SUMMARY MEASURES FOR ASSESSING THE RISK OF A NEW DISEASE: THE CASE OF THE SEXUAL TRANSMISSION OF HIV

by

SILVIA DECLICH

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfilment of the requirements for the
M.Sc. degree in Epidemiology

University of Ottawa

December, 1993

ISBN 0-315-95905-3

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

It is widely recognized that the development of most diseases involves a variety of factors. Moreover, the risk associated with each factor varies according to the presence of others, so that relying on a constellation of risk factors, rather than any single predictor, provides more accurate information. The appreciation of this multifactorial influence should lead to the development of summary measures, which attempt to synthesize the values of numerous variables into a single statement about risk of exposure or risk of becoming infected. The difficult task of choosing among existing instruments, or deciding to develop a new one, becomes even more arduous when we consider the lack of a consolidated body of knowledge concerning methods for risk assessment and summary measure development and the frequent use of an intuitive or an ad hoc approach. In fact, there is a surprising amount of confusion in the literature on methods for developing indexes of risk, and several disciplines have produced different strategies.

The search for an instrument for measuring risk of human immunodeficiency virus (HIV) infection has led to the study of risk assessment methods that have been developed by different health related areas; thus, the general focus of this thesis was modified to study methodology. The case of risk assessment for a new disease, such as HIV infection, is an application of the theoretical work and may be an example for other similar situations.

The overall goal of this work is to critically assess and clarify methodologies for risk assessment and summary measure development which may be useful in appraising the risk involved in sexual transmission of HIV. To achieve the goal of the study, the following objectives were established: a) to identify and describe the construction processes of summary measures used by different disciplines; b) to propose a classification of methods for developing summary measures of risk; c) to assess the strengths and weaknesses of each method; d) to

review the existing summary measures for risk of sexual transmission of HIV; and e) to discuss the adequacy of these existing instruments in assessing the risk of sexual transmission of HIV.

The work provides extensive background information on measurement principles, basic characteristics of summary measures, description of construction process, and assessment of measures' reliability and validity, with particular attention being paid to those aspects that regard summary measure of risk.

The source of information for the available methods for risk assessment which various disciplines have developed, was a systematic review of the literature, the raw material being papers containing indexes assessing risk. Five computerized searches were conducted on four different computer-based bibliographies through an innovative research strategy. Approximately 1,200 titles were retrieved as a result of the computerized searches, of which 170 papers were actually collected. Roughly the same number of articles was identified through references cited in the original material. After the application of three inclusion criteria to the collected material, about 160 papers were retained for review and appraisal. Papers containing information about the development procedure were selected and sorted into reasonably homogeneous groups, which resulted in the proposal of a classification for risk assessment methods.

The proposed classification identified the following methods: 1) the empirical method, which relies on collected data; 2) the judgemental method, which is based on the opinions of experts; 3) the psychometric method, which is built on a theoretical hypothesis about the construct; and 4) the mathematical method, which is based on an abstract representation in mathematical form of the phenomenon under study. Development strategies, procedures for assessing the validity of a new measure, and strengths and weaknesses are discussed for each method. Moreover, the methods are compared according to seven criteria, and the factors that should be considered in the choice of the appropriate method for risk assessment are highlighted.

A strategy similar to the one used in the search for risk assessment methods was employed in the search of the literature for existing indexes for the risk of HIV due to sexual behaviour. Fifty-one articles describing thirty summary measures (plus six alternative forms) were retrieved. Development process, measure characteristics, and information available on reliability and validity are described and discussed. Several trends in the existing indexes are found: first, a growing interest in developing these kinds of instruments; second, the presence of a small number of measures developed for predicting the future risk of acquiring HIV infection; third, a change of interest from exposure to HIV to the actual risk of having or of acquiring the infection; fourth, a transition from measures for homosexuals to instruments for heterosexuals and towards indexes specific for increasingly better defined populations; fifth, recently increased yet scarce attention paid to evaluating the reliability and validity of measures constructed; and sixth, an increasing use of standard methods as opposed to ad hoc approaches for developing indexes.

The proposed classification of methods has proven to be useful in organizing the existing summary measures for risk of HIV. It has also shown that certain methods are preferred to others, depending on the type of measure being developed and that specific factors influence the choice of method. The most important factor for a new disease such as acquired immune deficiency syndrome (AIDS) is the stage of the epidemic, given that the accessible knowledge and the availability of a sufficient number of cases of disease are essential for employing a certain method. A logical progression of methods over time is evident, with the mathematical method as a foundation of risk assessment, followed by the psychometric method supported by a theoretical hypothesis, and finally, the judgemental and empirical methods, which require a sufficient amount of experts' knowledge and empirical data, respectively. The assessment of risk for HIV infection seems to be reaching the final stage of this progression, with an increasing number of summary measures developed through empirical and judgemental

methods; a possible repetition of this sequence can be predicted.

Recommendations for improvements and future development based on the material and results from this study include the need of enhancing awareness with respect to developing methods, their individual advantages and weaknesses, and the factors that influence the appropriateness of a method. Researchers from different disciplines should contribute to the creation of a body of knowledge on risk assessment methods and promote the integration of methods, which could enrich the way in which factors are selected, weighted, and aggregated. Furthermore, it is necessary to standardize the principles, the strategies, and the terminology for summary measures, to produce guidelines for the development of risk indexes, to define criteria for the presentation of measures and the acceptance of articles for publication, and to identify criteria for evaluating risk assessment indexes.

This work is an attempt to organize, describe, and evaluate the available risk assessment methods. A broad perspective was taken to find methods that could be used for risk assessment, to clearly describe the factors influencing the choice of a method, and to organize them in a framework for a better understanding of risk index development and for increasing the number of high quality measures. The appraisal of risk for other multifactorial phenomena, such as the progression from HIV infection to AIDS or events related to other diseases, will improve the efforts aimed at establishing standard methodologies for risk assessment.

Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
APACHE	Acute Physiology And Chronic Health Evaluation
HIV	Human Immunodeficiency Virus
ICU(s)	Intensive Care Unit(s)
IVDU(s)	Intra Venous Drug User(s)
MESH	Medical Subject Headings
STD(s)	Sexually Transmitted Disease(s)
WHO	World Health Organization

ACKNOWLEDGEMENTS

This thesis was realized while I was supported by a fellowship from the Istituto Superiore di Sanita' (Rome, Italy) for on the job training at the Federal Centre for AIDS (FCA), later the HIV/AIDS Division, Bureau of Communicable Disease Epidemiology, Laboratory Centre for Disease Control (LCDC), in Ottawa, Canada. I am grateful to these institutions for providing me with this opportunity.

I would like to express my gratitude to the thesis supervisory committee: Tom Elmslie, my principal supervisor, for supporting me in conceiving, planning, and carrying out this work; Ian McDowell for believing in my abilities and for encouraging me to face new challenges in scientific thinking; and Rama Nair for having followed me from my first term as a special student, to my acceptance into the Master's programme, up to the realization of this work with support, concern, and encouragement.

I am grateful to the professors of the Department of Epidemiology and Community Medicine of the University of Ottawa for their assistance and their patience in understanding my stilted English. A special thanks to Mariella Peca, Anne Plourde, Randa Nooh, Jackie Kierulf, Adwoa Bentsi-Enchill, Margaret Cousins, and the other students and staff of the Department who amicably supported me during times of hardship and life in a foreign country.

Finally, I am indebted to Lucilla Marino and Mark Kanieff for the painstaking revision of the language and style of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

Risk assessment may be defined as "the process of quantifying the relationship between environmental or personal characteristics and the probability of the occurrence of some outcome"[1]. It is widely recognized that the development of most diseases involves a variety of factors. Moreover, the risk associated with each factor varies according to the presence of others, so that relying on a constellation of risk factors, rather than any single predictor, provides more accurate information. For example, the probability of becoming infected with human immunodeficiency virus (HIV) through sexual transmission involves several components: sexual behaviour exposing a person to the risk (e.g., sexual practices, condom use, and type and number of partners), number of persons infected in the population (i.e., prevalence of HIV in the considered population), and other less well-understood factors relating to susceptibility of the host, infectiousness of the infected partner, and infectivity of viral strains. The assessment of the risk involved in the transmission of HIV infection through sexual contact should then consider these factors and choosing just one of these (e.g., number of sexual partners), to distinguish between high and low risk individuals in a population, would not take into account other factors that could increase the risk (e.g., anal intercourse) or decrease it (e.g., consistent use of condoms). The appreciation of this multifactorial influence led to the development of risk assessment instruments, which attempt to synthesize the values of numerous variables into a single statement about risk of exposure or risk of becoming infected. However, choosing among existing instruments, or deciding to develop a new one, requires guidelines for the appraisal of a suitable risk assessment instrument.

Developing a summary measure means to aggregate multiple factors quantitatively into a composite risk estimate. It involves the choice of factors affecting the risk, the creation of an algorithm to combine these factors (accounting for a possible joint effect of multiple

contributors, such as counting, summing, multiplying, and weighting), and the validation of the instrument. This task becomes even more arduous when we consider the lack of a consolidated body of knowledge concerning methods for risk assessment and summary measure development and the frequent use of an intuitive or ad hoc approach. There is a surprising amount of confusion in the literature on methods for developing indexes of risk, and several disciplines have produced different strategies.

The search for an instrument for measuring risk of HIV infection has led to the study of risk assessment methods that have been developed by different health related areas; thus, the general focus of this thesis was modified to study methodology. The case of risk assessment for a new disease, such as HIV infection, is an application of the theoretical work and may be an example for other similar situations.

The current World Health Organization (WHO) estimates of more than one million acquired immune deficiency syndrome (AIDS) cases and of an additional 9 to 11 million individuals infected with the HIV[2,3] are illustrative of the alarming dimension of the AIDS epidemic. The predominant mode of HIV transmission is sexual intercourse, which is responsible for more than 75% of the cases among adults worldwide, while transmission through needle exchange, blood transfusion and blood products account for the remaining 25%[4]. The assessment of risk of infection through sexual contact may help in the efforts for preventing the spread of this disease, for which neither a cure nor a vaccine has yet been found. Risk assessment can be used to describe the distribution of risk levels among the population, to target groups at higher risk for preventive activities, to identify individuals for counselling on behavioural change, and to detect asymptomatic people for available treatment.

The overall goal of this work is to critically assess and clarify methodologies for risk assessment and summary measure development which may be useful in appraising the risk

involved in sexual transmission of HIV. To achieve the goal of the study, the following objectives were established: (a) identify and describe the construction processes of summary measures used by different disciplines; (b) propose a classification of methods for developing summary measures of risk; (c) assess the strengths and weaknesses of each method; (d) review the existing summary measures for risk of sexual transmission of HIV; and (e) discuss the adequacy of existing instruments in assessing the risk of sexual transmission of HIV.

CHAPTER 2 - SUMMARY MEASURES

The general concepts and the construction process involved in developing summary measures of risk are similar to those used for other types of summary measures. Thus, it would be appropriate to begin with a discussion of measurement principles and with a presentation of the basic characteristics of summary measures, as well as the steps involved in their development process and the requirements for assessing their validity and reliability.

"Measurement consists of rules for assigning numbers to objects to represent quantities of attributes. The term 'rules' implies that the procedures for assigning numbers must be explicitly formulated"[5]. When the phenomenon to be measured involves many factors, the measurement process also requires a judgement on the relative importance of the various factors and an algorithm to aggregate the information into a single measure. A summary measure can consequently be defined as a formalized and standard multi-component instrument for global assessment, not biased by subjective variation, where the process of measuring, "weighting and combining the individual factors is made according to defined rules"[6]. The advantage of summary measures over personal judgement is the objectivity and the communicability of the process and the results[5].

The development and use of summary measures have become increasingly common in several health related areas, such as clinical practice, clinical trials, epidemiology, health promotion, psychiatry, psychology, and health service evaluation. Despite this variety of uses, the general principles are similar. The desired characteristics of a summary measure (i.e., type, purpose, operational definition, and application) influence the process of development and the strategies for assessing validity and reliability. Determining the characteristics also assist those who want to judge whether a developed measure is appropriate to their needs. Moreover, assessing the degree of validity and reliability provides information on the quality of the

measure.

Since the use of summary measures is shared by several disciplines, terminology is an issue. Different terms can be found for designating similar things, and each discipline has produced its own jargon. For example, summary measure has been also reported as "index", "score", "scoring, staging, or rating system", "profile", "scale", "inventory", "classification", "criteria", and "test". In several sections of the present work, the reader will be alerted to differences and parallels in the usage of terms.

2.1. CHARACTERISTICS OF A SUMMARY MEASURE

Characteristics include the type of measure (which may range from simple inventories to complicated equations), the designed purpose, the conceptual basis and the operational definition on which the measure is based, and finally, the population to which the measure will be applied. In this first section, the characteristics of a summary measure will be presented, with special attention being paid to those aspects that have more interest for summary measures of risk.

2.1.1. Type of summary measures

A summary measure for a phenomenon that has multifactorial influences can have different forms. They are described below in order of complexity.

a. *Checklist*. A list of factors (or combination of factors) is provided. The factors are just counted when present. For example, the Needle Sharing Inventory (NSI) is a 59-item test developed to specify and assess behaviours associated with drug use and needle sharing, the context in which HIV transmission may occur. It may be used to identify individuals who are at high risk for HIV infection[7]. Alternatively, each factor is divided into levels (e.g., low, medium, and high) and the total number of high level factors is considered. With checklists it

is assumed that each factor acts independently and equally to increase risk.

b. *Classification* (or categorization). Single factors (or a combination of more than one factor) are grouped into classes on the basis of common characteristics. Individuals are assigned to these predesignated classes. The HIV infection-exposure risk index (VIRI) combines six different sexually related variables into 31 possible scenarios. These categories are ordered into five classes of increasing risk level, from no risk to high risk[8].

c. *Score* (or index). A value is assigned to each factor (or to sub-levels of factors). The final score is provided by the combination of these values according to a proposed model (additive, multiplicative, etc.). An example of a scoring system is the Apgar score, a well recognized method for clinical assessment of neonates. Five clinical signs (i.e., heart rate, respiratory rate, colour, muscle tone, and reflex response to nasal catheter) are each rated with three response options (0, 1, and 2); these ratings are then added to form an index which can range from 0 (no sign of life) to 10 (vigorous infant)[9].

d. *Probability equation.* An equation is developed to provide a probability estimate for an individual when his or her specific set of values is used. The Framingham coronary risk profile is a summary measure for calculating the conditional probability of cardiovascular events over a specified number of years[10].

The type of measure will influence the analysis that may later be performed (e.g., the probability estimate provides a continuous measure, while the score often produces a discrete one). While probability estimate can be easily converted into any of the simpler forms, it is very difficult to convert a simpler form to one of higher complexity. For example, converting a scoring system to probability requires the development of a conversion model or at least a table of empirical probabilities[11].

Comparison between risk equations and simpler forms suggests that the greater the simplification the less valid the result[4]. For example, checklists tend to overlook (or to

overstate) persons at high risk because of multiple marginal abnormalities and because most critical values used to dichotomize a continuous variable are themselves arbitrary[4].

2.1.2. Purpose of a summary measure

Purpose refers to the measure's intended use. Bombardier and Tugwell[6] note that clinical indices have different purposes corresponding to levels of medical decision making. An index can be used to make a diagnosis (*diagnostic index*), to arrive at a prognosis (*prognostic index*), or to assess efficacy of the treatment (*therapeutic index*)[6]. Similarly, Feinstein divides clinimetric measures into indexes of status (diagnostic criteria or graded ratings for magnitude or severity), indexes of prediction (prognostic estimates), and indexes of change (describing changes)[12].

The above distinctions may be suitable for clinical measures, but they are too narrow to be generally applied to any summary measure. Three broader categories have been proposed to include the potential applications of health status measures (table 1). These categories, which can be extended to other types of summary measures (including measures of risk), identify three possible purposes: discriminating among subjects, predicting a future condition or outcome, and evaluating changes over time[9,13]. A *discriminative index* is used to distinguish between individuals or groups on an underlying dimension (a continuum of health, illness, or disability); for example, intelligence tests used to distinguish between children's learning abilities, and screening or diagnostic instruments[13]. The discriminative purpose includes the diagnostic indexes. An index may have *predictive purpose* when it is used for predicting an event or a condition occurring after the instrument is applied (e.g., prognostic instruments). The outcome itself or a gold standard (often referred to as the criterion) is applied prospectively to determine whether individuals have been classified correctly[13]. This category is parallel to the prognostic purpose of clinical measures. Lastly, an *evaluative index* will measure the extent of longitudinal

change in an individual or group on the dimension of interest (e.g., quantifying the treatment benefit in clinical trial, comparing the effectiveness of alternate health care programs, or evaluating changes in individual behaviour)[13,14]. This category includes the therapeutic purpose of clinical measures.

This classification is useful for summary measures assessing HIV risk. A risk index may be used to distinguish individuals based on their present risk of HIV infection. In this case, the summary measure is expected to discriminate individuals according to levels of risky behaviours or to screen them for their risk of having the HIV infection (discriminative purpose). Differently, a developer may be interested in predicting the future risk of acquiring HIV infection in a certain period (predictive purpose). The test for detecting anti-HIV antibodies can be used as a gold standard to verify the predictions. Finally, an index could be developed to measure the longitudinal changes in at-risk behaviours which expose a person to HIV infection (evaluative purpose).

Table 1. Purpose of a summary measure

General purpose	Clinical purpose	Description
discriminative	diagnostic	distinguish individuals/groups
predictive	prognostic	predict future state
evaluative	therapeutic	evaluate changes

These purposes are not always mutually exclusive, but the requirements for maximizing one of the functions of discrimination, prediction, or evaluation may sometimes impede the others. As a result, an index developed for one purpose may not be the most appropriate for another[6], and the strategy for assessing the validity of an index for one purpose may differ and does not necessarily ensure that it can be used for the remaining ones[13]. For example, an index intended to predict the risk of hospitalisation will include those factors that best correlate with

the outcome (e.g., age and health status), whereas a measure developed to evaluate changes in the risk of hospitalisation will emphasize those behaviours and lifestyles (e.g., smoking), that, although related to the risk of hospitalisation, may be modified by the individual[15].

2.1.3. Conceptual basis and operational definition of a summary measure

Describing the purpose does not fully define a measure. Even within discriminative or predictive indices, the relevant content will depend on the specific question asked. The context for developing a summary measure is based on the conceptual basis to which the developer refers. This can be a particular theory or a wide body of knowledge, no matter how imperfect[16]. For example, an index of health status may be based on the negative view of health (i.e., absence of disease), on the concept of functioning (e.g., Functional Disability Measurements), or on the positive aspect of health (e.g., Quality of Well-being)[17]. The reference to a conceptual approach is critical for developing, interpreting, and evaluating a summary measure. This is a major problematic issue of several instruments, which lack a clear description of the underlying theoretical basis or an explanation of the derivation of the measure from its conceptual foundation.

Within the framework created by the conceptual basis, the specific object or event to be measured should be defined. This operational definition is particularly important when the object is not readily observable, when the phenomenon is multidimensional, or when the event is the result of many factors. For instance, the Acute Physiology and Chronic Health Evaluation (APACHE) scoring system defines severity as the risk of imminent death among patients hospitalised in intensive care units (ICUs), while the Patient Management Categories (PMCs) relate increasing case complexity with increasing costs[18].

In practice, the operational definition specifies the context and the level of study for the phenomenon, which will influence the selection of factors to be included in the summary

measure. For instance, it should indicate whether the interest lies in actual factors or in their precursors (e.g., demographic, social, cognitive, or environmental determinants), whether the attention is focused on causal factors or indicators of the phenomenon, and whether there is a preference for modifiable factors. Moreover, the operational definition should clarify whether a direct measurable outcome is present (e.g., the quality of care for mild disease cannot use mortality as outcome), which will influence the choice of the method used to develop the measure, and whether a gold standard exists for the phenomenon being measured.

