)	22 (22)	65 (63)
Q38 Been embarrassed	0 (0)	8 (8)	6 (6)	30 (29)	27 (27)	31 (30)
Social disability						
Q39 Avoid going out	1 (1)	0 (0)	0 (0)	2 (2)	14 (14)	85 (83)
Q40 Less tolerant of others	0 (0)	0 (0)	1 (1)	11 (11)	14 (14)	76 (74)
Q41 Trouble getting on with others	0 (0)	0 (0)	1 (1)	2 (2)	17 (17)	82 (80)
Q42 Irritable with others	0 (0)	0 (0)	2 (2)	7 (7)	14 (14)	79 (77)
Q43 Difficulty doing jobs	0 (0)	1 (1)	1 (1)	5 (5)	18 (18)	77 (75)
Handicap						
Q44 Health worsened	1 (1)	0 (0)	2 (2)	6 (6)	14 (14)	79 (77)
Q45 Financial loss	0 (0)	0 (0)	0 (0)	4 (4)	3 (7)	95 (93)
Q46 unable to enjoy people's company	0 (0)	1 (1)	0 (0)	6 (6)	20 (20)	75 (73)
Q47 Life unsatisfying	0 (0)	4 (4)	5 (5)	15 (15)	26 (25)	52 (51)
Q48 Unable to function	0 (0)	1 (1)	1 (1)	9 (9)	10 (10)	81 (79)
Q49 unable to work	0 (0)	1 (1)	1 (1)	6 (6)	10 (10)	84 (82)

Table 3: Severity, prevalence, weight, and impact score of each item in the short form OHIP

Conceptual domain	Severity[±]: item mean (SD)	Prevalence[§]: Number (%)	Item weight	Item impact
Functional limitation				
Q1 Difficulty chewing	4.34 (0.95)	77 (75)	1.69	5.54[*]
Q2 Trouble pronouncing words	3.56 (0.74)	48 (47)	1.04	1.74
Q4 Appearance affected	3.68 (0.93)	57 (56)	0.99	2.04
Q5 Breath stale	3.74 (0.86)	31 (30)	1.04	1.18
Q6 Taste worse	3.73 (0.94)	48 (47)	0.49	0.86
Q7 Food catching	4.37 (1.05)	92 (90)	0.97	3.82[‡]
Q8 Digestion worse	3.77 (0.93)	39 (38)	1.86	2.68
Q9 Dentures not fitting	4.65 (0.95)	79 (77)	1.34	4.83[†]
Physical pain				
Q11 Sore jaw	3.85 (0.97)	26 (25)	1.39	1.36
Q12 Headaches	3.79 (1.05)	14 (14)	1.49	0.77
Q15 Painful gums	4.03 (1.08)	63 (62)	0.89	2.22[*]
Q17 Sore spots	3.97 (1.03)	63 (62)	.0.85	2.09[‡]
Q18 Discomfort (dentures)	4.35 (0.97)	75 (74)	0.69	2.21[†]
Psychological discomfort				
Q19 Worried	3.49 (0.77)	37 (36)	3.12	3.95[*]
Q20 Self-conscious	3.85 (0.92)	52 (51)	0.97	1.90
Q22 Appearance	4.00 (1.00)	41 (40)	1.72	2.77[‡]
Q23 Tense	3.69 (0.90)	35 (34)	2.66	3.37[†]
Physical disability				
Q24 Speech unclear	3.67 (0.85)	42 (41)	1.35	2.04[†]
Q25 Others misunderstood	3.47 (0.73)	30 (29)	1.42	1.45
Q26 Less flavor in food	3.57 (0.77)	42 (41)	0.76	1.12
Q28 Avoid eating	4.33 (1.03)	63 (62)	0.80	2.14[*]
Q29 Diet unsatisfactory	3.86 (0.99)	36 (35)	0.94	1.28
Q30 Unable to eat (denture)	3.70 (0.88)	33 (32)	1.48	1.77
Q31 Avoid smiling	3.63 (1.02)	16 (16)	1.04	0.59
Q32 Interrupted meals	3.73 (0.93)	40 (39)	1.32	1.93[‡]
Psychological disability				
Q33 Sleep interrupted	3.67 (1.03)	6 (6)	2.27	0.49
Q34 Upset	3.43 (0.87)	21 (21)	1.05	0.74
Q35 Difficult to relax	3.61 (0.85)	18 (18)	1.90	1.21[†]
Q36 Depressed	3.77 (0.92)	22 (22)	2.27	1.85[*]
Q37 Concentration	3.53 (0.92)	15 (15)	1.83	0.95
Q38 Been embarrassed	3.50 (0.79)	44 (43)	0.67	1.01[‡]

Table 3: Cont'd

Social disability				
Q39 Avoid going out	4.00 (1.73)	3 (3)	0.98	0.11
Q40 Less tolerant of others	3.08 (0.29)	12 (12)	3.22	1.17*
Q41 Trouble getting on with others	3.33 (0.58)	3 (3)	1.99	0.19
Q42 Irritable with others	3.22 (0.44)	9 (9)	1.86	0.53†
Q43 Difficulty doing jobs	3.43 (0.79)	7 (7)	1.96	0.48‡
Handicap				
Q44 Health worsened	3.56 (1.01)	9 (9)	2.77	0.87†
Q45 Financial loss	3.00 (0)	4 (4)	0.86	0.10
Q46 Unable to enjoy people's company	3.29 (0.76)	7 (7)	1.22	0.28
Q47 Life unsatisfying	3.54 (0.78)	24 (24)	1.99	1.66*
Q48 Unable to function	3.27 (0.65)	11 (11)	1.98	0.70‡
Q49 Unable to work	3.38 (0.74)	8 (8)	1.18	0.31

± Item mean (SD) of the response codes 3 – 6.