2.1.4. Application of a summary measure

Similar measures may further differ according to the type of people for whom the instrument has been developed or on whom its validity has been tested. Depending on the phenomenon of interest, a measure may be specific to one population or one setting and not applicable to others. In these situations the characteristics of the population for which the measure is intended should be clearly stated (e.g., age group, gender, special diagnosis, and particular setting).

In fact, the factors to be included in a summary measure may vary for different populations. The Johns Hopkins Risk Equation showed that the risk factors for developing coronary heart diseases in young persons are not the same as those identified for adults[19]. Even when the factors are the same, the weights attached to the variables may differ. As an illustration, in the Framingham risk profile for cardiovascular disease some factors are equally important for both men and women, while others are weighed differently because the risk associated is diverse (e.g., women have higher risk due to glucose intolerance, whereas men show higher risk associated with left ventricular hypertrophy)[10]. In the field of HIV infection, for example, measures developed for use among homosexual populations are inappropriate for estimating the risk of HIV infection among heterosexuals, because of differences in risk factors,

in their relative importance, and in the levels of prevalence of HIV among potential partners.

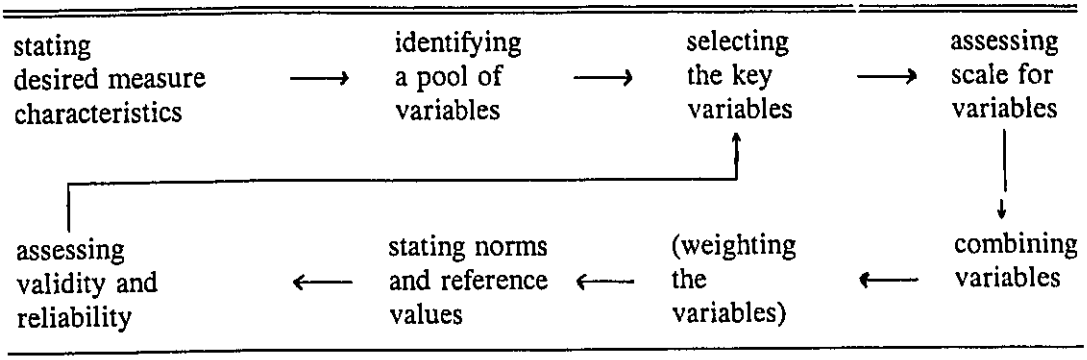
Generic measures, having a broad application across different diseases, may show poor performance in relation to a specific disease group. For example the APACHE II has been developed to quantify severity of illness and to predict patient outcome for people in ICUs. Application to a particular disease requires additional factors specific for that disease[20].

In conclusion, the purpose, together with operational definition and intended application, creates a framework for developing a summary measure and provides a guide for testing its reliability and validity.

2.2. DEVELOPMENT OF A SUMMARY MEASURE

The construction process of a summary measure, where several variables are incorporated, requires a sequence of operations, which roughly includes (table 2): stating the characteristics of the desired measure (type, purpose, operational definition, and application), identifying a pool of potential variables, selecting the appropriate variables, assessing the adequate scale for each variable, combining the variables into a single global measure (if necessary, assigning weights to the variables), stating norms for the standard use of the measure and reference values for interpreting the results, and finally, assessing validity and reliability of the measure. This construction process may be cyclic. The assessment of reliability and validity often shows that the measure needs to be improved; consequently the developer refines the selection of the variables and the subsequent steps.

Table 2. Development process for a summary measure



When decisions underlying each of these steps are made according to the subjective judgement or impressions of the developer, the resulting index cannot be replicated by others and adapted to different contexts, and may suffer from lack of credibility and subjective bias. Moreover, this *ad hoc approach* employs a strategy which makes sense only for the problem at hand, but is not applicable to other problems. The procedure of summary measure development is invented at the same time as the index itself. In this way, one prevents the replication, hence the scientific evaluation, of the results[21]. Although criticized by several authors[12,22,23], an ad hoc approach is frequently utilised in constructing summary measures, creating confusion and loss of scientific value in the field of summary measure development.

Unlike an ad hoc procedure, a standard method establishes a formal process for the development of an index[12,23]. Although it may seem superfluous, as these are requisites of every scientific method, it is important to state that: a) a method needs to be based on a theoretically sound foundation; b) a method must be transparent enough in order to allow other users to modify and to apply it; and c) a method must be replicable by the same researcher or by others and must produce results which are comparable. Unfortunately, there is no organized body of knowledge concerning methods for the development of summary measures. Various disciplines in the area of health sciences have developed different methods, with different strengths and limitations, and researchers are often not aware of all the methods available. This,

together with the absence of established guidelines for the development of summary measures development and with the negligence of certain researchers, may have resulted in the great number of indexes developed in an ad hoc manner.

In this section, some general issues concerning the development of summary measures will be described. These apply to the various methods which will be discussed in detail in the subsequent chapter. The problem of terminology recurs here, again with regard to the components included in an index. For example, the terms "factor" or "dimension" are often used interchangeably to define the elements included in a multi-component measure, whereas "attribute" is utilised by psycho-social scientists to designate a more abstract concept. Factors are assessed through one or more "variables" or "items".

2.2.1. Stating the characteristics of the desired measure and identifying a pool of variables

Before the development process begins, the investigator has to define the characteristics of the desired measure. As stated in the previous section, the purpose, the conceptual bases, the operational definition, and the intended application create the framework for the construction of the appropriate measure. An example using AIDS can be used to illustrate this point. A researcher wishes to develop a new measure to predict the individual risk of HIV infection in the next 5 years among adolescents. The conceptual basis is the theory about the disease to which he refers, such as that AIDS is a deadly illness due to a specific infectious agent, the HIV virus, that arises through transmission of that agent from an infected person to a susceptible host. This is not the only theory about AIDS and other researchers may use the less accepted point of view, proposed by Duesberg, that AIDS is not an infectious disease, but is instead caused by drugs like alcohol, hard drugs, and amphetamines, and not by HIV[24]. Then the developer decides what it is about AIDS that he needs to measure and the level of interest. For example, does he want to assess the general "risk of acquiring the infection" or the more

specific "risk of acquiring the infection through sexual transmission"? Once the object is chosen, he should provide an operational definition of it and explain the reference to its theoretical framework. The dimensions or factors that are considered should be clearly identified at this point (e.g., sexual behaviour, host susceptibility, partner infectiousness, viral strains virulence, prevalence of infection, etc.).

This sequence of logical steps helps the investigator (or a panel of experts) to identify an adequate pool of variables. These should be larger in number than those that will eventually be used and can be selected from variables previously used by other researchers, or created from clinical observation, theory, research findings, and expert opinion[25]. Considering the framework defined, an initial selection should be made ensuring that the measure will have important variables and that it adequately covers the domain under investigation. The criteria for inclusion will depend on the measure's purpose. For example, an index intended to discriminate between infants who require resuscitation at birth and those who do not, will include those variables that best separate the two groups, that are stable over time, and that are universally applicable to all infants at birth. Variables chosen for predictive purposes should be known to be associated with the criterion of interest (e.g., neonatal mortality and morbidity). Finally, a measure to evaluate the efficacy of resuscitation should focus on variables sensitive to physiologic changes and responses to medical intervention in distressed infants[9,13]. The relevance of the variables identified also depends on the population for which the measure is intended. Factors might differ between the general population and a special population[26].

2.2.2. Selecting the key variables

From the initial pool, a sub-group of variables should be chosen and the variables which do not contribute to, or actually detract from, the usefulness of the instrument should be excluded. This is accomplished in different ways, according to the chosen method. These

methods will be described in detail in the next chapter. Briefly, a group of experts may be asked to use their knowledge and judgement to evaluate the coverage and the importance of the variables in the initial pool. In mathematical modelling, the choice of the components is still made through a judgemental strategy, but mainly utilising the developer's subjective opinion. Another strategy relies on the statistical association with an external criterion to indicate the most important variables (although it involves a subjective decision as well, for the choice of the model). Finally, the approach commonly used by psychosocial researchers requires the preliminary measure to be pre-tested on a group of individuals and the utility of each item to be assessed for its psychometric properties, such as the frequency of endorsement (e.g., an item answered with the same response by most of the people is not useful), and the degree of homogeneity. Whichever method is used, the choice of the variables to be included should consider their credibility (face validity) and their comprehensiveness (content validity) in relation to the phenomenon under study.

2.2.3. Assessing scale for the variables

The scale associated with each variable permits its translation into numerical values. Scales for the variables may have been selected beforehand; if not, at this point they are demarcated. The choice depends on the nature of the phenomenon (qualitative or quantitative) and on the desired level of measurement[17,25]. The set of categories that constitutes a scale should be exhaustive (i.e., fully covering the range of possible expression), mutually exclusive, and should have realistic values for the population of interest[12].

Scales are classified according to the precision with which their categories can be ranked. A scale can refer either to a single variable or to the output of a composite index containing multiple variables[12]. Categories of a *nominal* scale have no order and the values that are assigned, for example to female or male, are just labels. In an *ordinal* scale, contiguous

categories can be arranged in increasing or decreasing order, but values are still assigned arbitrarily and the intervals between them cannot be actually measured in fixed and equal units. The Apgar score includes the variable "respiratory rate" with three possible levels: absent, irregular or slow, and good. The values, expressed in an ordinal scale of 0, 1, or 2, do not have a dimensional meaning[9]. An *interval* scale has adjacent categories separated by measurable equal intervals. However, a continuous variable may be divided into classes of frequency to which values may be assigned. Finally, in a *ratio* scale the intervals are constant and there is a true and meaningful zero, so that the ratio of two values can be calculated.

Usually, the developer, or a panel of experts, decides the scale, the available categories, and the numerical values to be assigned according to the requirements of the model chosen to combine the components and the desired analysis that could be done on the final measure[25]. The scale depends also on the measure's purpose. A predictive measure needs item scales which maximize correlation with the criterion measure, while an evaluative measure should ensure response sets with sufficient gradations to register changes[13].

2.2.4. Combining the variables through a model

A distinctive feature of summary measures is the synthesis of multiple variables into a single output expression that offers a global rating for the complex phenomenon under study. The simplest way to synthesize the components is simply to *count* the variables when they are present, which provides a checklist. A more complex form of aggregation is classification, where the variables are cited in categorical levels and the output measure is formed when these elements are combined into classes. For example, the three levels of the TNM (Tumor Nodes Metastases) staging system for cancer are formed logically through the various unions and intersections of the variables using the *Boolean algebra*[12].

In scoring systems and probability equations, numerical values are assigned to the

variables and then joined *mathematically*. Individual factors may be added, multiplied, or aggregated using more complex models. Assumptions are made about the way in which variables relate to the phenomenon under study, and an algorithm is applied in order to derive the final measure. The choice of the model includes considerations about the independence of the variables and the possible interactions among them. The most common mathematical procedure is the sum, although other models, such as multiplicative, binomial, regression models, have been employed. Possible models depend on the methods chosen, as will be described in the next chapter.

2.2.5. Weighting the variables

When the summary measure is obtained through an additive or some other mathematical model, variables may vary with respect to importance and contribute differently to the total score. In this case, the value of each component variable could be multiplied by it's own coefficient, called "weight"[25]. For example, the hospital Mortality Prediction Model (MPM) provides a probability estimate through the equation $-3.0+2.6x_1+1.6x_2+1.5x_3+0.7x_4+0.6x_5 +0.04x_6-0.5x_7+0.0001(x_8)^2$. In this formula the individual x_i values are the available rating for variables (such as consciousness (x_1) that could be rated 0 for no coma and 1 for coma, or age (x_6, that assumes the actual age values), and the coefficients are the relative weights[11,27].

Even with checklists, weights can be assigned to individual variables. The Therapeutic Intervention Scoring System (TISS) classifies patients by illness severity utilising a list of over 70 procedures[28]. These procedures are counted when present, but some of them contribute 1 point to the total score, while others contribute 2, 3 or 4 points. Moreover, different populations may require different weights for the same factors. Deciding the magnitude of the weights can be done using purely judgemental strategies or relying on empirical data with the aid of some statistical models, according to the method chosen to develop the summary measure.

2.2.6. Stating norms and reference values

Regardless of how the development process has been carried out, basic norms must be established for reproducibility of the measure, so that it will yield similar results when repeatedly applied by the same rater over time or by different raters simultaneously. Clear and objective decision rules for each step involved in deriving the final score will reduce the measurement errors due to the raters' variability when they apply the instrument, and it will provide an identical score for the same case[29]. First, indications should be stated to identify each of the variables and the associated scale. For example, if level of consciousness is a component variable in the MPM index of severity, the possible categories have to be stated and defined - a) consciousness: neither coma nor deep stupor; b) unconsciousness: presence of coma (unresponsiveness to painful stimulation) or deep stupor (minimal response to pain with decorticate or decerebrate posturing) - as well as the numerical value attached (0 or 1)[27]. Moreover, the developer should state the mechanisms of aggregation of the basic components to form the summary measure, and the output categories that emerge[12].

Finally, the developer should supply information to help with interpretation of the results. For some measures, cut-off points may define meaningful levels for the final score. Another possibility is to provide reference scores or average probabilities from the population for which the measure is designed in order to furnish values for comparing individual scores[17]. For instance, the estimated probability of coronary heart disease within 6 years, assessed for a 40 year-old man, may be compared with the average risk for Framingham participants at the same age[10].

2.3. QUALITY OF A SUMMARY MEASURE

Once the development process is finished, and before the measure can be used, it is necessary to validate the instrument. In fact, any measure, in order to be useful, must lead to

consistent results on repeated measurements (reliability) and fulfil its intended purpose (validity). The assessment of reliability and validity provides information on the quality of the summary measure and may lead to a refinement of the instrument through a reiteration of the development process with revision of some of the steps.

2.3.1. Reliability of a summary measure

"Reliability refers to the degree to which the results obtained by measurement procedure can be replicated"[30], regardless of how they should be interpreted. It is sometimes referred to as reproducibility, or repeatability. A highly reliable measure is one that leads to consistent results on repeated occasions because it does not fluctuate greatly due to random error[31]. Lack of reliability may arise from divergences between observers or instability of the attribute being measured.

There are several ways to examine the reproducibility of a measure administered by different raters or on different occasions. *Inter-observer reliability* refers to the agreement between the results obtained by different raters using the same measure. It is generally assessed from correlations between different raters' judgements of a sample of respondents. The intra-class correlation coefficient (ICC) for continuous data[32] and Cohen's Kappa (K coefficient) for categorical data[33] are appropriate measures. Typical ICC values fall in the range of 0.65-0.95, and values above 0.85 may be considered satisfactory[17], whereas a K coefficient greater than 0.75 represents excellent agreement and values below 0.40 show poor agreement[34]. *Intra-observer reliability* refers to the reproducibility of two measurements made by the same observer. It evaluates the variation which occurs within an observer as a result of multiple exposure to the same stimulus[25]. The stability of a measure is especially relevant when the measure is used to evaluate change over time. In this situation, stability (referred to as *test-retest reliability*) is usually estimated by administering the measure on two occasions separated

by a time interval sufficiently short so that it can be assumed that the underlying process is unlikely to have changed. Satisfactory correlation values fall between 0.85 and 0.95[25].

Reliability has been traditionally highlighted in psychometric research, where the test of observer variability is often carried out and reported before the index is used. In the clinical field, however, this formal test is often omitted and most of the well-known clinimetric indexes, such as Apgar score or TNM staging system, were developed and published without studies on reliability, so that it remains unclear whether differences in scores are attributable to variance between the persons studied or systematic differences in the application of the instrument[9,12].

Every situation should be carefully evaluated since the circumstances vary greatly among the objects or events of interest. For example, the assessment of reliability for risk indexes for sexual transmission of HIV merits discussion. First, these indexes are mostly based on sexual behaviours, the interpretation of which can easily vary among different users (e.g., classifying the frequency of sexual practices or condom use as never, sometimes, often, always), or involve observers' judgement (e.g., deciding if a sexual partner is anonymous, casual, or regular). Second, these measures are often developed for application in different sites by people with different backgrounds (e.g., outpatient departments, psychological counselling centres, sexually transmitted diseases (STDs) clinics, and drug user rehabilitation centres). Third, personal sexual behaviour may vary greatly even over a limited period of time or change due to educational messages or interventions. All these considerations suggest that the assessment of reliability should be carefully made when developing risk indexes for HIV infection. Nonetheless, testing reliability may be difficult, since the process presents several problems. For example, checking inter-observer variability, which requires that the individual undergo the same interview with different raters, would produce substantial discomfort for the participant because of the confidential nature of the questions asked. For a different set of reasons, the second test for intra-observer variability cannot be applied too soon after the first

because the interviewer may remember how he or she rated previously; however, a delay between tests will be pertinent only if the individual has not changed his or her behaviours. Intra-observer, as well as inter-observer, variability can be readily tested when raters work with a set of well recorded data, instead of performing interviews.

Internal consistency

The internal consistency refers to the homogeneity of the items included in the measure, that is, are all items tapping different aspects of the same phenomenon?[25]. A test with high inter-item correlations is homogeneous and is also likely to produce consistent responses[17]. Methods for verifying the internal consistency can determine whether each item correlates with the total score (item-total correlation)[5] and whether the items correlate with each other (split-half reliability, Kuder-Richardson 20, Cronbach's coefficient alpha)[31]. If all items measure the same dimension and a simple additive model is used, there should be a moderate correlation among the items, while if items are measuring different attributes or a non-additive model is used, internal consistency may be inappropriate[25]. Moreover, when indexes are developed through empirical method (e.g., stepwise regression), variables are eliminated if they do not make a sufficient contribution to the total index, hence the measure will not show internal consistency.

2.3.2. Validity of a summary measure

Validity is the degree to which a measurement measures what it purports to measure[30]. Assessing validity is a process used to determine the degree of confidence which can be placed on inferences made about people based on the result of the measure[25,35]. This is particularly important when the phenomenon under study is not directly observable and when knowledge about the relationship between the observation and what it reflects is incomplete[25]. Lack of

validity arises because of the presence of bias (systematic error), which produces a systematic distortion on measuring instruments[31].

There are many ways of testing validity, the choice depending on the purpose, the conceptual basis, and the operational definition of the instrument to be validated. Some types of validity are assessed through experts' judgement (face and content validity), while others are more empirical and require data (criterion validity). The terms used for the same validity issues vary and may generate confusion. Social scientists, often dealing with "subjective" phenomena (e.g., pain), distinguish among several types of validity[31]; while clinicians use different, although parallel, terms when evaluating a clinical test or instrument[36]. Tugwell and Bombardier propose the following table (table 3) for translating among terms[37].

Table 3. Validity of a summary measure

Social science terminology	Clinical terminology
Face validity	Credibility
Content validity	Comprehensiveness
Criterion validity	Accuracy
Discriminant validity	Sensitivity to change
Construct validity	Biological sense

Face validity

"Face validity simply indicates whether (on the face of it) the instrument appears to be assessing the desired qualities". The yardstick is "a subjective judgement based on a review of the measure itself by one or more experts"[25]. Face validity is sometimes used in selecting the items or variables, looking for those which appear (on the surface) to be measuring the topic in question. Proponents of face validity state that it may increase the acceptability of the instrument[25].

Content validity

Content validity refers to the degree to which the measure includes all the appropriate components of the phenomenon under study, as specified in the operational definition and in relation to the purpose[6,30]. As with face validity, it is assessed by experts' judgement. The higher the content validity, the broader are the inferences that can be validly drawn[25]. Content validity may also be used in devising the initial pool of factors, to ensure that the measure includes the relevant factors (content relevance) and adequately covers each domain under investigation (content coverage)[25].