§ % of those in the "with impact group"; i.e. those reporting each item occasionally, some of the time, most of the times, or all of the times.

* Highest impact score

† Second highest impact score

‡ Third highest impact score

Table 4: Question content of short form OHIP

Conceptual domain	Item impact short form
Functional limitation	Q1.Have you had **difficulty chewing** any food because of problems with your mouth or denture? Q2.Have you had **food catching** in your denture? Q3.Have you felt that your **dentures have not been fitting** properly?
Physical pain	Q1.Have you had **painful gums**? Q2.Have you had **sore spots** in your mouth? Q3.Have you had **uncomfortable dentures**?
Psychological discomfort	Q1.Have you been **worried** by dental problems? Q2.Have you felt **uncomfortable** about the appearance of your mouth or dentures? Q3.Have you felt **tense** because of problems with your mouth or dentures?
Physical disability	Q1.Has you **speech been unclear** because of problems with your mouth or dentures? Q2.Have you had to **avoid eating** some food because of problems with your mouth or dentures? Q3.Have you had to **interrupt meals** because of problems with your mouth or dentures?
Psychological disability	Q1.Have you found it **difficult to relax** because of problems with your mouth or dentures? Q2.Have you felt **depressed** because of problems with your mouth or dentures? Q3.Have you been a bit **embarrassed** because of problems with your mouth or dentures?
Social disability	Q1.Have you been **less tolerant** of your spouse or family because of problems with your mouth or dentures? Q2.Have you been a bit **irritable with other** people because of problems with your mouth or dentures? Q3.Have you had **difficulty doing your usual jobs** because of problems with your mouth or dentures?
Handicap	Q1.Have you felt that **your general health has worsened** because of problems with your mouth or dentures? Q2.Have you felt that **life in general was less satisfying** because of problems with your mouth or dentures? Q3.Have you been totally **unable to function** because of problems with your mouth or dentures?

Table 5: Discriminant validity: association between gender and each subscale score and the total score of OHIP-49 and impact-short form

Conceptual domain	OHIP-49		Impact short-form	
	Mean*	CI	Mean*	CI
Functional limitation score	-4.21†	(-7.14, -1.29)	-1.98†	(-3.46, -0.51)
Physical pain score	-3.61†	(-6.48, -0.74)	-1.48	(-3.09, 0.13)
Psychological discomfort score	-2.70†	(-4.93, -0.48)	-0.73	(-1.83, 0.36)
Physical disability score	-3.51†	(-6.28, -0.73)	-2.25†	(-3.45, -1.04)
Psychological disability score	-1.98‡	(-4.04, 0.08)	-1.05†	(-2.02, -0.08)
Social disability score	-0.75	(-1.75, 0.26)	0.12	(-0.29, 0.53)
Handicap score	-1.09 †	(-2.32, 0.14)	-0.46	(-1.07, 0.16)
Total score	-17.85†	(-31.04, -4.66)	-7.30†	(-12.68, -1.91)

*Mean of score difference (mean male scores – female mean score).
† Significant ($p < 0.05$)
‡ Borderline significance

Table 6: Effect of further shortening on construct validity: Correlation* between the change score of each subscale of the short – form OHIP and the change score of the satisfaction ratings.

Conceptual domain	3 items per subscale	2 items per subscale	1 items per subscale
Functional limitation	0.82	0.81	0.74
Physical pain	0.77	0.78	0.65
Psychological discomfort	0.68	0.64	0.61
Physical disability	0.75	0.72	0.71
Psychological disability	0.61	0.53	0.51
Social disability	0.50	0.50	0.40
Handicap	0.54	0.53	0.39

*all correlations were significant ($p < 0.01$)

Table 7: Effect of further shortening on responsiveness: Effect sizes calculated after each item reduction step.

Conceptual domain	3 items per subscale	2 items per subscale	1 items per subscale
Functional limitation	2.36	2.10	2.00
Physical pain	1.75	1.90	1.85
Psychological discomfort	1.50	1.53	1.92
Physical disability	1.75	1.96	2.3
Psychological disability	1.30	1.26	1.70
Social disability	1.70	2.30	2.10
Handicap	1.29	1.49	1.78

DISCUSSION & CONCLUSIONS:

The main aim of this study was to derive a short form OHIP for the edentulous population using three different methods. The second aim was to test the measurement properties of the short forms. A short form OHIP specific for the edentulous population was devised containing 22 items with good measurement properties.

In this study, internal reliability and regression methods did not perform well in reducing items. This could be explained by the moderate to high item-subscale correlation. In the internal reliability method, a subset of items that improve the value of Cronbach's alpha is chosen. The original OHIP-49 had very high α values to begin with and could not be improved significantly when this method was used (55). These findings were consistent with what Slade has reported with the OHIP-49, where all 39 items remaining after the removal of items with low response rate, were retained because of moderate to high α values ($\alpha = 0.94$ for item-total correlation, $\alpha = 0.66 - 0.89$ for item-subscale correlation) (95). Slade used item-total correlation, while we used item-subscale correlation, to calculate α values, but this slight difference in the approach should not make a large difference. Item-total correlation reflects homogeneity of the scale overall, while item-subscale is a way to assess the homogeneity of each subscale and also to assess how each item behaves within each scale. It is important to stress that the controlled entry methods that was used by Slade is not expected to perform in the same fashion for the same reason mentioned above.

However, although regression analysis has yielded satisfactory results with Slade's short form, the OHIP 14, it did not perform as well in our study. This could be attributed to differences in the heterogeneity of the two populations. In Slade's study, the

population was more heterogeneous than in this study because it was a cross-sectional sample of older adults with wide range of dental conditions and, more importantly, it included edentate and dentate participants. On the contrary, our sample was more homogeneous because it included only edentate participants. This meant that there is less variation in participants' responses to items than expected with a more heterogeneous sample. The objective of this method is to select items with the highest contribution to R2. However, in this study almost all items demonstrated equal contributions. Therefore, to further develop a short form using this method; it would have been necessary to intuitively choose among items with similar or very close R^2 values. For example, choosing three from four items with the same R^2 value, 0.988, in the physical pain subscale meant that we needed to make a decision that is not based on the method itself. In Slade study, the R^2 range was much wider, 0.004 – 0.56 and, therefore, item selection could be exercised with more confidence.

The item-impact method, on the other hand, does not rely on inter-item or item-total correlation. Instead, items identified most frequently and rated the most important by patients are selected for the short form (118). Hence, some consider it as an intuitive rather than statistical approach because it selects items that are most important to patients (94, 107). This method has been used previously by Locker & Allen to develop two short form OHIP questionnaires with 14 items using data from Canadian and British patients. The two forms were then combined to form the OHIP-EDENT that had 19 items (97). Twelve out of 19 items in the OHIP-EDENT overlapped with items in our short form. More interestingly, 10 out of 14 items, developed from the Canadian data, overlapped with items in this study. Moreover, one of the four items that did not overlap (Q1: painful

aching) was initially deleted in step 1 because it was considered to be a redundant item. The remaining three items were; item 20: self-conscious (psychological discomfort subscale), item 34: being upset (Psychological disability subscale), and item 46: unable to enjoy company (handicap subscale). The major reason for that is the within-population difference in item weights and in relative ranking of the OHIP subscale items. For instance, item 20: self-conscious was ranked 5th in both Québec and Ontario populations, but the Ontario weight was much higher, 1.55, than Québec's, 0.97. However, an almost complete overlap was observed in the functional limitation, physical pain, and physical disability subscales. These findings were expected because the impact of a disease on function and pain is somewhat more direct in nature than a condition's impact on psychological variables that depend on an individual's emotional makeup (13). These differences in the perceptual impact of edentulism on psychological aspects could be attributed to edentulism's higher prevalence in Quebec, 54%, as compared to Ontario and the rest of Canada (125). This could also suggest that there could be underlying cultural differences between the two populations. Moreover, age of participants in our data, 35 – 65 years, is less than that in Locker & Allen (Ontario population) study which was 50 years and older (126). Another possible cause for the differences in the psychological discomfort subscale, as Allison et al have explained, could be due to differences in individual interpretations of French and English words (119).

Although shorter measures are frequently desired, the advantage of the brevity of a questionnaire could be at the cost of the some of the measurement properties such as the internal consistency (114). Cronbach's α, like other internal consistency measures, are based on the number of items included in the measure, as well as the extent of the

correlation among them. Although the original OHIP was shortened by 50 %, the internal consistency of the short form, 0.73 – 0.86, is considered excellent. Cronbach's α should be above 0.7 but not higher than 0.9 as higher α may suggest a high level of item redundancy (127) and possible loss of content validity (114)

The approach we used to assess construct validity is called convergent validity, which means that a measure of related construct should result in the expected correlation, either positive or negative (127). Pervious studies indicate that patients, who are satisfied with the treatment they have received, would report fewer oral health related problems (109, 112, 122, 123). Therefore, we hypothesized that those patients who had positive satisfaction change scores (higher satisfaction ratings after treatment) would have a negative OHIP change score on each subscale (improved OHRQOL after treatment). In this study, the correlation between the change score of each subscale of the short-form OHIP and the change score of the satisfaction ratings was significant indicating that the short form maintained good construct validity. The short form did not have a similar ability as the original OHIP to discriminate between males and females. However, the total score of the short form discriminated between genders in the same direction and significance as the total score of the original OHIP. The social disability subscale scores of both forms, the short and original, were not significantely associated with gender. Out of the remaining 6 subscales; functional limitation, physical disability, and psychological disability subscale scores discriminated between genders in the same pattern as the OHIP-49, while physical discomfort, psychological discomfort, and handicap subscale scores did not. This may suggest that the items that were removed from the latter 3 subscales had better ability to discriminate in context of gender differences. In the context

of this study, discriminate validity refers to the ability of the short form to discriminate between those who received conventional treatment and implant treatment by means of their scores. On the other hand, using treatment variable to evaluate discriminate validity would be most appropriate. However, gender was used in this study because the variable gender was the only one significantly associated with the pre-treatment scores of the OHIP49. Therefore, the use of gender to assess discriminate validity is acceptable.