One way of establishing content validity is using a statistical technique known as factor analysis. It is appropriate when a measurement contains separate components, each reflecting a different aspect of the object (e.g., health). However, when factor analysis is used to examine only the internal structure of a test, rather than its correlation with other tests, it is similar to internal consistency, which is considered a form of reliability[17].

Criterion validity

Criterion validity refers to the degree to which the measure "correlates with an external criterion of the phenomenon under study"[30]. Criterion validity has also been referred to as correlational, empirical or statistical validity. For example, a new test and an established test are applied to a sample of people and the results are compared using an appropriate correlation statistic. The reason for developing a new test, when there already exists a "gold standard" (the external criterion), could be that the existing instrument is expensive, invasive, dangerous, or time-consuming. If the outcome is used as the external criterion, the test may anticipate and predict the outcome. The criterion may also be a diagnosis independently made by a clinician, if this is the accepted gold standard.

The criterion validity is usually divided into concurrent and predictive validity,

depending on the current or future existence of the criterion[31]. With *concurrent validity*, the new measure is correlated with the criterion measure, at the same point in time, and the level of agreement is observed. *Predictive validity*, on the other hand, concerns a future criterion which is correlated with the measure[31]. It is used, for example, when a test is intended to predict an aspect of future health status. It is assessed in prospective studies, where the outcomes are compared to the measurements made at the beginning of the study[17]. When the measures are dichotomous, criterion validity may be estimated through indices of sensitivity and specificity or correlation measures; if the measures are continuous, the Pearson correlation coefficient may be used[25].

Construct validity

Construct validity is defined as the degree to which the measure "corresponds to theoretical concepts (constructs) concerning the phenomenon under study"[30]. When a variable is not directly observable or there is no gold standard, several hypothetical constructs about the object measured may be proposed and tested, so that evidence from several validation procedures are assembled[17]. Construct validity, then, may be seen as an ongoing process[35].

For *convergent and divergent validity*, hypotheses (constructs) are formulated which state that the measurement will correlate with other methods and variables measuring the same concept. The measure should not correlate with dissimilar, unrelated variables[25]. Another example of construct validity is the *group differences or discriminant validity*, which may be used if an index is intended to distinguish between different categories of respondents. The measure may be tested by applying it to groups with extreme characteristics for the phenomenon under study and by analyzing the scores for significant differences[6,17].

Importance and appropriateness of the types of validity

Final acceptance of the validity of a measure depends on the collective judgements of persons knowledgeable in a specific field[35]. However, the different types of validity described above do not have the same importance or strength. The criterion validity is the more classic and clear type of validity. For the others a joint committee consisting of The American Psychological Association, The American Education Research Association, and The National Council on Measurement in Education recommends that content validity be viewed as an independent form of (internal) evidence to complement convergent and divergent (external) evidence supporting validity. At the same time, the joint committee does not consider face validity a form of validity at all, because it is not an appropriate or reliable basis for inference, since there are no empirical ways to examine it consistently[38].

The appropriate way to assess validity depends on the characteristics of the measure, in particular its intended purpose, and on the existence of an appropriate gold standard[13,17]. Risk measures can often rely on criterion validity (since there is usually an external criterion for the object or event that is measured) as well as on content or construct validity. Criterion validity, however, is not as straightforward as it seems. For concurrent validity the gold standard may be regarded as not very good or the index itself may become the definitive standard. In these circumstances, there is nothing with which the validity of the new measure can be compared. With predictive validity the main problem relates to the time lapse necessary to test the predictions made the first time the measure is applied. This implies a long and expensive prospective study and a delay in obtaining the validity results. Second, the predicted outcome may be altered by interventions suggested by the predictions itself (e.g., therapies and behavioural changes) or by extraneous and unknown factors. On the other hand, when a gold standard is not available, the completeness of the factors covered by the index should be addressed with content validity and the external evidence can be addressed with construct

validity either with a cross-sectional approach (relationship between index and external measures at the same point in time) or with a longitudinal approach (relationship over time)[13].

A clear definition of the conceptual basis underlying the measure allows extensive evaluation of the developed instrument, through the formulation of several hypotheses, which can be tested with construct validity. Moreover, since the measuring instrument itself is not being validated, but the measuring instrument in relation to what it measures, the choice of the strategy for testing validity depends on the population and the setting for which the measure is intended. It is important to consider all these factors so that the appropriate validation is conducted, otherwise the measure could be validated only partially or not at all.

Although the assessment of validity is an ongoing process where evidence is gathered over time, some basic evaluation should be presented together with the novel measure.

As stated in this chapter, the major problem in constructing summary measures regards the confusion in the literature concerning the available "methods" to develop these instruments. Different disciplines use dissimilar methods, and also within the same discipline several strategies are employed. Psychometrics is a well developed, although controversial, field for psychological instruments and educative tests[5,39], while more recently, principles and approaches have been established for other types of measurements, such as health status indicators[17] and clinimetric indexes[12].

One of the goals of the present work is to study the methods developed by different disciplines for the construction of summary measures of risk and to develop a classification of these methods. Organizing the existing material into generalizable categories could help in knowing more about the single strategy and could provide a picture of the potential procedures to develop summary measures of risk. This may enable researchers to assess existing risk assessment measures or to select the most appropriate method for developing a new one.

CHAPTER 3 - METHODS FOR RISK ASSESSMENT

The previous chapter describes the basic characteristics of a summary measure, the steps needed for its development, and the importance of assessing validity and reliability. Here, the focus will be narrowed to summary measures of risk for multifactorial phenomena, which must be assessed as a combination of factors.

A summary measure of risk implies that the component variables to be included be referred to as risk factors[22]. A risk factor is "an aspect of personal behaviour or lifestyle, an environmental exposure, or an inborn or inherited characteristic, which is known ... to increase the probability ... of a specified outcome"[30]. Sometimes risk factors are truly etiologic factors for the outcome, but often they are mere indicators (or markers) for the likely presence of unknown etiologic mechanisms[22].

The previous described characteristics of a measure are identical to those for a summary measure assessing risk. The operational definition will clarify which risk needs to be estimated and the body of knowledge to which the researcher refers. The risk can be related to a clear and measurable outcome, such as a disease or an event (e.g., death or hospitalisation), or it may be difficult to measure, as in the case of an exposure. The purpose of a summary measure can be, as stated, either discriminative (distinguishing between groups on an underlying dimension), predictive (predicting a future risk), or evaluative (measuring the change in the dimension of interest).

As introduced in the previous chapter, the development of a summary measure should be performed according to a standard method. However, as in other fields, the confusion regarding the available methods for risk assessment was the most surprising discovery upon review of the literature. A large number of risk assessment indexes have been developed with an ad hoc approach, or not appropriately evaluated. Through reviewing the literature, the

methods for risk assessment that various disciplines have developed, have been identified. Similarities and differences among the methods have been studied in order to decide how to organize and classify them. Finally, strengths and limitations of the methods are compared, together with factors influencing the choice.

3.1. LITERATURE SEARCH

The source of information for the available methods for risk assessment was a systematic review of the literature, the raw material being papers containing indexes assessing risk.

3.1.1. Methodology

Since several health related areas have produced summary measures of risk, different computer-based bibliographies have been explored. The search for medical literature was conducted using the MEDLINE system (one computerized search covering the period 1982-86 and one for 1987-91). One search was performed on a psychological data-base (Psych-LIT 1983-91), and two more for the sociological area (Sociological abstract 1963-91 and Family resources 1970-91).

As risk indexes have received little attention as a distinct entity, there is a lack of specific key words for locating them with the customary procedure. Words such as "risk index", "risk assessment", and "risk appraisal" are not available as MESH (Medical Subject Headings) terms. To overcome this difficulty, the following strategy was used to find the materials through computerized searches. The first key concept to express summary measure or assessment used "health status indicators" (a substitute for health status index or health risk appraisal) and "severity of illness index". Since these MESH terms were not comprehensive, the following words were searched as text word in the titles and the abstracts: "index(es)", "indices",

"scale(s)", "score(s)", "appraisal(s)", and "assessment(s)". A second key concept was needed to restrict the search to articles containing methodological elements. The MESH headings were: "models, theoretical"; "models, statistical" (which includes likelihood functions and linear, logistic, and proportional hazard models); "probability"; "statistics" (which include, among others, actuarial analysis, analysis of variance, cluster analysis, discriminant analysis, factor analysis, and regression analysis); "Bayes' theorem"; and "decision theory". The MESH subheading "methods" was also used. Cross referencing was performed with the concept of risk by the key words "risk", "risk factors", and "risk-taking".

All articles were searched for relevant secondary references, which were then pursued in a branching pattern. A further search for books was conducted, with similar key words, through the computer-based catalog at the Health Science Library and at the Morriset Library (humanities and social sciences) of the University of Ottawa.

3.1.2. Inclusion criteria for papers

Approximately 1,200 titles were retrieved as a result of the computerized searches. Because of the extensive use of text words and the lack of specific key terms, after the abstracts were scanned for appropriate content, only 170 papers were considered to be useful and were actually collected. Approximately the same number of articles was identified through references cited in the original material.

Since the interest was focused on summary measures for global risk assessment, three inclusion criteria were applied to the articles. In order to be selected, a paper needed to regard some kind of risk, imply a multifactorial assessment, and provide a summary measure. After this additional screening, about 160 papers were retained, two thirds of which described actual risk indexes and the remaining third which provided methodological discussion on some aspect of the development and evaluation process, or which compared different measures.

3.2. CLASSIFICATION FOR RISK ASSESSMENT METHODS

3.2.1 Development of a classification

The research was planned with the idea that, through the review and the appraisal of the material, the general principles and the classification for risk assessment methods would have emerged quite easily. All that seemed necessary was collecting the reports and then dividing them appropriately.

I started to sort the articles by their main topic, such as clinical indexes, health status measurements, severity of illness scores etc., but for many of them it was difficult to find a suitable class. Then, I tried to arrange them according to their functional role and the area of interest in categories like clinical judgement, risk factor assessment, health risk appraisal, decision making, environmental risk assessment etc. Again, this division did not exemplify the underlying methods, and it became clear that there was no simple way to classify them. I eventually decided to concentrate on the way in which the indexes were constructed, without regard to their object or purpose. Papers containing information about the development procedure were selected and sorted in reasonably homogeneous groups. This was an iterative process of assembling, further subdividing and dismantling the categories, looking for key criteria that pooled similar indexes together and different measures apart. In fact, each group should have been homogeneous enough to warrant internal differences. As a result of this work, a classification is proposed (table 4) with the intent of creating a framework for the development of summary measures of risk.

3.2.2. Proposed classification

The dictionary of epidemiology describes the characteristics that a classification ideally should have: "1) naturalness - the classes correspond to the nature of the thing being classified; 2) exhaustiveness - every member of the group will fit into one (and only one) class in the

system; 3) usefulness - the classification is practical; 4) simplicity - the classes are not excessive; and 5) constructability - the set of classes can be constructed by a demonstrably systematic procedure"[30]. The inductive process used (i.e., inferring the methods from the actual indexes found in the literature) should foster the naturalness of the classes. At the same time, the proposed classification is intended to be useful in helping the potential developer in choosing the best method according to his or her needs, as well as the consumer of an existing measure in understanding how the measure has been developed or how it can be modified for reaching the specified goal. The attempt was to keep the classification as simple as possible; at the end of the process, five methods for constructing summary measures of risk were identified, although subclasses have been created for each method.

The categories should act as a mechanism which helps both in classifying the available indexes and in providing a better understanding of problems and robustness of indexes developed with a certain standard method. However, as for every classification, some measures will fit well, whereas others will not. The criterion chosen to construct the classes was the different bases used for the development process. The following five main classes were created:

Table 4. Proposed classification of risk assessment methods

Risk assessment method	Development basis
1. Empirical method	based on collected data
2. Judgemental method	based on experts' opinions
3. Psychometric method	based on theoretical construct
4. Mathematical method	based on mathematical model
5. Ad hoc approach	none/unsystematic

3.2.3. Development basis

1. The *empirical method* relies on collected data to develop summary measures. The researcher applies statistical techniques to data-sets to identify variables that appear to be highly

correlated with the outcome of interest and to find their weights. The empirical method is commonly used by epidemiologists and biostatisticians.

2. The *judgemental method* relies on the opinion of experts to identify the variables and to determine their relative importance. Group process procedure and scaling techniques are commonly employed for obtaining and improving choices and estimates. This method has been developed mainly in the social science and economics areas.

3. The *psychometric method* is built on a theoretical hypothesis about the construct, which attempts to explain the phenomena. The preliminary instrument, based on the subjective judgement of the developer, is submitted to an empirical analysis for evaluation and revision. The psychometric method evolved from the field of psychophysics and is the prominent procedure in psychological and educative areas.

4. The *mathematical method* is based on an abstract representation in mathematical form of the phenomenon under study. The choice of the mathematical model is made a-priori by the researcher, who assumes that it can be used to describe the phenomena. The development does not incorporate information gained by collected data, although the developer refers to available knowledge and estimates of the parameters included in the model.

5. The *ad hoc approach* is used here to denote those procedures for which the development process is invented at the same time as the measure itself and is based only on the researcher's personal opinion. The ad hoc approach cannot be classified as a true method because of the lack of a standard and replicable strategy.

3.2.4. Criteria to assess the methods' strengths and weaknesses

Each method has strengths and weaknesses. Both the developer and the consumer should be aware of them in order to understand which are the pitfalls and the strengths of a summary measure developed according to a certain method. Several authors have suggested guidelines

for evaluating and selecting appropriate summary measures for clinical assessment[12,40], health status appraisal[17,41-43], or trauma severity scoring[21]. However, these criteria are not fully appropriate for comparing development methods. Gustafson[21] and Alemi[44] identified six features, some of which are similar to those for evaluating indexes, to be considered in appraising different methods.

Three criteria are used to evaluate each method with regard to the validity of the results. That is, which methods produce summary measures with more face, content, and criterion validity? In fact, the goal of any risk assessment method is to develop high quality measures, as discussed in the previous chapter. Two more criteria appraise the generalizability to other populations or settings and the modifiability to new or changed situations of the instruments produced by different methods. The sixth feature considered is the feasibility of the method itself in terms of time and resources required. Finally, since the importance of a method is to substitute ad hoc approaches (which are based on personal opinions and impressions) with standard procedures, one more criteria was added concerning the potential for subjective bias. Therefore, the following seven criteria (table 5) will be applied to each risk assessment method and will provide elements to evaluate and compare them.

Table 5. Criteria for evaluating risk assessment methods

a. potential for subjective bias
b. modifiability and adaptability
c. generalizability
d. credibility, acceptability or face validity
e. comprehensiveness or content validity
f. accuracy or criterion validity
g. feasibility

In the followings sections, each method will be examined in detail. First, a definition of the risk assessment method is provided together with the description of one summary

measure developed according to this method. Then, the available development strategies for each method, summarized in table 6, will be delineated. When several sub-methods or development strategies are presented, similarities and differences between them and the factors influencing their choice will be considered. Each section will also contain examples of possible procedures for assessing the validity of a new measure and a discussion about the strengths and weaknesses of the method.

Table 6. Risk assessment methods and method development strategies

Risk assessment method	Method development strategies
1. Empirical method	Univariate approach Multivariate approach
2. Judgemental method	Explicated approach Derived approach
3. Psychometric method	Integrated approach
4. Mathematical method	Binomial models Poisson models Other models

In the final section of the chapter, the methods will be compared according to the seven criteria identified above, and factors that should be considered in the choice of the appropriate method for risk assessment will be highlighted.

3.3. EMPIRICAL METHOD

3.3.1. Introduction

Definition

The empirical method uses empirical data to develop a summary measure of risk. Statistical techniques are applied to sets of data to select the variables that best correlate with

the outcome of interest[23] and to determine their relative contribution (weights)[6].

Although mainly based on empirical data, several steps of the development process will involve the developer's judgement, in addition to the preliminary evaluation regarding whether existing data-bases could be used or new variables should be collected. The developer must choose the model to aggregate the variables, based on assumptions regarding the relationship between variables and the outcome of interest (linear, multilinear, multiplicative, logistic, etc). The developer must also state the statistical criteria and the best way to enter the variables in a multivariate model (e.g. forward, backward, stepwise etc.) and judge whether there are variables that should be included for their biological credibility, and whether there is a need for interaction terms.

Description

The best known summary measure developed with an empirical approach is the Framingham equation for the prediction of risk of cardiovascular diseases over a period of time[4,10,45,46]. It will be used to illustrate this method. A cohort was established in 1948 and followed prospectively to provide data. The variables (risk factors) to be collected were decided by the researchers in advance. Assumptions were then made regarding how these factors relate to the risk of developing the disease. This functional relationship has been formalized by using a mathematical model: the logistic regression model. The model allowed for the selection of variables which were significantly related to coronary heart diseases, when the level of additional variables was controlled, and the estimation of their relative impact on the risk of disease as a function of antecedent factors. The final equation can be used to estimate the risk of coronary heart disease over a number of years.

3.3.2. Development strategies

Univariate approach

The univariate strategy to develop summary measures identifies risk factors and evaluates their individual contributions to the outcome. Variable selection is usually based on univariate statistical tests, such as chi-square or Pearson correlation coefficient (depending whether the outcome is dichotomous or continuous), while measures of association (e.g., relative risk) are used to assign weights. The underlying assumption is the independence of the factors.

An index of severity for alcoholic liver disease to predict the risk of mortality can be used to illustrate this development strategy[47]. Data were obtained from a prospective analysis of 253 patients with alcoholic liver disease who were followed up for one year. Variables with statistically significant prognostic value for mortality (chi-square, Mann-Whitney, and t test) were selected, without controlling for the effect of other variables. Developers decided to assign weights based on relative risks. The range of relative risk was subdivided into three categories, to each of which a value from 1 to 3 was assigned. An additive model was chosen for the final score.

A slightly different strategy was employed by Fedrick, who used an available data-base (the British Perinatal Mortality Survey) to calculate relative risks. All the variables with a relative risk greater than one were included and their relative risks were multiplied together to produce a composite score to identify pregnant women at high risk for spontaneous pre-term birth[48]. Univariate analyses were also used to create checklists for risk, regardless of the factors' weights[49,50].

Development strategies based on marginal relative risks, that is unadjusted effects, are likely to produce summary measures that are misleading in the identification of people at risk, because the effect of each factor is not controlled for the presence of the others. However, when

a large number of variables are available, univariate strategies have been employed to provide an initial screening of the factors associated with the outcome. Only those variables with a statistically significant association with the outcome are later used in multivariate techniques[50-54].

Multivariate approach

The simplicity of the univariate strategies has the major pitfall of not considering that the risk, associated with any individual risk factor, varies widely depending on the number and the influence of coexisting factors. Multivariate strategies can take into account the joint effect of multiple contributions[45]. They also have the advantage that both risk-increasing and risk-decreasing factors can be considered[22], and that variables with biological credibility and interaction terms can be taken into account. The most popular strategies for selecting, weighting, and combining several variables into a summary measure are based on regression methods. These methods characterize a relationship between a dependent variable (outcome) and one or more independent variables. Other methods (such as logit, probits, etc.) may be employed, but they have not been used lately to develop summary measures of risk in the articles reviewed.

a) Multiple linear regression analysis

Linear regression is applied to continuous dependent variables, although the independent variables may be continuous or discrete. The multiple linear regression function ($y = a+b_1x_1+....b_nx_n+e$) expresses the dependent variable as a linear function of a number of predictor variables. Coefficients represent the predicted change in y per unit change in each independent variable holding the others constant, while 'a' is the intercept of the line. The model requires that the dependent variable have a normal distribution and that the variance of

y be the same for any x.

This strategy was employed to develop a decision rule for predicting seat belt use from data on 3,108 Tennessee residents. The data-base was randomly divided into derivation and validation sets. The dependent variable was the self-reported seat belt use, while the independent variables were based on personal health information. Stepwise multiple linear regression was used to select those variables associated with seat belt use (percentage of time used), although the assumption of normality for the dependent variable was not verified. The regression coefficients were then used as the weights for each of the independent variables in a final score. Cut-off scores were decided in order to classify people at different risks for seat belt use[55].

b) Linear discriminant analysis

The linear discriminant function model has frequently been used for binary outcome, but it had been originally planned to distinguish polytomous unranked nominal categories[56]. On the other hand, it may not be appropriate when many of the independent variables are binary or ordinal scaled[27]. This method involves deriving a linear combination of those independent variables which vary greatly among individuals belonging to different subgroups, compared with the small variation among individuals in the same subgroups. The linear discriminant function may be expressed as $f = a+b_1x_1+\ldots b_nx_n$ and can be computed for any individual whose set of x values is known. The resulting discriminant score is compared with a threshold value to classify the individual into one of several risk categories.

An example of this method is the development of a malignancy index for renal cell carcinomas[57]. In a stepwise procedure, the histologic and clinical characteristics of the tumours in 55 patients yielding the maximum separation of the two risk groups of patients (good and poor prognosis) were selected and their weights were determined. The same strategy has been used for calculating a discriminant score to classify patients in high/low risk groups for lack of

appropriate dental care[53], for adult respiratory distress syndrome development[58], and for ischemic heart disease development[59]. It has also been used to develop a score to identify intravenous drug users (IVDUs) at high risk for acquiring HIV infection[60]. The general strategy may be applied with some modification, such as transforming discriminant coefficients into integer weights or choosing a multiplicative model if this produces a better predictive value[61].

Discriminant function analysis was also used to create a checklist of antepartum and intrapartum characteristics for the classification of infants as at high or low risk for abnormal outcomes (e.g., birth weight, gestational age, neonatal status, etc). Variables were entered into the stepwise discriminant analysis in the order of their ability to discriminate between normal and abnormal outcome; those resulting in the greatest increase in the overall multivariate F ratio for the difference between group centroids were entered first. The final checklist contained only variables found to be predictive of the outcome[62].

c) Logistic regression model

The logistic regression model is more appropriate when the outcome (used as dependent variable) is a binary event, such as alive or dead. The model describes the probability (p) of an event occurring during some specified time interval as a function of a set of independent variables: $p = 1/[1+\exp-(a+b_1x_1+\ldots+b_nx_n)]$ where 'a' is the constant for the intercept (reflecting the general level of risk) and 'b's are the regression coefficients (reflecting the weighted contribution of each factor). The underlying assumption is that the log odds for the occurrence of the outcome is a linear function of the variables included in the model[63].

The most common use of the logistic model in summary measure development is the production of a risk equation from the analysis of a prospective study. For example, a forward stepwise model fitting process was used to identify independent predictors and significant interactions from a pool of potential variables in a 7-year prospective study on

institutionalization[64]. A similar strategy was used in determining risk of cardiovascular disease[4,10,45,46], and risk of hospital mortality for burnt[65], injured[54], and general patients[51,52,66].

When the data set derives from a retrospective or cross-sectional study, logistic model can still be used to derive a risk score. The application of a linear logistic regression model to perinatal data on 8,240 deliveries identified 18 variables significantly associated with prematurity. Using the resulting equation, each patient was classified into either a high or low-risk group by defining a general cut-off score[50]. A score for risk of heart attack used coefficients derived from multiple logistic regression just as an indication of the importance of each factor. In this application, the actual weights were rounded and the constant was not used[67]. Logistic regression was employed in another study to select those intrapartum factors which better predict infants with low Apgar score[68]. The final checklist contained 10 factors which were just counted.

In the HIV field, the logistic model has been used for providing a probability equation to calculate the risk of being positive for HIV among homosexuals[69] and for developing a checklist to classify heterosexual partners of infected people according to their level of risk of being infected[70].

d) Cox's proportional hazards model

Cox's regression model is intended to evaluate the entire dynamic pattern of a survival curve, rather than a static outcome, and the prime interest is usually a time-to-event outcome. The model assumes that the risk of disease is a dependent function of a set of variables which themselves may change with time. It allows one or more predictor variables to be related to the risk of an outcome, and uses the follow-up information from all participants, including those who are not followed until the outcome[56,71].

A prospective study in Japan was undertaken to determine global estimates of prognosis

in patients with liver cirrhosis, using clinical and laboratory data. Variables possibly relating to the prognosis were screened separately by multiple regression analysis in a stepwise process, and six were found to reflect the prognosis. Cox's proportional hazard regression model (in a stepwise manner) identified three independent variables and was then used to determine the contribution of these variables to survival: $h(t)=h_0(t)\exp(b_1x_1+\ldots+b_nx_n)$. The risk or hazard for death, which describes the instantaneous probability of outcome in a short interval of time, is a function of a basal or underlying death risk $h_0(t)$ and of the independent variables x_1 to x_3 weighted with the corresponding regression coefficients b_1 to b_3. A prognostic score ($\log_e[h(t)/h_0(t)] = b_1x_1+b_2x_2+b_3x_3$) for a given patient is calculated, where a higher score means a higher risk or poorer prognosis. The patients were divided into high, medium, and low risk groups, according to their score[72].

Cox's proportional hazard regression was employed to predict coronary heart disease and hypertension in young adults followed for 35 years. Risk factors selected from univariate analyses (Kaplan-Meier survival curves) were entered into the Cox's model, which provided the equation to calculate the probability of developing the outcome within 30 years[19].

3.3.3. Validation strategies

Face validity

Measures developed with an empirical approach require that a judgement be made concerning the credibility of the model (linear, multiplicative, etc.), of the assumptions made (independence, normality, etc.), of the technique (chi-square, Pearson correlation, linear regression, discriminant analysis, etc.), of the variables' scale, and of the cut-off points (for checklist and scoring systems). Moreover, in the case of multivariate strategies, the way variables are entered into the model (e.g., forward, backward, stepwise process) may be appraised for face validity. In fact, the use of different selection processes and criteria of

statistical significance may lead to different results, especially when some variables are highly intercorrelated[73].

Content validity

Since empirical approaches select variables from collected data, testing for content validity involves a judgement regarding two stages of the development process. First, an analysis of the comprehensiveness of the collected variables should be made, and then an evaluation of the importance of the variables (and the interaction terms) included by the statistical technique or forced into the model by the researcher because of their biological credibility should be carried out.

Content validity is also viewed as a form of internal validity of the measure[38]. Any multivariate statistical technique is based on a series of assumptions (e.g., a normal distribution for variables is frequently assumed). The data on which the technique is based should satisfy these assumptions. A critical step in assessing the suitability of the model is to examine its fit, or how well the model describes the data from which it was developed. Assessing the goodness-of-fit of the model usually entails two stages: computing a statistic providing a summary measure of the errors (e.g., standard error of the regression, R-squared, deviance, Pearson chi-square, etc.), and examining the individual values of the errors (regression diagnostic)[6,74].

Criterion validity

Although a measure developed with an empirical approach may be validated using a "gold standard" (such as a clinical judgement or a diagnostic test) often the "gold standard" is the outcome measure, which has been used to develop the measure. An index developed from a data-base with a predictive purpose should be tested on another data-base for its predictive ability to classify the individuals (using sensitivity, specificity, etc)[51,61] or to assess the outcome

(assessment of fit of the expected probabilities to the observed experience)[11]. Often risk scoring systems are validated by referring back to the same set from which they were derived (back-validation)[22], or by splitting the data randomly into two portions, one for model development and the other for model validation (cross-validation)[75]. Since by definition, a statistical index best fits the data from which it was estimated, the crucial test is on a new data-base. This is very expensive and time-consuming, especially when external data-bases are not available and new data have to be collected (temporal validation as Miller called it)[75]. Criterion validity is also required when investigators wish to ascertain whether a model developed at some other site is appropriate for prediction in their own particular setting[75].

Construct validity

Construct validity involves an assessment of how well the results of a new instrument agree with other existing measures in a manner which is consistent with theoretically derived hypotheses concerning the construct that is being measured. However, this type of validity is rarely assessed in empirically developed measures, which are mostly tested for criterion validity. Nevertheless, comparison with other empirically derived measures may be important in light of the phenomenon of multicollinearity. In fact, "if the selection of the components and the relative weights given to each component varies depending on the statistical method used as well as the clinical data-set on which it is estimated, the results then would not show any convergent construct validity"[6].

3.3.4. Discussion of the empirical method

Similarities and differences between univariate and multivariate approaches

The univariate and multivariate approaches share the basic characteristics of the empirical method; specifically, the application of statistical techniques to collected data both for

selecting variables correlating to an outcome and for estimating their weights. Major differences are related to the simpler development required by univariate strategies, which, on the other hand, have the disadvantage of not considering the influence of coexisting factors.

Multivariate strategies appraise the joint effect of all relevant factors, including variables forced into the model and interaction terms; however, they require a large amount of data, and usually utilise complex models (which have several assumptions to be satisfied). The univariate approach may serve as a useful first step for multivariate strategies, but its usage as an autonomous approach is largely questionable because the underlying assumption of independence among risk factors is usually not true and the use of marginal risks produces biased results.

Factors influencing the choice of technique

In deciding which technique should be used, the following factors have to be considered. Techniques diverge in their ability to provide different types of measures (e.g., not all of them can produce probability estimates). In addition, each technique has its own theoretical distinctions for the dependent variable (binary or continuous outcome). Furthermore, the efficacy of a multivariate technique can be affected by the statistical distribution of the data (e.g. normal, bimodal, exponential, etc) and by the mathematical structure of the relationship between the variables and the outcome (linear, multiplicative, logarithmic, etc)[76].

Strengths and weaknesses of the empirical method

The following discussion on the strengths and weaknesses of the empirical method refers to the criteria identified in section 3.2.4. It will focus mainly on the multivariate approach, since the univariate strategy is based on criticizable assumptions, already discussed, that make problematic its use as an autonomous approach.

a) potential for subjective bias

The main strength of the empirical method is the reduction of subjective bias in the development process. Indices based on appropriate statistical analysis of a large data-base tend not to have arbitrariness due to the researcher in the selection of variables and in the choice of weights assigned[61], although there could still be biasing effects due to sample selection. However, subjective bias may be introduced in those decisions where the judgement of the developers intervenes, especially in the choice of a multivariate statistical model to fit the data. This is a crucial step and therefore the model adequacy has to be assessed carefully.

Once the model is chosen, other decisions may introduce bias in the development process, such as the choice of the variables for inclusion in the model and the selection of the statistical criteria and the way to enter the variables in the model. Finally, it may be appropriate to begin the multivariate analysis with a set of variables known to be related biologically to the outcome under study and to force them into the model or to evaluate whether products of two or more variables should be included in the model in order to assess possible interactions among these. However, the inclusion of such interaction terms may introduce bias that cannot be detected due to the complexity of the model and the collinearity among variables.

b) modifiability and adaptability

Including new factors or changing the assigned weights to modify a measure for another population or situation is difficult for instruments developed with empirical method. Empirically developed measures are quite inflexible and can only be modified with difficulty. For example, in estimating the probability of an event for a particular configuration of significant predictors, when not all risk factors are known, the most acceptable approach is to refit a new regression model using only variables for which data are available and to use the coefficient from the new model to calculate the probabilities[64].

c) generalizability

Statistical parameters (even if derived from a large data-base) tend to be influenced by the data from which they are derived. Therefore, generalizability of an empirical-based index to different populations is suspect and often criticized[23,54]. There exists a debate between those who argue for population-specific measures[29,50] and those who prefer indexes developed from larger samples[29,51]. Another recurrent methodological issue related to the generalizability is whether data on outliers should be trimmed[18].

d) credibility, acceptability or face validity

In order to have face validity, a measure must contain risk factors which are known to be related to the outcome of interest[51]. Selection of variables performed with correlation methods may sound unclear and the measure may lack credibility and acceptability. Clinicians frequently consider the face validity of empirical approach less than satisfactory[6].

e) comprehensiveness or content validity

Comprehensiveness of the variables is limited to those that have been collected, since empirical method cannot be used to determine in advance which are the data to be collected[6,22]. With multivariate strategy, content validity is affected by multicollinearity problems when independent variables are highly correlated with each other. In this situation, there is a somewhat arbitrary selection of factors in the model, which leads to the absence of a definitive, single multivariate solution. This may also contribute to a decrease in face validity. A further limitation of multivariate strategies depends on the fact that techniques for assessing the goodness-of-fit can aid the researcher in examining the fit of the model to the observed data, but they cannot give conclusive evidence regarding internal validity and the adequateness of the model, and the relationship between the phenomena and the risk may still not comply with the

assumption of the statistical model used[18].

f) accuracy or criterion validity

Among the few studies comparing measures developed through different methods, a review of indices for risk of preterm birth found that the predictive power of scoring system was independent of the way the score was constructed and of the employment of sophisticated statistical techniques or arbitrary judgements to assign the weights[22]. Another study comparing judgemental and empirical severity of illness systems for predicting mortality in hospitals does not suggest that any one method always performs better than others, if tested in a wide range of settings. Both empirical and judgemental methods seem to be accurate and each method has specific characteristics that may influence the choice of one method over another[11].

g) feasibility

It is difficult and expensive to collect data-bases large and representative enough to provide meaningful coefficients[23]. The availability of collected data may help the feasibility of this method; but often the absence of data-bases, the poor quality of the existing data, or the lack of important variables, require the collection de-novo, which is expensive and time consuming.

Summary

In conclusion, the empirical method requires the presence of a quantitative outcome, and a large amount of data. The major strengths are the reduction of subjective bias in the development process and the accuracy of the results. The weaknesses of this method are related to its poor modifiability and feasibility, especially when new data have to be collected. Generalizability, credibility, and comprehensiveness greatly depend on the way the empirical

method is employed.

3.4. JUDGEMENTAL METHOD

3.4.1. Introduction

Definition

The judgemental method is based on subjective opinions of a panel of experts[44] to identify the principal factors involved in the index and to weight them[21]. A general assumption underlying this method is that experts can be effective estimators of probability, value, or preferences[14,77]. Methods for selecting the panel, conducting the discussion (group process techniques), and obtaining the estimation (scaling techniques) have been developed in order to correct possible bias in the experts' subjective judgement and to improve the validity of indices developed according to this method[23]. Factors aggregation may be based either on a simple additive model or on more sophisticated models (e.g. derived from Bayesian theory and Multiattribute theory).

The developer using the judgemental method generally should decide how the panel is selected, which group process procedure and scaling techniques are preferable, and which is the most appropriate model to combine the factors. The initial pool of variables could be conceived either by the researcher or by the panel, while the criteria for the variable selection and the weights are based on the experts' knowledge. The judgemental method does not require a dependent variable, and it can be employed to develop predictive as well as evaluative and discriminative indices.

Supporters of this method argue that there are some phenomena (such as severity of illness or quality of care) which are "subjective" or "judgemental" by nature. The phenomenon of interest must therefore be defined by the opinions of experts[23], and the decision concerning factors and their contribution to an index should mainly be a subjective activity[78]. Indices based

on experts' judgement may be capable (at least partially) of controlling other factors which affect the phenomenon under study (e.g. quality of care when making severity assessment), whereas for empirical indices, produced to correlate with an outcome, it is difficult to factor out these other influences[23]. The choice of a comprehensive and mutually exclusive set of variables according to the experts' knowledge and to a specific theory about the phenomenon may help in explaining the phenomenon and in testing hypotheses.

Description

One of the first applications of the judgemental method to risk indexes was the development of a summary measure to discriminate between burnt patients based on their severity of illness[79]. Gustafson designated a panel of six experts which were involved in sessions conducted with nominal group technique. First, the judges identified variables considered to be important in evaluating severity of illness and developed criteria measures using nominal and interval scales. Next, the judges were involved in a quantitative estimation session, in which they determined the severity function associated with the continuous variables and calculated the relative importance of each factor using category scaling for interval variables and magnitude estimation for nominal measures. The factors were finally aggregated using an Additive Multiattribute Utility (AMAU) model to measure illness severity.

3.4.2. Panel of experts

Selection of the panel of experts

The experts or judges can be researchers, clinical experts, patients, patients' relatives, or others[14]. In choosing the panel, it is essential that estimators who understand the subject matter be selected (i.e., judges should have expertise on the impact of the risk factors on the phenomenon being studied). Studies show that the knowledge about the subject improves the

accuracy of the estimates[44]. A short training phase for the experts (e.g. in general theory of probability) may further improve calibration[44]. Heterogeneous groups produce a higher proportion of high-quality solutions, while homogeneous groups have been found to facilitate group performance because of the reduced likelihood of interpersonal conflict. This suggests that a heterogeneous group is preferable if its negative effects can be controlled[80].

Group process techniques

The panel has, at first, the qualitative task of identifying criteria considered important in selecting the appropriate variables to be included in the summary measures. The second task is a quantitative evaluation of the relative importance of each chosen factor. A process of group estimation, discussion, and re-estimation may reduce inaccuracies in individual estimates of the weight of different factors by as much as 33%[81]. Unless a group functions effectively, errors might be made that could damage the quality of the index. For each factor the individual estimates are averaged, often using the geometric mean because it reduces the influence of any one extreme estimate[44,77,79,82,83].

Group processes used to develop qualitative models and conduct quantitative evaluation can be classified into five main types.

a) Unstructured or interacting group

"Generally an informal and unstructured group process with minimal direction supplied by its leader"[23,84]. Group size should not be larger than seven participants in order to function properly[80].

b) Consensus group

This group functions with the guidance of some behavioural rules. For example,

participants should avoid arguing for personal rankings and should abstain from changing their mind just to elude conflict. Participants understand the others' position, but discussion does not force consensus. Modification is encouraged only if a participant supports the logic of such a change. The group is often conducted by a group facilitator[23,85].

c) Nominal group

This involves a structured process conducted in four phases: silent generation of ideas in writing, feedback from group members to record each idea in a terse phrase, discussion, and rank-order rating on items by individual group members[23]. The meeting strategy guarantees different processes for each phase of creativity, balances participation among members, and assures that consensus is not forced[80,86]. This group can accommodate a maximum of approximately nine members[80].

d) Delphi group

"Unlike the previous three group processes, where all participants meet to engage in discussion, Delphi participants may never see each other. The Delphi process is quite similar to that of the nominal group, except that it is conducted via an anonymous questionnaire"[23] interspersed with summarized information and feedback of opinions derived from earlier responses. There is no limit to the number of participants in a Delphi group[80].

e) Integrative group

Gustafson proposes a new structured group process which integrates several existing group processes. He suggests the use of a group facilitator and the separation of idea generation (model building session) from idea evaluation (quantitative estimation session). This technique is based on a "estimate-talk-estimate" process, where group members formulate their own model

or estimates before discussion takes place; changes in position should be encouraged, but consensus is not forced[23,44,81].

Agreement among experts

When a group process is complete, it is important to determine whether the experts are acting as one (i.e., to what extent the experts in the group agree with each other)[6]. Among the possible measures to evaluate the extent of agreement, Kappa statistics have been suggested with values greater than 0.75 indicating excellent agreement, and values below 0.40 representing poor consensus[77].

3.4.3. Development strategies

Explicated approach

In the explicated approach to developing summary measures, the experts that make up the panel are interviewed about the risk assessment process that they are using. In this case, judges have to explicate their judgement defining a set of rules for selecting the variables and for assigning their relative weights[6,23]. This explicit approach assumes that the panellists can make a small number of extremely precise (error-free) assessments. A model is then chosen, generally by the developer, to synthesize the experts' explicit judgement and to derive a global measure. Some of the most commonly used models are illustrated below.

a) Simple additive model

A simple additive model has been used by several scoring systems measuring severity of illness developed through a judgemental approach, such as APACHE[87,88], APS[89,90], and TISS[28,91]. Severity is a multidimensional phenomenon, which cannot be measured directly. Often it is defined as the likelihood of death or residual impairment as a result of disease[92],

although different indicators have been included in various indices.