Content validity was not directly assessed, however, procedures such as selecting three items per subscale were undertaken to improve the content validity. Moreover, content validity was indirectly assessed by testing responsiveness. If the short form has a good responsiveness, it implies that the content validity is acceptable.

Responsiveness can be assessed using two methods. The first one is to compare the scores on a scale before and after expected change of a treatment with known efficacy (128). To determine the significance and the magnitude of the change, paired t-test and effect sizes can be used. The second method is to use anchor measures, e.g. Patient's global transition ratings, and then relate changes in a scale scores to these measures (110). The second method has the advantage over the first in that patients who report change in the anchor measure can be used to measure responsiveness. In this study, we applied the second method, using the satisfaction ratings as the anchor measure, to calculate effect sizes. Furthermore, 20-mm of increase in the post-treatment satisfaction ratings was considered as the minimum of meaningful change.

Two important issues in the methodology of this study were; the deletion of redundant items (in the initial item reduction) and making decisions on the final number of items retained in the short-form. Items that had high item-item correlation (> 0.7) with

one another were considered to be redundant (e.g. item 21: miserable and item 23: tense) and it seemed appropriate to remove those items that measure the same construct initially before applying different shortening methods. A similar approach has been followed by Juniper et al for the construction of the Mini Asthma Quality of life Questionnaire (121). However, deciding which item to be deleted among two highly correlated items was approached differently. While Juniper et al have approached it statistically by eliminating the item with the lowest item-total correlation; we approached it intuitively by selecting the item on the basis of experts' judgment. The latter approach has been frequently used during the item reduction phase to ensure face validity of new instrument construction (127). Thus, using this approach during the shortening process seemed logical.

The second issue was how many items should be retained in the short-form. At first, three items per subscale were selected, according to the highest impact score, to ensure that there is no huge effect on measurement properties (94, 106, 120, 121). Therefore, a short-form with 21 items was produced with good measurement properties. Up to this step, the final number of items, 21 items, is considered satisfactory with only 2 items more than the OHIP-EDENT. With further reduction of items, the concern was how the psychometric properties would be affected? Thus, we assessed the construct validity and responsiveness as the number of items reduced from 21 to 14 and, finally to 7 with equal numbers of items per subscale. For example, there were 3 items per subscale in the first step of item reduction, then 2 items per subscale in the second step, and finally one item that had the highest impact score per subscale in the third step. The purpose of this procedure was to achieve the shortest form that possessed better measurement properties than the alternative short forms. The approach of establishing the minimum number of

items for a responsive and valid scale was described by Moran et al to shorten the Chronic Respiratory Questionnaire (CRQ). They argued that item retaining or deletion should be based on the consequent effect on the measurement properties rather than approaches such as factor analysis or inter-item correlations (129). Further, they specified that the latter approaches are contradictory to the purpose of the evaluative instruments, which is measuring within-person change over time, because these approaches depend on the size of the between-person variance.

The effect of further reduction on construct validity is consistent with what Moran et al reported. Each item measures a specific aspect of subscale construct and, consequently, deleting more items could lead to compromised construct validity. On the other hand, the effect size will increase with further reduction, except for the functional limitation subscale. The overall increase in responsiveness may be attributed to the fact that the item-impact method, as mentioned earlier, the items with the highest impact score are kept, i.e. the most important to the patients, with further reduction. This means that the items with lower impact score may act as a noise. Although the effect size of functional limitation subscale decreased with further item reduction, this reduction was relatively small and the effect sizes are still large.

In this situation, it seems that if we place the 21 item short-form at one end of a continuum and another with 7 items at the other end, better construct validity but worse responsiveness fall at the 21 item end and worse construct validity, but better responsiveness, at the 7 item end. Making the best decision among the three short forms; 21, 14, 7 items, involves a trade off between construct validity and level of responsiveness. However, it should be mentioned that even though effect sizes of the 21

item short-form were less than those of the 14 and 7 item short-forms, they are considered high according to Cohen's guidelines for the interpretation of effect sizes.

The process of deriving a short form and evaluation of the psychometric properties should not be conducted in the same sample. Instead, a new and independent sample should be used to test the measurement properties (106). Given that we used an existing database, we developed and tested the short form on the same sample. Another related limitation, was the failure to assess test-retest reliability, a measure of reliability over time, which has been reported to be affected by such significant reductions (127). In order to assess this type of reliability, a measure is applied at two different points in time during which it is assumed that there is no change. The scores are then compared, and the closer the scores are, the more reliable the measure. Because the original OHIP-49 was not applied at two points of times between which there was no change, it was not possible to the test-retest reliability.