The APACHE II system is based on the hypothesis that the severity of acute disease can be measured quantifying the degree of abnormality of multiple physiologic variables[93]. Knaus and colleagues used a nominal group process to choose the variables, to assess the scales, and to assign the weights. The weighting system is based on a scale of 0 to 6 points, and the different abnormalities are treated as independent factors. The final additive model includes 13 physiologic variables, plus age and chronic health conditions: Severity $= w_1 + \ldots + w_{13} + w_{age} + w_{chronic\ health}$. This index has been used for two purposes: to discriminate patients along the continuum of severity (which reflects or exposes to the risk of death) and to predict hospital mortality.

b) Multiattribute Utility (MAU) theory based model

"The essence of the MAU theory is to build an operational, quantitative measure of a subjective, complex, multidimensional concept by capturing the judgements, values and preferences of experts"[78]. Utility theory originated in the field of economics for quantifying and analyzing the way in which people make their choices[94]. In the last thirty years, scientists from a variety of disciplines have contributed to establish a theoretical basis for constructing multiattribute utility functions. "These mathematical developments establish conditions under which it is possible to represent a decision-maker's preference over a set of multiattribute outcomes using explicit and relatively decomposed functional forms"[95].

The use of MAU theory in regard to health risk indices was first proposed by Gustafson, who adapted it to develop severity indices[79]. The MAU theory assumes that the phenomenon of interest is some aggregate of individual criteria, and a function S can be assessed by decomposing S to its more easily measurable components ($S_{x1,\ldots,xn}$). In brief, MAU approach requires experts to identify a set of mutually exclusive and exhaustive set of factors

$(x_1,\ldots,x_n)$ which they consider important in judging the phenomenon (S), and to suggest measures for assessing each factor (qualitative phase). Judges then proceed to the quantitative session, where they should estimate the weights (w_i), which indicate the importance of each factor, identify the levels of each factor, and develop a utility measurement scale $u(x_i)$, which reflects the relative contribution of each level of that factor to the phenomenon.

The Additive Multiattribute Utility model (AMAU) takes the following form: $S_{(x1,\ldots,xn)} = \Sigma_i [w_i\, u(x_i)] + \Sigma_j [w_j\, x_j]$, where $S_{(x1,\ldots,xn)}$ is the summary measure, x_i is the extent to which a quantitative variable is present, x_j is the presence or absence of a qualitative variable, w is the relative weight with $\Sigma w_i + \Sigma w_j = 1$, and $u(x_i)$ is the utility function associated with the $_i$th variable[96]. A second model can be derived from this theory: the Multiplicative Multiattribute Utility model (MMAU) which is represented by the equation $S_{(x1,\ldots,xn)} = \{-1 + \Pi_i[1 + KK_i\, u(x_i)]\}/K$ where K and K_i are constants chosen so that $K = -1 + \Pi_i(1 + KK_i)$[96]. From these two general models, others can be developed, such as the Quasi-Additive Utility model and the Multilinear Utility model[95,96]. An assumption of independence among factors underlies all these models, while specific independence assumptions are necessary conditions for the use of each model[95].

The additive model has been used to develop indexes for burn severity[79], for trauma severity[21], for ischemic heart disease severity[23], and for risk of hospitalisation within three years[15]. The AMAU is the simplest model, and it has been shown that it performs quite well both when compared with the MMAU[23] and when some interdependence among factors exists[97]. However, Fryback and Keeney used MMAU in severity index development for trauma[98] and Alemi found that the multiplicative index more accurately simulates the experts' judgements and has more face validity for an index estimating prognosis for AIDS patients[82]. In fact, the additive index balances the impact of one or more severe conditions with the normal condition in the remaining categories. This contradicted the experts' intuition about severity of AIDS, where severe conditions in one category is sufficient for a poor prognosis. In the

multiplicative index the presence of a poor score in any category is sufficient to produce an overall poor severity score.

c) Bayes' theorem based model

Bayes' theorem is an optimal probabilistic method for revising prior opinion about the risk of a particular event, as additional information about the event becomes available, to form posterior probabilities[77]. One version of Bayes' formula makes use of the concepts of odds and likelihood ratio (odds-likelihood ratio form). The posterior odds of having a disease (or developing it) 'D' given a symptom (or risk factor) 'R' is provided by the prior odds multiplied by the likelihood ratio of observing that symptom (or risk factor) given the disease:

$$\frac{P(D|R)}{P(nD|R)} = \frac{P(D)}{P(nD)} \times \frac{P(R|D)}{P(R|nD)}$$

posterior odds = prior odds X likelihood ratio

The theorem can be used with several factors ($R_1,\ldots,R_n$), although it requires conditional independence of the factors[99]:

$$\frac{P(D|R_1,R_2,\ldots,R_n)}{P(nD|R_1,R_2,\ldots,R_n)} = \frac{P(D)}{P(nD)} \times \frac{P(R_1,R_2,\ldots,R_n|D)}{P(R_1,R_2,\ldots,R_n|nD)}$$

then the posterior odds can be transformed into a probability by the following: probability = posterior odds/ (1 + posterior odds)[77].

A judgemental approach, using Bayes' model, has been employed to develop a predictive index for the assessment of modifiable health risk for hospitalisation within three years[44]. The panel identified a list of modifiable risk factors and the levels for each factor. The judges then evaluated possible conditional dependence among factors; if dependent factors were

identified they were combined into a broader one, while factors for which the impact was manifested by another factor were dropped. Probability estimates of the impact of individual risk factors on hospitalisation likelihoods [$P(x_1|D)$ and $P(x_1|nD)$] were then elicited from the experts by asking, "Of 100 adults who have been hospitalised at least once in the last three years, how many of them have the risk factor x_1?" and, "Of 100 adults who have not been hospitalised in the last three years, how many of them have the first risk factor x_1?". The ratio of the responses to this pair of questions gave the likelihood ratio, while the average probability in the population was taken as a priori probability of hospitalisation. Through the Bayesian model, the posterior odds were then estimated for any combination of variables[44].

A slight variation was used for developing a short-term risk index for falling in the elderly, where the judges were required to estimate both the likelihood ratio and the prior odds ("Of 100 elderly, how many have fallen within 30 days?")[77]. The Bayesian model has also been employed to produce an index for the diagnosis of thyroid disease. In this case the prior odds were calculated actuarially on the basis of analysis of medical records[83].

Derived approach

A second approach has been used to develop summary measures through the judgemental method. The judges are not asked to evaluate the importance of the factors one by one, but rather to make a global judgement on a profile, which includes various factors, and to assign a global score.

Embedded in this global judgement is an implicit estimate of the relative value of the individual factors. These implicit weights can subsequently be estimated using statistical techniques such as regression analysis, where the dependent variable is the global judgement and the independent variables are the individual factors. The regression coefficients represent the weights implicitly given by the judges and are usually called "derived weights"[6]. The

derived approach assumes that the judges can make a relatively large number of holistic assessments, each of which may include a random error component.

The Hybrid Regression Model (HRM), which utilises the same utility functions used by the MAU models, has been considered the most appropriate[23]: $S_{(x1,...,xn)}=C+w_1x_1+...+w_nx_n$, where $w_1...w_n$ are the regression coefficients and C is the constant (sometimes omitted)[6]. This approach has been used to develop a severity index for ischemic heart diseases. After the panel had reached a consensus on a small number of variables, a set of hypothetical patient profiles was constructed. These profiles were employed to obtain global scores to be used in regression analysis to construct a derived model. A hypothetical patient is represented by specific levels of the selected variables, and the number of profiles depends upon the number of variables and levels. Panellists rated the profiles in terms of their relative severity[23].

The difficulty with this approach is that as the number of factors or levels for factors increases, the possible combinations to be compared increases dramatically. A condensed version consisting of only two levels (high and low) per factor, called fractional factorial experimental design (FFED), has been proposed to overcome this problem[14,26]. This design was employed for developing a scoring system to discriminate women on their risky behaviours associated with the risk of HIV transmission[100,101]. Another example of summary measure for assessing the risk of exposure to HIV, developed with a derived approach, is the HIV Virus Infection-exposure Risk Index (VIRI). Six sex related variables were combined in a tree format producing 31 possible scenarios. The panel of experts was asked to rank order these combinations according to their level of risk. After three rounds performed with the Delphi technique, the judges reached a high level of agreement on a classification with five categories of risk[8,102-104].

3.4.4. Scaling techniques

Several scaling techniques have been developed to obtain subjective judgement in order to quantify weights and utilities. The simplest is to *rank order* the factors, or factor levels, and to assign to them a numerical value. This has been used in assigning the points to the APACHE II system of severity[93].

Among the refinements to this approach, the most frequently used in developing risk indices are category scaling and magnitude estimation, which are derived primarily from the psychometric discipline[105]. With *category scaling*, the least important and the most important factors are considered the extremes. Each intermediate factor is assigned a value relative to the two points on a continuous rating scale or to a category on an equal-interval category scale[26,95]. This technique has been applied for interval-scaled measures in a burn severity index[79].

When using *magnitude estimation*, the judges select the most important factor, then compare it with the other factors, and determine how much less important in a ratio sense they are compared with the first one. This technique has been claimed to yield ratio level measures[26,106]. An application can be found in a risk index for hospitalisation[15].

The same scaling techniques can be used to provide a subjective judgement for the profiles in the derived approach. On the other hand, studies employing the Bayes' model assess the probabilities of each factor by directly asking the judges[44,77].

3.4.5. Validation Strategies

Face validity

A judgemental-based index must be evaluated for its face validity by experts other than those who developed it. They can likely be the researchers or the clinicians who will use it.

The global appraisal, regarding whether the index appears to measure the desired object in relation to its purpose, should be followed by an opinion about the appropriateness of the

panel selected, the group process procedure used, and the scaling technique chosen. Finally, the model to aggregate individual components should seem suitable for the phenomenon under study, and the required assumptions should be met[6,82].

Content validity

As for face validity, judges other than those who developed the summary measure should evaluate content validity (i.e., comprehensiveness and the importance of the factors included in the index). This can be accomplished using their knowledge or comparing the risk factors elicited and incorporated into the models with risk factors presented in other studies[77].

As a test of internal validation for judgemental indices, the explicated and derived approaches can be contrasted[6]. For example, one might ask the panel to rate a number of hypothetical patients (derived approach) to be compared with the scores obtained using the developed model (explicated approach) for the same hypothetical patients. Correlation among the results may help in determining if the model can predict the panel's insights and in reassuring that it has captured the judges' expertise[15,44,77].

Criterion validity

If the index is developed to predict an outcome (e.g, risk of hospitalisation), this can be used as "gold standard" and accuracy can be tested by comparing the model prediction with the objective data. This validation procedure requires prospective data collection and considerable time[44,77,107].

On the other hand, when there is not a directly measurable outcome, another panel of judges can be employed as criterion for external validation to check how well a model developed by one group of experts predicts judgement from another group. In this case the explicated and derived approaches are contrasted between the development panel and the

validation panel[6].

Construct validity

Index results could be associated with convergent (or divergent) indicators to test its construct validity when a gold standard does not exist. Other indirect measures of severity may be used, such as physicians' judgement, cost of care, survival rates[21,79,93] or results from other studies[77,79].

A further construct validation could test one model developed by a group of experts against one developed by another group. Such efforts ensure that the expertise being modelled is not narrowly defined[44,79]. Bombardier suggests that, since different types of judgements should lead to similar results, the contrast between explicated and derived approaches within the same panel could be considered a test for construct validity[6], instead of internal validity.

3.4.6. Discussion of the judgemental method

Similarities and differences between explicated and derived approaches

Both approaches have the basic features that characterize the judgemental approach: the use of a panel of experts to identify the factors and to determine their relative weights. The derived approach permits the investigators to evaluate combined effects on a dependent variable of two or more factors simultaneously. However, this procedure becomes extremely time consuming when the number of possible profiles is large, and it puts a heavy cognitive burden on the judge, who may not be able to give appropriate estimates for all the profiles[95].

The advantage of the explicated procedure is that it greatly reduces the number of subjective judgements required to assign weights and utility values[95]. In practice, the derived approach is often used for validation purpose, while its employment for index development is quite rare.

Factors influencing the choice of the model

After the explicated or derived approach has been chosen, the developer should decide which aggregating model is preferable. The following discussion will focus on differences among the three available families of models for the explicated approach (i.e., simple additive, MAU, and Bayes' models).

The first distinction is related to the type of summary measure needed. A model based on Bayes' theory is preferred when a probability equation is needed, while both MAU based and simple additive models can be employed for obtaining only scoring systems or simpler forms of index (e.g., classification and checklist). The second factor to consider is that models based on MAU theory and Bayesian statistics have the advantage, compared with the simple additive model, of being supported by a theoretically sound foundation which improves the likelihood that indices developed according to the model will be valid. However, the two theories require certain assumptions to be made. All MAU models assume the existence of a function that can be assessed by decomposing it to its components and require that each component be independent. Each model has specific independence assumptions which are necessary and sufficient for its existence[95,96]. The Bayes' theorem also requires conditional independence of the factors considered.

Strengths and weaknesses of the judgemental method

a) potential for subjective bias

Although the judgemental method uses a panel of experts instead of a single individual, the main criticism of this method is its potential for subjective bias[99,108]. The supporters of this method assert that group process techniques and guidelines (e.g. selecting "expert" judges, providing them with short training, anchoring estimates to known objects) may correct the systematic bias in model choice, variable selection, and weight estimate[44,77]. However, in using

group process, one must still consider the agreement among the group's members[77].

b) modifiability and adaptability

Judgement-derived summary measures are not based on any specific population and may be modified or adapted to a specialized or different environment[21,23,44]. However, a new panel of experts, or the same panel that developed the original measure, should be used to ensure that subjective bias is not introduced during measure modification.

c) generalizability

When the expertise of the judges is not limited to a specific institution or population, models based on their judgement may be more general and require less revision than models based on a specific data base[44]. The panel selection is then crucial, although the generalizability of an index is limited by the studies and the knowledge already available on the phenomenon.

d) credibility, acceptability or face validity

An index developed with the judgemental method may appear to be valid (face validity) because people understand what is being scored and how the measure is constructed. In empirically derived indices, inter-variable correlations and design variations lead to variable selection and weighting schemes which may not be understandable[44], especially by those who are not familiar with statistical techniques[79]. This means that the judgemental measures may seem more credible and would likely be more acceptable by a broader number of researchers. A review of face validity of trauma severity indices created by different methods suggests that those developed through a panel of experts sound more reasonable to physicians. The factors seem to be important, the weights make sense, and the additive approach to combining scores looks reasonable[21].

e) comprehensiveness and content validity

The judgemental method identifies factors through consultation with a number of knowledgeable experts. In other methods the choice of the variables to be collected or the final selection of the factors to be included is also subjective, but the principal researchers themselves identify the important factors. Reliance on a panel of experts and use of a systematic approach increase the likelihood of identifying a broad, but mutually exclusive set of factors[44,77].

f) accuracy or criterion validity

A number of studies suggest that these subjective measures are at least as accurate as measures developed with other methods. For example, an index for diagnosis of thyroid disease, based on Bayes' model with likelihood ratios estimated by a panel, performed similarly to a model based on 1,328 abstracts[83]. Results of another study using a severity index for trauma developed using a MAU model were similar to those obtained from an empirically derived score and two others developed with an ad hoc approach[21]. A third comparative study showed that APACHE II had a prognostic accuracy for acute pancreatitis patients similar to empirical scoring systems[107]. On the other hand, another work was unable to demonstrate any significant difference between APACHE II predictions for patients admitted to ICUs and physicians' or nurses' clinical assessment[20], showing that the scoring system has little to add to the subjective decision making process.

g) feasibility

Judgemental summary measures are cheaper and faster to construct with respect to empirically developed instruments[23]. Subjective methods cut the data requirements because data are not needed for the development of the model. Any data collected can be used to test the model[44]. However, there are costs and time required for the judges' meetings, which increase

if a large panel is selected. Gustafson[83] compared cost and time for developing a model based on judgements of four experts and a model derived statistically from retrospective analysis of 879 medical records showing similar accuracy. The costs to obtain the subjective estimates were $64 in reimbursement for two hours time per person (2 hours x 4 persons x $8 per hour), while costs to extract information from the records were over $1,400 (20 minutes per record x 879 records x $5 per hours). The required time could be measured in hours instead of months.

Summary

The judgemental method requires experts (hence prior knowledge and background data and information), while a dependent variable is not necessary and data collection is needed only for validation.

In the context of risk indices, the judgemental approach is appreciated for its modifiability and generalizability and because a wider range of investigators can acknowledge the easily understandable face and content validity. Feasibility depends largely on the size of the panel and on the cost and the number of required meetings.

Opponents point out the problems related to the potential for subjective bias. In addition, we have to take into account that when little is known about the phenomenon, there are no "experts" for employing this method.

3.5. PSYCHOMETRIC METHOD

3.5.1. Introduction

Definition

The psychometric method is based on a theoretical hypothesis about the construct to be measured, together with the empirical analysis of the psychometric properties of the preliminary measure[16,109]. The development process is built on the subjective judgement of the researcher,

but the initial instrument is submitted to a pre-test which helps to revise the summary measure and correct it for possible subjective bias. Psychometrics originated from the field of psychophysics[110] to measure, using subjective judgement, qualities that cannot be objectively or physically measured, such as psychological traits, attitudes, achievement, and subjective states[17].

A construct could be viewed as the attribute or the dimension to be measured[5,111]. Individual factors (often assessed through questions or 'items') are assembled to obtain a global measure. The classic theory of reliability assumes that a measure is a combination of the true score and an error score[25]. The reduction of the random error, then, increases the psychometric properties of the instrument (i.e., the precision of the instrument to measure the construct of interest).

The psychometric approach has recently been applied in developing summary measures of behaviours related to exposure to health risk, and the following sections will focus on this use. Usually, the purpose of measures developed with the psychometric method is to discriminate people along a continuum, although evaluative as well as predictive measures can be developed. The method does not require a dependent variable. The theoretical approach used in the psychometric method, to identify factors which are relevant to a hypothesis about the construct, may help explain the exposure to the risk, beyond simply describing it[17].

Description

Since summary measures that focus on generic behaviours account for only a small part of the variance in injuries of children, Speltz and colleagues hypothesized that a better measure could be obtained by considering specific behaviours related to injuries. "These risky behaviours do not always lead to injury in children, but increase the probability of exposure to hazardous materials or situations (e.g. climbing/jumping off the furniture, leaving the house without

permission, playing with sharp objects)"[112].

Several factors (behaviours) were identified from the literature and through consultation with experts. The answers to specific items to assess behaviours were scaled on a 5-point scale, ranging from 0 to 4, according to the frequency with which the child had engaged in the activity in the previous 6 months. The Injury Behavior Checklist global score is the total sum of 24 selected items. The instrument was pre-tested on a sample of children's parents, and their scores were checked for the form of distribution, internal consistency, and discriminant power. The results were considered satisfactory, and no further revision has been made. The instrument could be employed to discriminate children at low/moderate and high risk of subsequent injuries. The preliminary results have demonstrated that specific risky behaviours can explain subsequent injury better than the generic problem behaviours[112].

3.5.2. Development strategy

Theoretical construction

A new measure should spring from a hypothesis regarding the relationship between behaviours and risk, from a theory regarding the construct. The hypothesis suggests a number of factors which might be used to measure the general construct "risky behaviour"[16,113]. The developer proceeds to the creation of questions to assess each factor. Items with face and content validity may be selected from relevant literature, clinical observations, research findings, expert opinions, and existing measures. Each item should be associated to an appropriate scale, according to the kind of possible responses (categorical, continuous) and the desired level of measurement (nominal, ordinal, interval, ratio).