Therefore, we recommend further assessment of the measurement properties of the short form using an independent sample and an appropriate design to allow for testing the test-retest reliability.

The results of this study suggest that the 21 item short-form developed using the item-impact method maintained good measurement properties when used as an evaluative measure for implant therapy in the edentate population. Moreover, the 14 item short-form can also be used for the same purpose. However, one must bear in mind the trade off between lower construct validity with higher responsiveness. It is vital to ensure whether or not the short-form and its measurement properties are appropriate for the purpose of

measuring and the population to which it is applied. Establishing validity should be a continuous process every time a measure is used in different contexts.

REFERENCES:

1.Wood-Dauphinee S. Assessing quality of life in clinical research: from where have we come and where are we going? J Clin Epidemiol 1999; 52:355-63.

2.Pigou AC. The Economics of Welfare. London: MacMillan; 1920

3.Kozma A, Stones MJ. Predictors of happiness.J Gerontol. 1983 Sep; 38(5):626-8.

4.Carp FM. Impact of improved housing on morale and life satisfaction. Gerontologist 1975 Dec; 15(6):511-5.

5.Berelson B. The population problem: conceptions and misconceptions. Anesth Analg 1971 Jul-Aug; 50(4):481-9.

6.Zamfir G, Rugina V, Mihail G. The role of certain socio-sanitary factors in conditioning the quality of man's life. Part One. Sante Publique 1984;27(3):197-211

7.Watts MS. Ecological health and quality of life now and forevermore. Calif Med 1970 Nov; 113(5):55-77.

8.The WHOQoL Group. What quality of life? World Health Organization Quality of Life Assessment.World Health Forum. 1996; 17(4):354-6.

9.Farquhar M. Elderly people's definitions of quality of life. Soc Sci Med 1995 Nov; 41(10):1439-46.

10. Mendol WF, Pelligrini RV. Quality of life and coronary artery bypass surgery patients. Soc Sci Med 1979; 13A:457-61.

11. B. Spilker .Quality of life assessments in Clinical trials. Raven Press. Ltd. New York; 1990.

12. Rod O'Connor. Issues in the measurement of health-related quality of life. Center for Health Program Evaluation Working paper 30. 1993 (cited 2003 June 21), from: URL

http://www.buseco.monash.edu.au/centres/che/pubs/wp30.pdf

13. . Bowling A. Measuring health: A review of quality of life measurement scales, 2nd ed. Buckingham; Philadelphia: Open University Press; 1997.

14. Fitzpatrick R, Davey C, Buxton M J, Jones D R. Evaluating patient-based outcome measures for use in clinical trials. Health Tec Ass 1998; 2(14):i-iv, 1-74.

15. McDowell I, Newell C. Measuring health: A guide to rating scales and questionnaires, 2nd ed. Oxford Uni. Press: New York and Oxford; 1996.

16. Bowling A. Measuring disease: a review of disease-specific quality of life measurement scales, 2nd ed. Buckingham; Philadelphia, PA: Open University Press; 2001.

17. Rogerson RJ. Environmental and health-related quality of life: Conceptual and methodological similarities. Soc Sci Med, 1995; 41:1373-82.

18. Rogerson RJ, Findlay AM, Morris AS. Indicators of quality of life: Some methodological issues. Environment and Planning, 1989; 21:1655-66.

19. Bowling A .What things are important in people's lives? A survey of the public's judgments to inform scales of health-related quality of life. Soc Sci Med, special issue 'Quality of life' 1995; 10:1447-62.

20. Spitzer WO. State of Science 1986: Quality of life and functional status as target variables for research. J Chronic Dis 1987; 45:765-71.

21. Meenan RF, Pincus T. The status of patient status measures. J Rheumatol 1987; 14: 411-14.

22. Dubos R. Mirage of health: Utopias, progress and biological change. New York: Harper.

23. Baumann BO. Diversities in conceptions of health and physical fitness. J Health Hum Beh 1961; 3:39-46.

24. Cox BJ, Blaxter M, Buckle ALJ. The health and life style survey. London: Health Promotion Research Trust; 1987.

25. World Health Organization. The First ten years of the World Health Organization. Geneva: World Health Organization, 1958

26. Goldsmith SB. The status of heath status indicators. Health Ser Rep 1972;87:212-20.

27. Schipper H, Clinch J, Powell V. Quality of life assessments in clinical trials. New York: Raven Press. Ltd.; 1990.

28. Goldsmith SB. A reevaluation of heath status indicators. Health Ser Rep 1973;88:937-41

29. Rosser RM. Recent studies using a global approach to measuring illness. Med Care 1976; 14 (suppl):138-47.

30. Greer S. The psychological dimension in cancer treatment. Soc Sci Med 1984; 18:345-9.

31. Kaplan R. Social support and social health. In Saranson I, Saranson B, ed. Social support theory, research and application. The Hague: Nijhoff; 1985.

32. Partick DL, Bergner M. Measurement of health status in the 1990's. Ann Rev Public Health 1990; 11:165-83.

33. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995; 273:59-65.

34. Hunt SM. The problem with health-related quality of life. Qual Life Res. 1997; 26:205-12.

35. Gill TM, Feinstein AR. A critical appraisal of quality-of-life measures. JAMA.1994; 272:619-26.

36. Albrech GL. Subjective health assessment. In Jenkinson C, ed. Measuring health and medical outcomes. London: UCL; 1994.

37. Mapi Research Institute. The Quality of Life Instruments Database. Available from: URL

http://www.qolid.org/

38. Ware J, Sherbourne CD. The MOS 36-items short for health surveys (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473-83.

39. Gilson BS, Gilson JS, Bergner M, Bobbit RA, Kressel S, Pollard WE, Vesselago M. The Sickness Impact Profile. Development of an outcome measure of health care. Am J Public health 1975; 65: 1304-10.

40. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May; 28(3):551-8.

41. Kaufman S. The emerging role of health-related quality-of-life data in clinical research. Part 1: The increasing importance of quality – of – life research. Clinical researcher 2001; 1(5): 2-6.

42. Lewis C. Grappling with QoL: Patients, FDA and drug companies struggle to link therapies with well – being. FDA Consumer magazine 2001, March – April (cited 2003 August 14) from: URL

http://www.fda.gov/fdac/features/2001/201_life.html

43. Burke LB. Acceptable evidence for pharmaceutical advertising and labeling DIA workshop on pharmacoeconomic and quality o life labeling and marketing claims

2000; Oct 3. (cited 2003 August 14) New Orleans (LA) from: URL http://www.fda.gov/cder/ddmac/brklDIA1000/index.htm

44. Till JE, Sutherland HJ, Meslin EM, Is there a role for performance assessments in research on quality of life in oncology? Qual Life Res 1992; 1:31-40.

45. Epstein AN. The outcomes movement-will it get us where we want to go? N Eng J Med 1990; 323:266-70.

46. Torrance GW, Thomas WH, Sackett DL. 'A utility maximization model for evaluation of health care programs'. Health Services Research 1972; 7:118-33.

47. Kaufman S. The emerging role of health- related quality-of-life data in clinical research. Part 2: Basic concepts and terminology of quality-of-life research. Clinical researcher 2001; 1(6): 38-43.

48. Torrance GW. Measurement of health state utilities for economic appraisal: a review. J Health Econ 1986; 5:1-3.

49. Revicki D, Kaplan R. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. Qual Lie Res.1993; 2:477-87.

50. Feinstein AR. The theory and evaluation on sensibility. In: Clinimetrics. New Haven: Yale University Press; 1987.

51. Wright JG, Feinstein AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. J Clin Epidemiol 1992; 45(11):1201-18.

52. Fava GA. Ethodological and conceptual issues in research on quality of life. Psychother Psychsom 1990; 54:70-6.

53. Deyo RA, Patrick LP. 'Barriers to the use of health status measures in clinical investigation, patient care, and policy research'. Med Care1989; 72 (3 Suppl):S254-68.

54. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health-related quality of life instrument in Asthma: development of a questionnaire for use in clinical trials. Thorax1992; 47:76-83.

55. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Comm dent health 1994; 11:3-11.

56. Cleeland CS. Assessment of pain in cancer. Measurement issues. Advances in Pain Research and Therapy 1990; 16:49-55.

57. Beck A, Ward C, Medelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch General Psychiatry 1961; 4:561-71.

58. Jenkinson C, Peto V, Coulter A. Measuring change over time: a comparison of result from a global single item of health status and multi-dimensional SP-36 health status survey questionnaire in patients presenting with menorrhagia. Qual Life Res 1994; 3:317-21.

59. Ruta D, Garratt A. Health status to quality of life measurement. In: Measuring health and medical outcomes (Jenkinson C, editor). London: University College London Press.

60. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire – an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol1987; 14:446-51.

61. Kirshner B, Guyatt G. A methodo;ogical framework for assessing health indices. J Chron Dis 1985; 38(1): 27-36.

62. Somervell PD, Kaplan BH, Heiss G, Tyroler HA, Kleinbaum DG, Obrist PA. Psychologic distress as a predictor of mortality. Am J Epidemiol. 1989; 130(5):1013-23.

63. US Department of Health and Human Services. Healthy people 2010. US Department of Health and Human Services. Government Printing Office, 2000:8.

64. Slade G.D. Oral Health-related quality of life: Assessment of oral health-related quality of life. In: Inglehart MR, Bagramian RA, ed. Oral health-related quality of life. Illinois: Quintessence Publishing Co. Inc; 2002

65. Gift H.C, Atchison K. A., Dayton C. M. Conceptualizing oral health and oral health related quality of life. Soc Sci Med 1997; 44 (5):601-08.

66. Inglehart MR, Bagramian RA. Oral health related quality of Life: An Introduction. In: Inglehart MR, Bagramian RA, ed. Oral health-related quality of life. Illinois: Quintessence Publishing Co. Inc; 2002

67. Inglehart RF. The silent revolution. New Jersey: Princeton University Press; 1977.

68. Inglehart RF. Cultural shift. New Jersey: Princeton University Press; 1990.

69. Davis P. Complaince Structure and the delivery of health care: The case of dentistry. Soc Sci Med. 1976; 10:329 – 35.