In order to combine the individual factors, several psychometric models have been proposed, but the most frequently used is the linear or summative model. It assumes that the sum of item scores has an approximately linear relationship with the construct in question[5,113].

Empirical evaluation

At this stage of measure development, a preliminary instrument is available for empirical evaluation. It should be administered to a sample of people, representative of the group for whom the instrument is intended, and the resulting responses are studied to evaluate the psychometric properties of the summary measure, and to revise it when necessary[5,16,114].

There are many approaches to item analysis and subsequent item reduction in the history of psychometrics[115]. In the 'rational method', items are selected according to their face validity through subjective judgement and knowledge, with no empirical analysis. The 'criterion method' chooses items discriminating two or more extreme groups, with little regard for their face and content validity. 'Homogeneity approach' focuses on the statistical relationships between items. Jackson found that the optimal approach is to use a combination of the three basic methods ('integrated approach') because it provides a measure with the highest validity in a wide variety of contexts[16,109,113].

Different aspects should be considered when using the integrated approach. A first screening of the items could be made by a panel of experts, before the empirical evaluation. For the Stress Assessment Inventory (a measure of behavioural and cognitive factors which reduce the risk of stress and subsequent psychological and physical illness) the initial 1,000 items were reduced to 300 by three health professionals according to their face and content validity, lack of clarity, and probable bias[116]. Once the instrument is administered to a sample of people, items might be checked with respect to their frequency of endorsement: an item answered in the same direction by most of the people is not useful and does not improve the measure's psychometric properties[25].

The linear model assumes that each item adds something to the others when they share a common construct. This implies that the empirical evaluation of the items should determine the degree of homogeneity or 'internal consistency' of the measure. Each item is related to the

total score through measures such as Pearson product-moment or point-biserial correlation, and those items which correlate most highly are considered the best. They have more variance related to the common factor, and they add more to the test reliability. The items included in the instruments are then considered in terms of correlation with each other, and the Kuder-Richardson formula 20 or Cronbach coefficient alpha are computed[5]. An example is the HIV Infection Prevention Scale developed for assessing self-reported sexual behaviours among heterosexuals. Sexual practices with a known or assumed risk for contracting HIV were chosen based on a review of the literature. Thirty five items were written as clearly as possible (including popular slang for the practices) for a large public. The initial instrument was administered to 150 students, and items that did not receive the full range of possible responses, that did not contribute to the reliability of the measure, or that did not reach the criterion level of 0.25 item-total correlation were eliminated[117].

Since the common purpose of measures developed through this method is to discriminate people along a continuum, the degree of homogeneity has to be balanced with the discriminatory power of the instrument. This may be at the expense of lowering the homogeneity among items[25]. For example, Speltz identified discriminating items by testing the statistical significance of each item in distinguishing between children at high, moderate, and low liability of injuries[112]. On the other hand, when a measure is intended to evaluate changes in the construct of interest, the most important factors will be the modifiable ones, and the relevant items will be those that are sensitive to change[13]. The empirical pre-test may lead to deletion of some items or to an iterative process of revision, including item generation, item analysis, and item reduction.

The described item analysis does not take into account the possibility of giving different weights to each item. This can be based on researcher judgement or on item-total correlation results. Occasionally, a different form of weights is adopted. Instead of weighting the items,

scores are associated to different possible answers to an item (e.g., 5-point scale from 0 to 4 assessing the frequency of a behaviour), and these scores are summed to achieve the global score. More frequent behaviours are weighted more than others[7,112]. However, summary measures developed with psychometric method seldom use weights, and it has been suggested that using differential weights makes little difference[5].

Multi-attribute measure

Sometimes, the construct of interest has more than one attribute, which are measured by different sub-measures. In this case there is a general score and specific scores for each sub-measures. The HIV Risk-taking Behaviour Scale (HRBS), for example, was constructed so that separate sub-scale scores for injecting drug use and sexual behaviour, two common modes of transmission of HIV relevant to the heterosexual population, could be derived[118].

The item analysis can be performed as an extension of the total-item procedure; items which correlate more highly on scales other than the one they belong to are eliminated. Factor analysis can be also employed. This technique identifies clusters of related variables. Ideally there should be one cluster for each sub-measure, and items which load more highly on the wrong sub-measure or on two or more sub-measures are discarded because they are measuring something other than what the developer intended[5,25].

The Needle Sharing Inventory is a summary measure for behaviours which expose an individual to the risk of HIV infection, through drug use and needle sharing. The developers identified several factors, and each was assessed by several items. When the preliminary instrument was administered to a sample of patients, factor analysis was employed to explore the internal structure of the measure. Items that did not load onto the intended factors and did not provide important information were discarded[7].

3.5.3. Validation Strategies

Face validity

In psychometrically derived measures, face validity is involved in each step of item generation, analysis, and selection. After a measure is constructed, face validity is also assessed by experts other than those who developed it. The judgement on the extent to which the instrument looks to assess the construct of interest involves assessment of both the individual items and the measure as a whole.

Content validity

Like the previous form of validity, the developer has to ensure content validity in the procedure of construction[5]. Four principles have been proposed for item creation and item analysis: "content saturation (items should be clearly related to construct, relevant, reflects positive and negative examples of construct, and exemplify behaviour that occurs with sufficient frequency), convergent and discriminant properties (items should not overlap with different constructs, and experts should be able to sort items into groups based on the construct definitions), item wording (items should be clear and unambiguous, concise and simple, understandable, and test only one behaviour), and response style (items should be worded so that most respondents will not give the same answer, so that they are free from extreme levels of social desirability, and so that they don't ask for very sensitive information)"[16]. Once the development process is finished, experts may be asked to judge how well the sample items of an instrument represent the construct.

Content validity also refers to the internal structure of the measure. Factor analysis is a way of assessing the homogeneity of a measure intended to assess a single-attribute construct. When the construct includes several components, factor analysis reveals the structure of the whole measure and shows how closely the observed pattern corresponds to the hypothesized

factors[5,7,116].

Criterion validity

Criterion validity is an issue when the purpose of a measure is to predict an outcome. However, for measures that assess behaviours which represent exposure to a health risk, a criterion or accepted gold standard does not exist. In this case, demonstrating that instrument scores are related to an outcome would just give us an indication of validity. This should more properly be referred to as evidence for construct validation.

Construct validity

Construct validation is particularly appropriate for measures developed through psychometric methods, usually discriminative measures assessing attributes for which there is no adequate criterion. This form of validity "was introduced in order to specify types of research required in developing tests for which the conventional views on validation are inappropriate"[111]. The process of gathering construct-related evidence for validity begins within the measure's development through specifying the construct of interest[16]; hypotheses are then formulated to validate the construct, and evidence from several procedures are assembled.

The relationship of the Needle Sharing Inventory with another method (direct self-reporting) to assess the same construct has supported its construct validity[7]. Convergent validity could also concern the relationship of a measure to other methods or instruments for measuring the same attribute. Discriminant validity of the Injury Behavior Checklist was evaluated by examining the mean scores of extreme groups for injury liability[112].

3.5.4. Discussion of the psychometric method

Strengths and weaknesses of the psychometric method

a) potential for subjective bias

In the development of measures through the psychometric method, subjective bias may be introduced in the construction of the preliminary instrument which is entirely based on the researcher's knowledge and judgement. When the initial instrument is submitted to empirical evaluation and adequately revised, this may correct some of the potential bias.

b) modifiability and adaptability

The theory underlying the construct may evolve as new knowledge becomes available. New or modified theories may lead to the inclusion of new factors in an existing measure or to substantial revision of an instrument. The theoretical development may be easily modified, although the amended index should undergo a new empirical pre-test if changes are considerable.

c) generalizability

Psychometric derived measures have a high degree of generalizability because they are based on a general theory about the construct to be measured. Concerns may arise from the characteristics of the group on which the measure has been tested. If the sample is representative of a large population, the measure will have a wider applicability.

d) credibility, acceptability or face validity

The simplicity of the linear model, the identification of factors based on a theoretical framework, and the selection of items based also on their face validity, all increase the credibility of a measure. Since the psychometric method has just begun to be used in risk

assessment and health-related fields, its acceptability may be greater among people with psychological or social science backgrounds.

e) comprehensiveness and content validity

Content validity is a major concern during instrument development through psychometric method. In these measures careful attention is paid to clearly defining the construct, describing the content domain, and selecting the content sample. Screening of the items, with regard to their comprehensiveness and relevance, is often conducted by experts[16]. The development process emphasizes the content validity of the measure.

f) accuracy or criterion validity

The main purpose of measures developed through psychometric method is to describe risky behaviours and to discriminate people along the continuum of risk of exposure due to their behaviours. A criterion for assessing the measure's accuracy rarely exists; therefore, the validation process consists of gaining evidence for construct validity.

A comparative study of personality inventories (developed through psychometric method), validated against some existing tests, revealed that these measures were significantly more valid than those developed with a strategy based only on empirical analysis of various psychometric properties of the items or with a strategy based on random selection of the items[115]. These findings support the validity of measures developed with the psychometric method and point out, beyond the empirical evaluation, the importance of the theoretical construction.

g) feasibility

The theoretical development of the preliminary instrument is quick and inexpensive,

since it is based on the researcher's work. Subsequent evaluation requires that the measure be administered to a sample of representative individuals, and this requires time and resources. However, the sample for initial revision does not need to be as large as that for the development of empirically derived measures.

Summary

The psychometric method is especially appropriate for developing discriminative summary measures of phenomena which cannot be objectively assessed. The development process includes a theoretical construction of the instrument and an empirical evaluation.

The utilisation of a general theory provides wide generalizability to the measure and increases the confidence in its content and face validity, while the empirical analysis may correct possible subjective bias. On the other hand, the extension of the empirical analysis and subsequent revision influences the feasibility and modifiability of a measure.

The difficulty in assessing criterion validity, because of the absence of a criterion or gold standard, and the need to rely substantially on construct validity often complicates the judgement on accuracy and the comparison with measures developed through other methods.

3.6. MATHEMATICAL MODELLING METHOD

3.6.1. Introduction

Summary measures developed with mathematical modelling method can be used to discriminate or to make predictions, although their value lies in their contribution to a better understanding of the phenomena under study. Modelling efforts are useful per se in organizing information, in stimulating explicit statements of assumptions, and in providing guidance on additional data needed from epidemiological and behavioural studies to improve understanding[119,120]. Mathematical simulations can be used to perform theoretical experiments

with different models and parameters, when epidemiological data are inadequate and knowledge is insufficient, or in case of time, monetary, or ethical constraints. These simulations can guide our thinking and be used for generating hypotheses[121-123].

Mathematical modelling is a wide and complex field, applied in epidemiology for a variety of purposes. This section will focus on a particular use: the modelling of sexually transmitted diseases. Mathematical modelling has been employed in this area mainly to predict the spread of epidemics. However, our interest is in its application in the development of single equations for estimating individual risk of infection.

Definition

The mathematical approach is a method that relies on an abstract "representation of ... a relationship in mathematical form in which equations are used to simulate the behaviour of the phenomenon under study. A mathematical model usually consist of two parts: the mathematical structure itself and the particular constant or parameters associated with it"[30].

The role of the researcher in developing summary measures through mathematical modelling method is substantial. Based on the available information about the phenomenon (e.g. how the disease is transmitted), components that contribute to the phenomenon under study are identified, and a mathematical model that can best represent their relationship is developed. The assumptions required are stated, and the best estimates are chosen, among those available, for the parameters which are to be included in the model. All decisions are made a priori according to the knowledge of the researcher. There is no further support, from data or experts, for selecting variables or weighting them, although validation of the model could help in refining it.

Description

Mathematical modelling for estimating the risk of infection for gonorrhoea has progressed in relation to the knowledge available about the transmission of this disease. Until the late 1970s, research focused on the number of partners with whom a person had unprotected sexual intercourse[124-126]. Unlike previous studies on the risk of transmission of gonorrhoea, Hooper examined risk of transmission in relation to the number of exposures as well as the number of partners. He assumed that the relationship between the number of partners, the number of contacts with each partner, the per-contact probability of transmission, and the probability of a partner being infected could be described by a binomial model[127]. This model was used to estimate the number of infected men and validated comparing them with the observed infected men. The hypothesis of a significant relationship between the risk of transmission of gonorrhoea and the frequency or the number of sexual contacts was tested. This raised new questions about the influence of factors such as variability in virulence of the organism, in host susceptibility, and in environmental conditions.

3.6.2 Development Strategies

The following models can be used in studying risk for any STD, or indeed for any disease transmitted by enumerable activities, although comments and specific equations are focused on HIV infection. This section illustrates different models and discusses the selection of the components to be included and the required assumptions. Estimates for the parameters associated to the model are not presented. These are chosen by the developer among those available from studies designed specifically for this purpose.

Mathematical models, in whatever form, provide an equation for calculating the risk of exposure to HIV or the risk of HIV infection. Theoretically, the equation can also be used

to develop other types of summary measures (e.g. classification, checklist, score), although this is rarely the case.

Binomial models

The following is a description of one family of mathematical models created to estimate an individual's probability of being infected by a sexually transmitted agent, as a function of parameters describing his/her sexual behaviour during a determined period[121]. Derivation of these models is based on elementary principles of probability: binomial expansion and conditional probabilities[121]. A binomial model assumes that each independent observation can only have two possible outcomes (being infected or not being infected).

a) Basic models

A rudimentary model to estimate the probability of an uninfected person acquiring an infection through a single contact with a partner chosen at random [p(inf)=bP] involves at least two components: the probability of transmission per sexual act (b) and the prevalence of infection in the population (P)[128]. Since biologic models of infectivity suggest that the number of contacts is associated with the probability of transmission[129], the model should be expanded to consider the (n) contacts. The model becomes $p(inf)=P[1-(1-b)^n]$, where (1-b) is the probability of not being infected after one contact with an infected partner; elevated to the power of n, it becomes the probability of remaining uninfected after n contacts with the same partner. The factor in brackets represents the probability of being infected under the same conditions. The probability of the partner actually being infectious is P, and the product of the two probabilities is the probability of being infected when the partner's status is unknown[121,128].

A slightly different model assumes that the infectiousness of a partner varies, and that the partner cannot be considered as infectious or non infectious for all contacts:

$p(inf)=1-(1-bP)^{n}$[130]. This requirement can be met for diseases (e.g., gonorrhoea) that have a high cure rate, a high reinfection rate, and for which the risk of infection is modelled over a long period compared to the average time between contacts[127].

When considering several partners, the model[121,127,128] to compute the individual probability [p(inf)] of being infected for a given number of contacts by partners of unknown status is $p(inf)=1-\{1-P[1-(1-b)^{n}]\}^{m}$ where: (P) is the proportion of individuals who are infected in the population from which the individual selects his or her sexual partner (assumed constant); (b) denotes the infectivity per sexual contact, i.e. the conditional probability of infection after one sexual contact given an infected partner (assumed constant); (n) is the number of contacts with each partner (assumed to be the same for all partners); (m) represents the total number of partners during the period.

The model treats partners and contacts as trials in a binomial experiment, which requires that these be considered as independent events. Another assumption stems from having to multiply together the probability of being infected $[1-(1-b)]^{n}$ and the probability of any partner being infectious. This is a valid operation only if partners are selected independently of whether they are infectious or not.

b) More complex forms

In the basic model the infectivity per contact (b) and the prevalence of infection in the population of partners (P) are assumed to remain constant during the period of time; a partner's infectious status remains the same for all contacts with the considered individual, and the number of contacts (n) is the same for all partners. More complex forms of the same model allow to be taken into account different infectivities (b_j) for each type of sexual contact (e.g., oral, genital, anal): $p(inf)=1-\{1-P[1-\Pi_j(1-b_j)^{nj}]\}^{m}$[121]; subgroups with different prevalences (P_i) of infectiousness: $p(inf)=1-\Pi_i\{1-P_i[1-(1-b_i)^{ni}]\}^{mi}$[121]; and different numbers

of contacts (n) per each sexual partner: $p(inf)=1-\prod_{j=1}^{m}\{1-P[1-(1-b)^{nj}]\}$[131]. These modifications can be combined into one model.

Poisson models

For focusing on the probability of infection over a given time period, instead of considering the probability of infection for a given number of contacts, a Poisson model has been proposed[131]. The number of contacts is considered the random variable following a Poisson distribution with expectation *lt*, where (t) is the time period and (l) represents the expected number of contacts per unit time.

The probability of becoming infected over a period of time (t) becomes $p(inf)=1-(1-P)+Pexp^{-blt}$ for one randomly selected partner and $p(inf)=1-exp^{-bPlt}$ for different partners, where (P) is the probability that a randomly selected partner is infected, which is assumed constant, and (b) represents the probability of infection per contact, which is also considered constant. The model requires that the number of contacts per unit time be the same throughout the entire time interval and that the contacts be distributed randomly over the time and be independent events.

Other models

When the probability is estimated over a long period of time, more sophisticated model should be used because, as time progresses, the prevalence of HIV is not constant. A person selecting multiple partners from the same pool will be choosing, over time, partners who are more likely to have been infected. Assuming that the prevalence grows exponentially, that one new partner is selected in each unit of time, and that there is one contact per partner, Sandberg has proposed the following model $p(inf)=1-(1-bP)(1-bP\ exp^{t1})(1-bPexp^{t2})...(1-bPexp^{tm-1})$, where (b) is constant, but (P) is allowed to grow exponentially with time[132].

The more complex models overcome some of the problems related to the simplicity of the basic forms. However, some studies on HIV transmission, indicating biological mechanisms which bring about variations in susceptibility and infectiousness, suggest that the assumption of a constant and homogeneous probability per contact (b) must be abandoned in favour of more flexible formulations, able to allow for variability in transmission rates[133,134].

Other studies[133,135,136] found that the probability of transmitting HIV in a partnership appeared to be unrelated to the number of sexual contacts. Thus, rather than assuming that the probability of transmission increases with the number of sexual acts in a relationship as in most models, some authors consider the probability of transmission to be fixed at some average value per partnership of a given kind[136-138]. As the knowledge on the underlying biological mechanisms in the transmission of HIV progresses, more sophisticated models have been developed to include the new acquisitions.

3.6.3. Validation Strategies

Face validity

Determining the suitability of a risk estimate, developed through mathematical modelling method should involve both the mathematical structure and the associated parameters selected by the developer. This means evaluating, in the light of the current knowledge on the disease, whether the model can simulate the phenomenon, whether the assumptions stated are plausible (e.g. independence of the events or assuming that the prevalence or the infectivity are constant), and whether the value chosen for the parameters are the most realistic among those available.

Content validity

Looking at content validity constitutes judging whether the developer has included all

the relevant or important aspects for predicting the risk. Experts should assess the importance and comprehensiveness of the components incorporated in the risk equation, based on information gained from other studies on the biological mechanisms underlying the phenomenon.

Criterion validity

The predictive ability of the mathematical model should be tested against an external criterion, usually an empirical data set. The risk estimates produced by fitting the model into the data are compared with the observed risk. When the expected values obtained by the model correspond to the actual risks, the model is appropriate, and a better understanding of the phenomenon is achieved. When deductions and observations do not correspond, the model will require revision[121,133,136].

Construct validity

When the measure assesses a risk which does not have a gold standard or a direct measurable outcome, such as the risk of exposure that may or may not result in infection, the convergent and divergent validity with other measures can still be tested. Blower[122] has proposed exploring the behaviour of a model by generating particular scenarios through computer simulations and then comparing the results with other studies.

In absence of a gold standard, a mathematical model intended to discriminate among different categories (e.g., no risk, low and high risk) may be applied to data from extreme groups, instead of individuals, which form a continuum. Clusters with significant differences in risk estimates provide support for the model.