70. Dunnell K, Cartwright. Medicine takers, prescribes and hoarders. London: Rutledge and Kegan; 1972.

71. Gerson LW. Expectations of 'Sick Role' exemptions for dental problems. Can Dent Assoc J 1972; 10:370 – 72.

72. Cohen L, Jago J. Toward the formulation of sociodental indicators. Int J Health Serv 1976; 6: 681 – 87.

73. Bonito A, Iannachione V, Jones S, and Stuart C. A study of dental health related process outcome associated with prepaid dental Care. Final Report, Part I, DHEW Contract No. HRA 231 – 760093, Research Triangle Institute, Research Triangle Park. NC; 1984.

74. Cushing AM, Sheiham A, and Maizels S. Developing socio – dental indicators-the social impact of dental diseases. Comm Dent Health 1986; 3:3 – 17.

75. Ettinger RL, Oral diseases and its effect on the quality of life. Gerodontics 1987; 3: 103 – 106.

76. US Department of Health and Human Services. Oral Health in America: A Report of the Surgeon General. NIH publication 00-4713. Rockville, MD: US Department of Health and Human Services, National Institute of Dental and Craniofacial Research, National Institute of Health, 2000:7.

77. Blalock H. "Measurement and Conceptualization Problems" American Sociological Review 1979; 44: 881-94.

78. Gift H.C., Atchison K.A. Oral health, health, and health-related quality of life. Med Care 1995; 33(11): NS57-NS77.

79. Patrick DL, Erickson P. Health status and health policy. Quality of life in health care evaluation and resource allocation. New York: Oxford University Press; 1993

80. Locker D. Measuring oral health: A conceptual framework. Community Dental Health 1988; 5:3-18.

81. Gift HC, Atchison KA, Dayton CM. Conceptualizing oral health and oral health-related quality of life. Soc Sci Med 1997; 44(5):601-08.

82. McNeil DW, Rainwater AJ 3rd. Development of the Fear of Pain Questionnaire—III. J Behav Med 1998; 21(4):389-410.

83. Terrell JE, Nanavati KA, Esclamado RM, Bishop JK, Bradford CR, Wolf GT. Head and neck cancer-specific quality of life: instrument validation. Arch Otolaryngol Head Neck Surg 1997; 123 (10):1125-32.

84. Cunningham SJ, Garratt AM, Hunt NP. Development of a condition-specific quality of life measure for patients with dentofacial deformity: I. Reliability of the instrument. Community Dent Oral Epidemiol 2000; 28(3):195-201.

85. Jokovic A, Locker D, Stephens M, Kenny D, Tompson B, Guyatt G. Validity and reliability of a questionnaire for measuring child oral-health-related quality of life.. J Dent Res 2002; 81(7):459-63.

86. Slade GD, Strauss RP, Atchison KA, Kressin NR, Locker D, Reisine ST. Conference summary: assessing oral health outcomes--measuring health status and quality of life. Community Dent Health 1998; 15(1):3-7.

87. The Oral Health Impact Profile from: URL http://www.qolid.org/public/OHIP.html

88. Hunt RJ, Slade GD, Strauss RP. Differences between racial groups in the impact of oral disorders among older adults in North Carolina. J Public Health Dent 1995 Fall; 55(4):205-9.

89. Murray H, Locker D, Mock D, Tenenbaum HC. Pain and the quality of life in patients referred to a craniofacial pain unit. J Orofac Pain 1996 Winter;10(4):316-23.

90. McNaugher GA, Benington IC, Freeman R. Assessing expressed need and satisfaction in complete denture wearers. Gerodontology 2001 Jul; 18(1):51-7.

91. Hegarty AM, Hodgson TA, Lewsey JD, Porter SR. Fluticasone propionate spray and betamethasone sodium phosphate mouthrinse: a randomized crossover study for the treatment of symptomatic oral lichen planus. J Am Acad Dermatol 2002 Aug; 47(2):271-9.

92. Allen PF, McMillan AS.A longitudinal study of quality of life outcomes in older adults requesting implant prostheses and complete removable dentures. Clin Oral Implants Res 2003 Apr;14(2):173-9.

93. Awad MA, Locker D, Korner – Bitensky N, Feine JS. Measuring the effect of intra-oral implant rehabilitation on health-related quality of life in a randomized clinical trial. J Dent Res 2000; 79(9):1659-1663.

94. Locker D, Allen F. Developing short-form measures of oral health-related quality of life. J Pub Health Dent 2002; 62(1):13-20.

95. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997 Aug;25(4):284-90.

96. Locker D. Effects of non-response on estimates derived from an oral health survey of older adults. Community Dent Oral Epidemiol. 1993 Apr; 21(2):108-13.

97. Allen F, Locker D. A modified short version of the oral health impact profile for assessing health-related quality of life in edentulous adults. Int J Prosthodont 2002 Sep-Oct; 15(5):446-50.

98. US Department of Health and Human Services. Oral health in America: A report of the Surgeon General. Rockville (MD): US Department of Health and Human

Services, National Institute of Dental and Craniofacial Research, National Institute of Health, 2000.

99. Kelly M, Steele J, Nuttall N, et al. Adult Dental Health Survey: oral health in the United Kingdom in 1998. London: UK Office for National Statistics, 2000.