3.6.4. Discussion of the mathematical modelling method

Factors influencing the choice of the model

All the models presented for providing a summary measure, and the many others that can be generated, share the features that are characteristic of the mathematical approach: the a-priori choice made by the developer that the phenomenon can be simulated by that equation. The decision about which form is to be used should be a balance between the complexity of the model and its ability to better represent the behaviour of the risk. More complex models require more assumptions, and their adherence to reality is difficult to test. On the other hand, as the knowledge about an event progresses, models need to become more complex, so that they can incorporate the new information. Moreover, when a large number of parameters is included (e.g., prevalence, infectivity, condom protection, susceptibility), or when variability for the parameters is allowed (e.g., different prevalences for several groups), the developer should review other studies for acquiring appropriate estimates for each of them, which is not an easy task.

Strengths and weaknesses of the mathematical modelling method

a) potential for subjective bias

The mathematical method is strongly based on decisions made by the developer. Although choices are based on current knowledge on the phenomenon, two researchers may propose dissimilar equations by interpreting differently the same available information. Bias could be introduced in each step: the choice of the components, the decision as to the form of the model, and the evaluation of the best estimates for the parameters.

b) modifiability and adaptability

Since the development does not require collection of data or meeting with experts, a

model can be easily modified to suit new acquisitions or be expanded to incorporate other components. Even the estimates for the parameters can be easily substituted with values more appropriate for another population or with more precise figures produced by new studies.

c) generalizability

Measures developed through a mathematical method are based on a theoretical model for describing the phenomenon. They do not arise from a specific sample of people. The same model can be applied to different populations unless biological mechanisms are different; for example, an increased susceptibility may alter the relative importance of the number of partners and the number of contacts[127].

d) credibility, acceptability or face validity

Doubts as to whether the phenomenon can be predicted by the proposed model could detract credibility from the measure. In addition, discussions on the possible violation of the assumptions (e.g., independence of events) by the nature of what is known about the condition or about the over-simplification of certain hypotheses (e.g., a parameter fixed at a constant value) are rarely conclusive. Acceptability is likely to vary between researchers who are familiar with mathematical models and those who are not.

e) comprehensiveness or content validity

As discussed in the section on factors influencing the choice of the model, the complexity of a model has several drawbacks. In general, the comprehensiveness of the components included is limited by the model complexity.

f) accuracy or criterion validity

Mathematical models attempt to describe the risk by trying to reproduce the actual mechanisms of transmission in mathematical form. They are often used to perform simulation studies when epidemiological data are scarce and the transmission process is not completely understood. In these conditions, the evaluation of the accuracy of the model or the validation of the assumptions could be delayed until sufficient data become available. However, it has been proposed[122,139] that predictions produced by a model can be investigated through sensitivity analysis, which describes changes in results if one systematically alters the assumptions. When little is known, sensitivity analysis provides a range of plausible probabilities, assesses the variability in the predictions, determines confidence intervals of the estimates, and evaluates which are the key variables for the phenomenon[122].

g) feasibility

Feasibility is regarded as a strong point of the mathematical methods. The development process is quick and inexpensive, since it does not require collection of data, the meeting of experts, or empirical pre-tests.. However, some of the forms of validity (criterion) still require time and money to obtain empirical data when no adequate data-sets are available.

Summary

Mathematical modelling method requires a baseline knowledge about the disease, yet it does not need a dependent variable or collected data. The few requirements in terms of money and time for the development process, the simple modifiability of the produced measure, and its generalizability contribute to the agility of the method, which is greater in comparison to that of other methods.

Weak points are represented by the possible subjective bias introduced, the poor

credibility of the results, and the small number of variables included. These weaknesses results in difficulties with respect to accuracy of the estimates, which are often considered in terms of range provided. Mathematical method may be seen as a foundation for the process of developing adequate summary measures since it helps in explaining the phenomenon, in creating hypotheses, and in stimulating explicit statements of assumptions.

3.7. DISCUSSION OF METHODS FOR RISK ASSESSMENT

The proposed classification identified four methods of risk assessment: 1) the empirical, which acquires information from collected data; 2) the judgemental, which relies on the opinion of experts; 3) the psychometric, based on theoretical hypothesis about the construct; and 4) the mathematical, which uses a probability model to simulate the phenomenon.

3.7.1. Similarities and differences among the methods

The *mathematical and empirical methods* may appear to be similar in the way they use mathematical equations. For both methods, the investigator chooses the functional model that best represents the phenomenon of interest, under certain assumptions. The developer who uses a mathematical method goes further and decides, based on the available knowledge, which are the components of the phenomenon and the best values for the parameters included in the model. By contrast, the empirical method is mainly based on data. The model is used for screening the collected variables to evaluate which one best correlates with the outcome of interest, for determining the importance of each variable by noting its impact on the statistical variance of the data, and for aggregating the components in a summary estimation of risk.

The *psychometric method* shares its theoretical construction with the *mathematical method* and its empirical evaluation with the *empirical method*. However, the philosophy behind the psychometric method is quite different. Instead of looking for complex models that best

describe the relationship among the components or that best fit the data at hand, it focuses on the selection of items and assumes a simple additive model, where each item adds something to the others because they are chosen to share a common construct.

There is an analogy also between *psychometric and judgemental methods* because both rely on subjective judgement. In spite of this resemblance, the two methods adopt different strategies in correcting possible bias. Psychometrically developed measures are submitted to an empirical pre-test (i.e., administered to a sample of representative people), which helps the investigator to revise the instrument. In the judgemental method, decisions are made by a panel of experts and not by a single investigator. Moreover, several methods for selecting the panel, conducting the discussion, and obtaining the estimations are applied to improve the resulting measure.

3.7.2. Strengths and weaknesses of the methods

The description of each method highlights its strengths and weaknesses, evaluated according to the seven criteria chosen.

The main strengths of the *empirical method* are the reduction of subjective bias in the development process and the accuracy of the results, although the fitting of the chosen model to the data should be carefully evaluated. Generalizability, credibility, and comprehensiveness strongly depend on the representativeness of the population, the quality of the variables collected, and the user's familiarity with statistical techniques. Difficulties include the scarce adaptability of the resulting instrument to different populations and the high cost in terms of financial resources and time required for development.

The choice of variables made by judges may be more comprehensive and understandable than that for measures developed with an empirical method; the index may also seem more credible and be more readily accepted by other investigators. Moreover, the use of a panel of

experts in the *judgemental method* increases the generalizability of the resulting measure, provided that the expertise of the judges is wide enough and not limited to a specific population. Several studies show a comparable accuracy with the empirically developed scoring system. Cost and time are related to the number of panel meetings required for the development and modification of an index. However, the main problem of the judgemental method is its potential for subjective bias, which depends on the panel and which can be in an unknown direction.

The utilisation of a theoretical hypothesis about the construct allows more confidence to be placed in the generalizability, the content validity, and the credibility of measures developed with the *psychometric method,* while some correction for possible bias, due to subjective judgement, is provided by the empirical analysis of the instrument and its subsequent revision. The feasibility of the method and the modification of a measure for different applications are limited by the necessity of pre-testing the instrument with a sample of population. Moreover, the frequent absence of an external criterion yields uncertainty in assessing the accuracy of the results.

Lastly, the *mathematical method* involves a quick and inexpensive development process. The resulting measures are quite generalizable and are easily modifiable by changing the parameters. Nonetheless, weaknesses can be identified in the potential for bias introduced by the developer, in the limited number of variables included in the model, and in the possible scarce acceptability among researchers not used to mathematical models. In addition, the estimates are often considered in terms of the range provided, and not for their accuracy.

This comparison of the methods cannot be conclusive, because, beyond the reduction of subjective bias, there may be different underlying considerations and points of view, and the importance given to each criteria could vary according to the situation. For example, the credibility may be of great value when the index is employed by clinicians not accustomed to summary measures; yet, it may be irrelevant when employed by the developers themselves.

Accuracy is more important when the risk assessment is designed as a diagnostic test for a preventable or treatable condition than when it is used solely for screening. Again, an index developed to describe or explain a risk should include all the factors known to influence the phenomena, while an instrument designed just for making predictions could include only those factors that best correlate with the outcome. Generalizability and modifiability depend on the use of the instrument and influence the range of its applicability. Finally, feasibility is largely based on available resources and time.

The following table (table 7) is an attempt to summarize the above described strengths and weaknesses of each method. Because of the scarce amount of research that has investigated this subject and the lack of studies that have compared the methods for each of the identified criteria, this table should be taken just as an indication of the intrinsic differences among the methods. Moreover, the study design, the data collection, the experts' selection, and the choice of the population sample (in other words, the quality of research) could completely change this interpretation of strengths and weaknesses. Three plus signs signify good performance with respect to the criteria, two signs indicate a medium or problematic performance, and while one plus signifies poor performance.

Table 7. Comparison of methods

Criteria	empirical	judgemental	psychometric	mathematical
reduction of bias	+++	+	++	+
modifiability	+	++	++	+++
generalizability	++	++	++	+++
credibility	++	+++	+++	+
comprehensiveness	++	+++	+++	+
accuracy	+++	+++	++	+
feasibility	+	++	++	+++

When considering reduction of bias, the empirical approach is the superior method,

whereas overall, the mathematical method is the most practical because of its high feasibility, modifiability, and generalizability. On the other hand, judgemental and psychometric methods are appreciated for their content and face validity by a larger number of researchers.

3.7.3. Factors influencing the choice of methods

Beyond the intrinsic strengths and weaknesses, other factors should be considered when choosing among methods for risk assessment. The existence of a direct measurable outcome, used as a dependent variable, is required by the empirical method, which employs regression models to select and weight the variables, while this is not a condition for the others. Measures developed with the empirical method have a more practical use, yet they are less helpful in explaining the phenomenon studied. Using statistical techniques to select variables and their weights may lead to a measure that is highly predictive of the outcome, but that is often not very instructive with respect to the underlying mechanisms. The other approaches, which have a theoretical basis, choose the components and evaluate their importance in light of a hypothesis on the object of study. The subsequent use of these measures may help in testing the hypothesis and in advancing the understanding of the phenomenon, as opposed to simply describing it.

A further factor should be considered when developing a measure of risk, especially for a new disease: the stage of the epidemic. If the number of cases is restricted because of the early stage of diffusion of the disease, the use of the psychometric, and especially empirical method, is limited by the fact that they require data to develop a summary measure of risk. On the other hand, when little is known about the disease, there are no "experts" for the judgemental method. This has been the limiting factor in developing summary measures of risk for HIV infection, as will be discussed in the following chapters.

3.7.4. Risk assessment circle and multi-disciplinary approach

The discussion up to this point has demonstrated that there does not exist a perfect method that is applicable to every phenomenon, but rather that the usefulness of a method changes over time and according to the situation. Risk assessment must be considered as an iterative process involving various methods, where each method contributes, at different times, to the development of the most effective measures of risk. In the early stages of an epidemic, there are not enough data for an empirical approach and not enough knowledge for experts' judgement. The mathematical method may then form a starting point for studying the new phenomena, for suggesting hypotheses, and for stimulating the understanding. When data become available, measures developed with mathematical models can be validated and hypotheses can be tested. The new theories about the phenomenon may be employed in constructing psychometrically derived measures, which can be empirically pre-tested with the data available. As the number of disease cases increases, a simple empirical instrument can be developed, while the knowledge accumulated over time helps judges to improve their expertise and to develop better indexes.

An ideal progression of methods to be used can then be suggested, with the mathematical method being employed first as a foundation of risk assessment, followed by the psychometric method supported by a theoretical hypothesis that is tested through the empirical evaluation and that contributes to the understanding of the particular phenomenon. Finally, the judgemental and empirical methods could be employed to provide measures with higher accuracy. This sequence can be repeated in a continuous cycle, contributing to the scientific understanding of the phenomenon and producing superior instruments for measuring the risk. The assessment of risk for HIV infection is probably reaching the final stage of the first cycle, with an increasing number of summary measures developed through empirical and judgemental methods.

A further comment about the methods regards the trade-off between the different strengths of each method, which investigators have to face in choosing the best approach. The distance between the disciplines, which have developed different methods, has precluded a fruitful integration that could prevent the extremes of either judgement without empirical evidence or blind empiricism that does not consider biological significance.

A multi-disciplinary approach can combine the strengths of each method. For example, a panel of knowledgeable experts may support decisions implied in the empirical approach, such as the choice of the variable to be included in the model other than those selected by statistical techniques, or the appraisal of biologically meaningful cut-off points in interpreting the final score. Moreover, theoretical hypotheses may suggest which variables need to be collected. By contrast, empirical data can be used to provide evidence for experts' decisions about the importance of the components.

CHAPTER 4 - SUMMARY MEASURES FOR RISK OF HIV

A decade after the first cases of an illness subsequently defined as AIDS were reported, more than 400,000 cases worldwide have been reported to the WHO. Since these numbers underestimate the true extent of the epidemic due to the fact that many cases remain unrecognized or unreported, WHO estimates that over one million cases of AIDS have actually occurred. An additional 9 to 11 million individuals are estimated to be infected with the HIV, the etiologic agent of AIDS[2].

HIV-infected individuals are not uniformly distributed worldwide: there are at least six million cases in Africa, one million each in North America, South America, and Asia, and half a million in Europe[3]. These individuals represent the asymptomatic reservoir of HIV infection. By the year 2000, as many as 20 to 30 million persons may be infected with HIV[4].

In major cities throughout the Americas, western Europe, and sub-Saharan Africa, AIDS has become the leading cause of death for both men and women aged 15 to 49 years. As a result of vertical transmission, infant and child mortality rates have increased more than 30% beyond those previously projected[4].

The predominant mode of HIV transmission is sexual intercourse, which is responsible for more than 75% of the cases worldwide. Of these, the large majority - by a factor of 4:1 - were heterosexually transmitted[3]. While transmission of HIV through homosexual contact has been common in developed countries, there is now growing evidence that heterosexual transmission is increasing in Europe, North America, and Latin America. Although the prevalence of HIV among heterosexuals in western countries is still low, this is the largest susceptible group for the spread of infection.

This chapter presents existing instruments in the field of risk assessment for sexual transmission of HIV. Since this infection has appeared very recently in the human population,

at least in epidemic form, this review is able to cover the entire period from the time of its discovery to the present, following the progressive spread of the epidemic in different groups and the related acquisitions of knowledge and empirical data about it.

4.1. LITERATURE SEARCHES FOR SUMMARY MEASURES FOR RISK OF HIV

A strategy similar to the one used in the search for risk assessment methods (reported in chapter 3) was employed to look for indexes for the risk of HIV due to sexual behaviour.

4.1.1. Methodology

Five searches were performed for existing summary measures in the medical (MEDLINE 1983-87 and 1988-1992), psychological (Psych-LIT 1983-91), and sociological literature (Sociological abstract 1963-91 and Family resources 1970-91). The first key concept was identical to that used in the general search. The MESH terms considered were "health status indicators" and "severity of illness index", and words searched in titles or abstracts were "index(es) (not index case)", "indices", "scale(s)", "score(s)", "appraisal(s)", "assessment(s)". The concept was cross referenced with the MESH terms "risk", "risk factors", and "risk-taking". However, two new key concepts were introduced to restrict the field to the HIV/AIDS area ("HIV infections", "HIV", and "acquired immunodeficiency syndrome") and to sexual behaviours ("sexual partners" and "sex behaviour").

To look for risk estimates through statistical or mathematical models, two more MEDLINE searches (one for each period) were conducted, substituting the first concept with the MESH terms "models, theoretical", "models, statistical", "probability", "Bayes' theorem". A further investigation on AIDSLine (1980-1992) was performed for abstracts presented in the last six International Conferences on AIDS (Washington 1987, Stockholm 1988, Montreal 1989, San Francisco 1990, Florence 1991, and Amsterdam 1992). Since abstracts contained too few

information for reviewing an instrument, a specific search by principal author's name was performed for each relevant abstract, in order to find possible subsequent articles (some of the authors were contacted).

4.1.2. Inclusion criteria

The inclusion criteria for papers described in chapter 3 were also applied to this search (i.e., regards some kind of risk, implies a multifactorial assessment, and provides a summary measure). Moreover, articles to be included had to focus on the topic of HIV/AIDS and on sexual transmission. The decision to not include papers related to other STDs was made because of the differences in infectivity and in availability of treatment, which renders the risk assessment models quite different. There were no restrictions for risk groups; however, the main focus had to be on sexual transmission (although measures which combined transmission through sexual contact and needle sharing were included). Instruments assessing knowledge of risk, attitudes towards risk, and perceived risk were excluded because the interest was in the actual transmission. Moreover, abstracts were not included in the review unless additional material was provided by the author or subsequent papers were found.

A total of 51 articles and abstracts (describing 30 summary measures and 6 alternative forms) are reviewed in the following section (the appendix provides a detailed review). Of these, 29 were retrieved with computerized searches (18 from MEDLINE, 2 from Sociological abstract, and 9 from AIDSline), 12 were found through primary references, and 5 via research based on abstracts. The remaining 5 are unpublished material obtained from the authors.

4.2. REVIEW OF THE EXISTING SUMMARY MEASURES FOR RISK OF HIV

This section provides a synopsis of the existing measures for risk of HIV; the complete description of each instrument can be found in the appendix (which also provides a short

presentation of each measure, the list of factors included, the method of administering, and some elements of the development process). In this review, the format includes the underlying method used for the development according to the classification (as proposed in chapter 3) and the characteristics and qualities of each summary measure (as described in chapter 2).

General

- Measure identifier number: number assigned in alphabetical order (except for alternative forms which follow the primary measure with a decimal number)
- Author: the principal author of the first published paper
- Year: year of the first published paper

Development process

- Underlying method: the method for developing the measure according to the proposed classification

Measure characteristics

- Type of measure: form of the summary measure
- Purpose: purpose of the measure in a categorical form
- Operational definition: definition of the risk the measure is intended to assess
- Application: population to which the measure has been applied or is intended for, according to sexual preference and gender

Measure quality

- Reliability: information available on reliability
- Validity: information available on validity

To present the results in a tabular form, these factors have been codified according to the following table.