100. World Health Organization. WHO oral health country/area programme. Malmo University, Sweden: WHO Division of Non-communicable Diseases/ Oral Health, WHO Collabrating Centers; 2003.

101. Thompson GW, Kreisel PS. The impact of demographics on aging and the edentulous condition on dental care services. J Prosthet Dent 1998; 79 (1):56-9.

102. Ship JA, Duffy V, Jones JA, et al. Geriatric oral health and its impact on eating. J Am Geriatr Soc 1996; 44:456-64.

103. Joshipura K, Ritchie C, Douglass C. Strength of evidence linking oral condition and systemic disease. Compend Contin Eduuc Dent Suppl 2000: 12-23.

104. Slade GD, Spencer AJ, Locker D, et al. Variations in the social impact of oral conditions among older adults in south Australia, Ontario, and North Carolina. J Dent Res 1996; 75: 1439-50.

105. Feine JS, Carlsson GE, Awad MA, Chehade A, et al. McGill consensus statement on overdentures. Montreal, Quebec, Canada. May 24-25, 2002. Int J Prosthodont 2002;15(4):413-4.

106. Coste J, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 1997 Mar;50(3):247-52.

107. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. J Clin Epidemiol 1997; 50(3):233-8.

108. Locker D, Jokovic A, Clarke M. Assessing the responsiveness of measures of oral health-related quality of life. Community Dent Oral Epidemiol 2004 Feb; 32(1):10-8.

109. Awad MA, Lund JP, Shapiro SH, Locker D, Klemetti E, Chehade A, Savard A, Feine JS. Oral health status and treatment satisfaction with mandibular implant overdentures and conventional dentures: a randomized clinical trial in a senior population. Int J Prosthodont 2003 Jul-Aug; 16(4):390-6.

110. Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, ed. Quality of life and pharmaeconomics in clinical trials, 2nd Ed. Lippincott – Raven Publishers. Philadelphia; 1996

111. Fayers PM, Hand DJ. Factor analysis, causal indicators and quality of life. Qual Life Res 1997 Mar;6(2):139-50.

112. Awad MA, Lund JP, Dufresne E, Feine JS. Comparing the efficacy of mandibular implant-retained overdentures and conventional dentures among middle-aged edentulous patients: satisfaction and functional assessment. Int J Prosthodont 2003 Mar-Apr;16(2):117-22.

113. Statistical Analysis System (SAS): Version 8, 3rd ed. Cary, NC: SAS Institute; 1996.

114. Streiner D, Norman G. Health measurement scales. A practical guide to their development and use, 3rd ed. New York: Oxford University Press; 2003.

115. Portney LG, Watkins MP. Statistical measures of reliability. In foundation of clinical research: Applications to practice, 2nd ed. New Jersey: Prentice Hall Health; 2000.

116. Moore DS, McCabe GP. Introduction to the practice of statistics, 4th ed. W.H. Freeman and Company, New York; 2003.

117. Feinstein A. Clinimetrics. New Haven: Yale University Press; 1987: 141-166

118. Cook DJ, Guyatt GH, Adachi JD, Epstein RS, Juniper EF, Austin PA, et al. Development and validation of the mini-osteoporosis quality of Life Questionnaire (OQLQ) in Osteoporotic women with back pain due to vertebral fractures. Osteoporosis Quality of Life Study Group. Osteoporosis International 1999; 10(3):207-13.

119. Allison P, Locker D, Jokovic A, Slade G. A cross-cultural study of oral health values. J Dent Res 1999; 78(2): 643-9.

120. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. The Canadian Med Assoc J 1986; 134:889-95.

121. Juniper EF, Guyatt GH, Cox FM, Ferrie PJ, King DR. Development and validation of the Mini Asthma Quality of Life Questionnaire. Eur Respir J 1999; 14:32-8.

122. Boerrigter EM, Geertman ME, Van Oort RP, Bouma J, Raghoebar GM, Van Waas MAJ. Patient satisfaction with implant-retained mandibular overdentures. A comparison with new complete dentures not retained by implants – A multicenter randomized trial. Br J Oral maxillofac surg 1995; 33:282-8.

123. Meijer HJA, Raghoebar GM, van't Hof MA, Geertman ME, van Oort RP. Implant – retained mandibular overdentures compared with complete denture: A 5 – year follow – up study of clinical aspects and patient satisfaction. Clin Oral Implants Res 1999; 10:238-44.

124. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1977.

125. Brodeur JM, Benigeri M, Naccache H, Olivier M, Payette M. Trends in the level of edentulism in Québec between 1980 and 1993. J Can Assoc 1996; 62(2): 156-60, 62-6.

126. Locker D, Salde GD. Oral health and quality of life among older adults: The Oral Health Impact Profile. J Can Dent Assoc 1993; 59:830-44.

127. Nunnally JC. Psychometric theory, 2nd ed. McGrawth – Hill: New York; 1978.

128. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Controlled Clinical Trials 1991; 12 (suppl):142-58.

129. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items of items for a responsive, valid, health – related quality of life instrument. J of Clin Epidemiol 2001; 54:571-579.