Table 8. Coding for the variables

- Measure identifier number
- Author
- Year
- Underlying method
  - 5 - empirical
  - 4 - judgemental
  - 3 - psychometric
  - 2 - mathematical
  - 1 - ad hoc

- Type of measure
  - 1 - checklist
  - 2 - classification
  - 3 - score
  - 4 - probability equation

- Purpose
  - 1 - discriminative
  - 2 - evaluative
  - 3 - discriminative + evaluative
  - 4 - predictive

- Operational definition
  - 1 - risk of exposure
  - 2 - risk of HIV infection

- Application: population
  - 1 - general population (homosexuals and heterosexuals)
  - 2 - homosexuals
  - 3 - heterosexuals (males and females)
  - 4 - heterosexual males
  - 5 - heterosexual females

- Reliability
  - 0 - none
  - 1 - inter-observer agreement, intra-observer agreement, test-retest, or internal consistency

- Validity
  - 0 - none
  - 1 - content validity
  - 2 - construct validity
  - 3 - criterion validity

Table 9. Existing measures for risk of HIV in alphabetical order

code	author	year	met	typ	pur	def	pop	rel	val
1	Bassman[117]	1991	3	3	1	1	3	1	0
2	Biglan[140]	1990	3	3	1	1	3	1	0
3	Calzavara[141-2]	1991	1	3	3	1	2	0	2
4	CDC[143-5]	1987	1	3	3	1	2	0	2
5	Crowe[69]	1991	5	4	1	2	2	0	0
6	Darke[118]	1991	3	3	3	1	3	1	1
7	Eisenberg[131]	1989	2	4	4	2	1	0	0
8	Elmslie[100-1]	1991	4	3	1	1	5	0	0
9	Europ mal[70]	1992	5	1	1	2	4	0	3
10	Europ fem[70]	1992	5	1	1	2	5	0	3
11.1	Fineberg[146-7]	1988	2	4	4	2	1	0	2
11.2	Schneider[148]	1991	2	4	4	2	1	0	3
12	Gladis[149-150]	1990	2	2	1	1	1	1	2
13	Hanson[151]	1992	3	3	1	1	3	1	0
14	Hearst[152]	1988	2	4	4	2	3	0	0
15	Hooykaas[153-4]	1989	1	3	3	1	3	0	0
16.1	Iguchi[155]	1991	5	1	1	2	1	0	3
16.2	Iguchi[60]	1991	5	3	1	2	1	0	3
17.1	Joseph[156-8]	1987	1	2	3	1	2	0	0
17.2	Trocki[159]	1990	1	2	1	1	3	0	0
17.3	Windle[160]	1989	1	2	1	1	3	0	0
18	Joseph[161-3]	1990	1	2	3	1	2	0	2
19.1	Kelly[164]	1990	1	3	1	1	2	0	0
19.2	Kelly[165]	1990	1	3	2	1	2	0	0
20	Martin[166]	1986	1	2	2	1	2	1	0
21	McQueen[8,102-4]	1991	4	2	3	1	1	0	2
22	Meyer[167-8]	1991	1	2	1	1	1	0	0
23	Montgomery[169]	1991	1	1	1	2	1	0	0
24	Nyamathi[170]	1992	1	2	1	2	5	0	0
24	Ostrow[171-2]	1990	1	2	3	1	2	1	2
26	Rapkin[173]	1990	1	1	1	2	4	0	0
27	Sandberg[132]	1989	2	4	4	2	1	0	0
28	Shaw[174-5]	1986	1	3	1	1	1	0	0
29.1	Stall[176]	1986	1	3	3	1	2	0	0
29.2	Leigh[177]	1990	1	3	1	1	1	0	0
30	Windle[160]	1989	3	1	1	1	3	1	0

A total of 30 measures were found in the 10 years covered (1983-92), in addition to 6 measures which have been considered as variants of a primary measure. In the first three years (1983-85), no index has been constructed, and only a few summary measures were published between 1986 and 1988 (3 indexes in 1986, 2 in 1987, and 2 in 1988). Since 1989, 29 indexes have been developed (81% of the measures reviewed) with a clear annual increase in the number of indexes (5 indexes in 1989, 9 in 1990, 11 in 1991, and 4 in the first half of 1992*). Although there is always a delay between the construction of an index and the time of publication, the large number of measures found in the last four years shows a growing interest or need for risk assessment instruments in the field of HIV.

4.2.1. Type of measure

The most common type of summary measure among those developed for HIV is the scoring system (14 indexes), while ten measures provide a classification in two or more risk groups. The remaining twelve instruments furnish a simple checklist of risk factors (6 indexes) or an equation to estimate the probability of HIV infection due to sexual behaviours (6 indexes). In some cases a measure provides two kinds of output, such as an equation and a classification[131], or a score and a classification[60].

4.2.2. Purpose

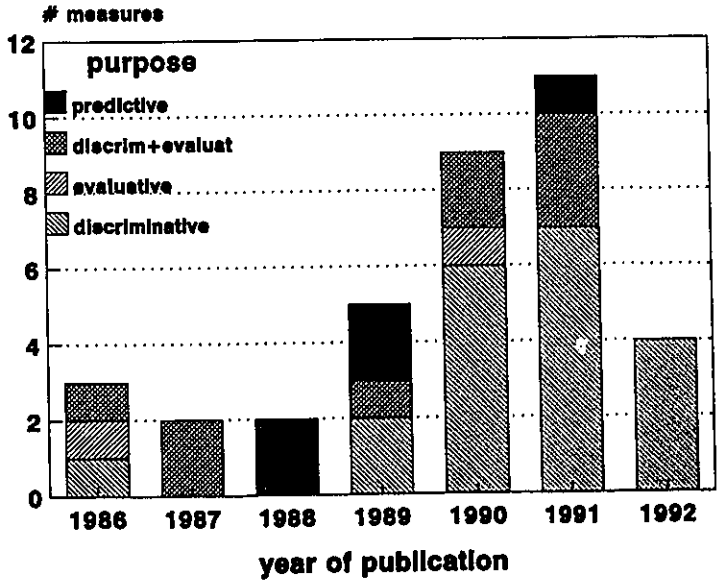
Five measures were created for predicting the future risk of HIV infection. Their intent (reported in the appendix in conversational style, with sentences appropriated from the original papers) was to estimate the individual probability of acquiring the infection. These instruments

---

* The final search on MEDLINE was performed on the October 1992 release of the CD Plus. Although there is no a deadline for inclusion of articles since this widely depends on the frequency of publication of the journal, it can be estimated that the search includes all articles up to June 1992 and some articles from July and August of the same year.

in general have an educational use: the estimation of a future risk may stimulate changes in the risky behaviours. On the other hand, there are 20 indexes classified as discriminative measures. They are not expected to predict the future risk of infection, but to distinguish individuals according to their present level of risk. Some of these indexes (7 measures) have the practical use of identifying the people that are most likely to be positive for HIV. In the other 13 measures, risky behaviours are combined in a summary measure that can help to describe the risk of exposure existing in a population, to study the behaviour predictors, and to determine the association of risky behaviours with other factors, such as knowledge, attitudes, beliefs, drugs and alcohol use, or other behaviours. Two measures were developed with the specific purpose of evaluating longitudinal changes in risky behaviours which expose to risk of HIV and to study the predictors of changes and the relationship between changes and factors such as risk perception or knowledge of HIV status. Since a summary measure can be used cross-sectionally or longitudinally, there are 9 more instruments having both discriminative and evaluative purposes.

Figure 1. Measure purpose by year



The distribution by year (figure 1) shows that most of the measures are produced with a discriminative and evaluative purpose. This indicates that the main interest is still in describing the risk of exposure to HIV, in screening people for HIV infection,

or in evaluating changes in risky behaviours. Since 1988, some measures for predicting the future risk of infection have been developed, but they are still rare.

4.2.3. Conceptual basis and operational definition

The conceptual basis for any operational definition of risk of HIV is the available knowledge about the infection. All the measures reviewed refer to the theory of AIDS being an infectious disease caused by the HIV virus, transmitted by an infected person to a susceptible host. For communicable diseases, the risk and severity of infection is determined not only by the exposure to the infectious agent, but also by the infectivity and virulence of the agent, by the size and dose of the inoculum, by the infectiousness of the infected person, and by the susceptibility of the host[178]. Moreover, the indexes, selected with the focus on sexual transmission, adopt the view that exposure consists of contact with human fluids which are known to carry the HIV virus (semen, cervical secretions, blood, and possibly saliva) during sexual intercourse.

Within this framework created by the conceptual basis, there are two main operational definitions of risk, with regard to sexual transmission of the HIV, that can be identified in the existing measures, although the articles rarely state clearly what the instrument intends to assess. The first refers to the *risk of HIV infection* through sexual transmission. Theoretically, this means that one has to consider all the factors that influence the risk of HIV; however, the 13 measures which use this operational definition include only some of the factors. An interesting aspect of these measures is the inclusion of non modifiable factors or factors that have no known causal relationship with the risk of HIV (e.g., age, use of nitrates, use of enemas, history of giardiasis), but that correlate strongly with the outcome. In this case, the outcome was infection with HIV, which can be assessed with serologic tests for detecting anti-HIV antibodies, considered the gold-standard. The other 23 indexes were developed to assess

the *risk of exposure to HIV* due to sexual behaviour. In this case the summary measure was intended to measure only the sexual behaviours which put a person at risk of being exposed to the virus (e.g., type of practice, frequency of intercourse, condom use, number and type of partners), without regard for the other factors or for the prevalence. In the case of risk of exposure, there is neither a directly measurable outcome nor a gold standard.

The level of study for all the measures is in the actual factors and not in possible predictors, which are not included in any index. The factors considered are mostly causal factors, although some markers are used, such as T4 lymphocyte count for stage of disease and for degree of infectiousness, type of partner for the probability of meeting an infected person, and history of STDs for promiscuous behaviour.

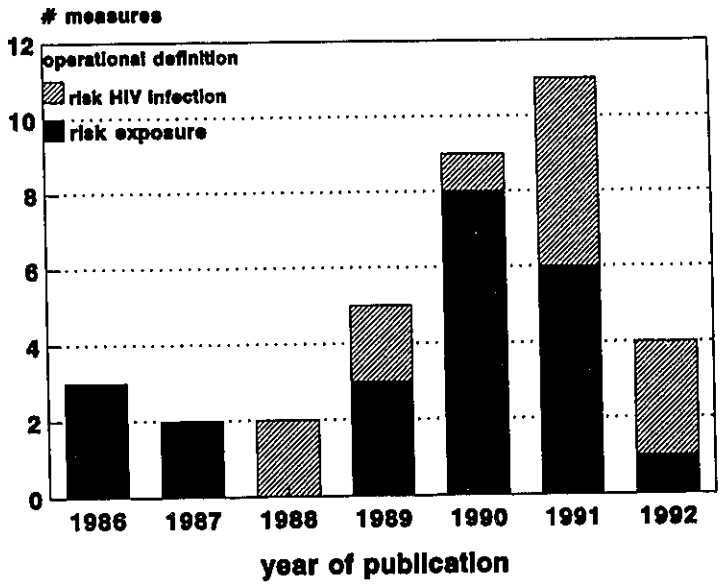


Figure 2. Operational definition by year

The distribution of the indexes by time (figure 2) shows that there is a change of interest from exposure to HIV to the actual risk of having or acquiring the infection. In fact, the measures assessing the risk of HIV infection appear later on and their proportion by year increases with time. Another variation in the operational definition of the indexes regards the population considered. Almost 60% of the measures developed for the general population assess the risk of infection, as opposed to 35% and 10% of the instruments constructed for heterosexuals and homosexuals, respectively.

4.2.4. Application

Some of the reviewed measures are specific for homosexual (10 indexes) or for heterosexual (14 indexes) individuals, while others have been developed for the entire sexually active population (12 indexes), which include both homosexuals and heterosexuals. Since the two genders differ in their behavioural risk factors and related susceptibility, 4 measures are specific for heterosexual females and 1 for heterosexual males. Moreover, some indexes are developed for specific categories (see appendix) such as sexual partners of HIV infected persons (3 measures), for which the type of partner is already determined, or IVDUs (4 measures), who are considered to be at risk for HIV infection also because of needle sharing. It is important to recognize for which population the indexes are intended, because the risk factors included in a summary measure for risk of HIV and their weights differ greatly depending on the sexual preference and the gender, while age, which is an important factor for many diseases, seems to have no relevance.

A further element to be taken into consideration is the prevalence of infection in the specific population. The relative importance of the risk factors varies with the proportion of persons infected. Most of the measures are specific for countries (see appendix) where the major part of the infected individuals are either homosexuals or IVDUs (27 for North America, 4 for Europe, and 1 for Australia), while the rate of infection among the rest of the population is low. In these situations, a factor such as the type of sexual partner assumes great relevance, because of the higher probability of meeting an infected person in certain groups (i.e., homosexuals, bisexuals, and IVDUs), compared to non drug-using heterosexuals.

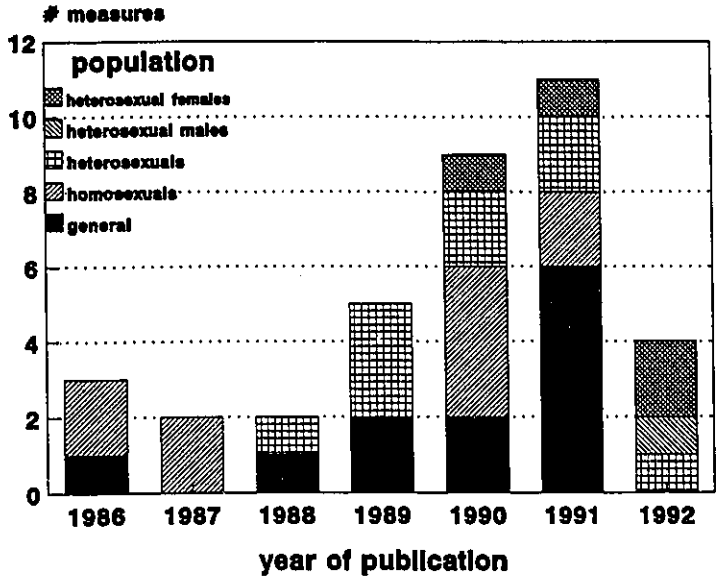


Figure 3. Population by year

Figure 3 shows the number of measures per population (based on sexual preference) by year. As, of 1988, the focus of measures has shifted from homosexuals to heterosexuals. Moreover, specific measures for heterosexual women appear in 1990, while the only measure for heterosexual men was developed in 1992.

Regarding the setting (see appendix) in which the indexes were applied, 13 were performed in the community, while 11 were in health facilities (outpatient clinics, STD clinics, family planning centres, primary health settings, and hospitals). The remaining 12 were employed in other settings such as colleges or gay bars, or the setting was not specified. This is an important element when considering the application of an index, since the measure validity may vary, for example, among college students and persons visiting STD clinics, who are at a higher risk for HIV infection.

4.2.5. Reliability

Only 8 of the 36 measures have been tested for their reliability. Internal consistency was evaluated for all 5 indexes developed with the psychometric method, yet it was not mentioned in any of the reports on indexes constructed with the empirical and judgemental methods. This finding is reasonable, since internal consistency evaluates the homogeneity of items measuring

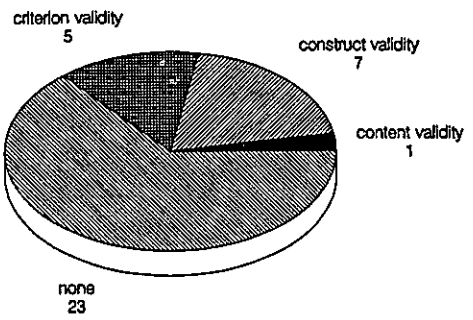
the same dimension when an additive model is used, which is often the case with psychometrically derived measures.

One index was evaluated for different aspects of reliability (i.e., test-retest, inter-observer agreement, and internal consistency)[118], one reports test-retest only[166], and one makes a vague reference to questionnaire reliability[172]. These data provide a discouraging picture with respect to the attention being paid to assessment of reliability.

4.2.6. Validity

Not even half of the existing indexes have been validated, and almost none have been tested for more than one type of validity. There is just one index[60] for which both construct and criterion validity are reported. From figure 4, it can be further noted that content validity was assessed for only one measure. This is surprising, first because this type of validity is a form of internal evidence that should complement criterion or construct validity, and second because its assessment is relatively easy and does not require data collection. Nonetheless, in the last three years, the number of validated measures has increased. Three measures were validated in 1990, 6 out of 11 in 1991, and 2 out of 4 in 1992.

Figure 4. Type of validity



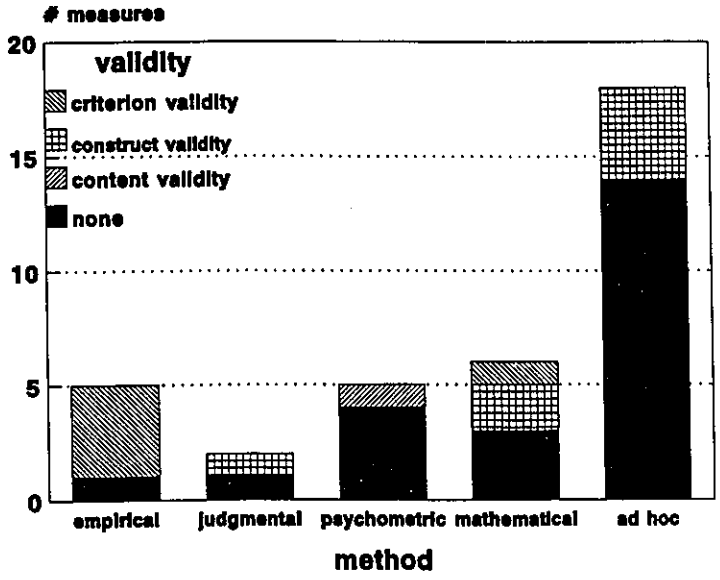


Figure 5. Type of validity by method

Figure 5 shows the number of measures validated by developing method. The proportion of measures not validated is higher for those developed with an ad hoc approach, confirming the scarce attention paid to all methodological aspects by this approach.

Although there are too few validated measures to appreciate the factors which influence the choice of the type of validity, it can be seen (figure 6) that measures intended to assess the risk of infection can be validated both with criterion and construct validity since the test for anti-HIV antibodies can be used as gold standard. On the other hand, for measures intended to study the risk of exposure to HIV, an external criterion is usually not available and measure quality is assessed with construct validity. In fact, the use of antibodies status as criterion is problematic because being

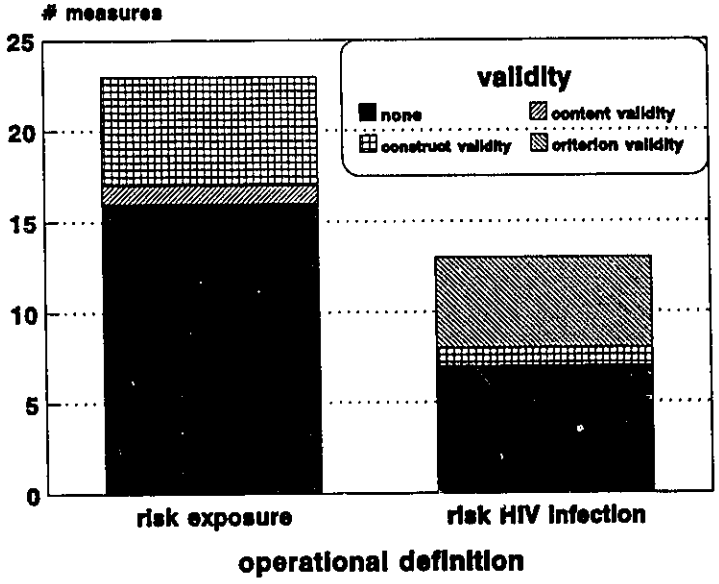


Figure 6. Type of validity by operational definition

"exposed to the risk" does not have the same meaning as being positive for HIV. Positive antibodies status could be used, since HIV-infected individuals represent a subset of the at-risk population, while negative antibodies status cannot be used for estimating false positives (assessed through an index), because seronegative individuals could be exposed to the risk but luckily uninfected, or they could be infected but still negative due to the time lapse between infection and development of antibodies.

4.2.7. Method used for the development

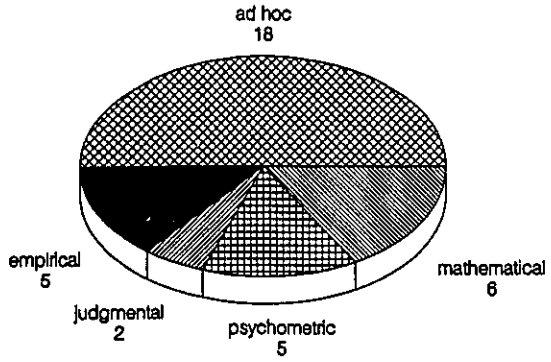


Figure 7. Developing methods

Among the 36 measures reviewed, half were constructed in an hoc manner with no standard method (figure 7). The distribution of methods used to develop the remaining 18 measures is more or less identical, with the exception of the judgemental method, which was adopted in only 2 instances.

When evaluating the HIV measures classified under the category 'ad hoc approach', it is not possible to infer what method was used in 8 of them since little or no information is given about the development process. Papers describing 6 of the indexes just mention that the measure is based on existing research and guidelines reported in the literature. The development process is described in details for only 4 measures; it shows that these indexes are based only on the researcher's personal opinion, with no systematic development basis.

However, the distribution of the types of method employed during the considered years is not random. Figure 8 shows that indexes developed with mathematical method appeared first (1988), followed by those developed with psychometric method (1989), judgemental method (1991), and empirical method (1991).

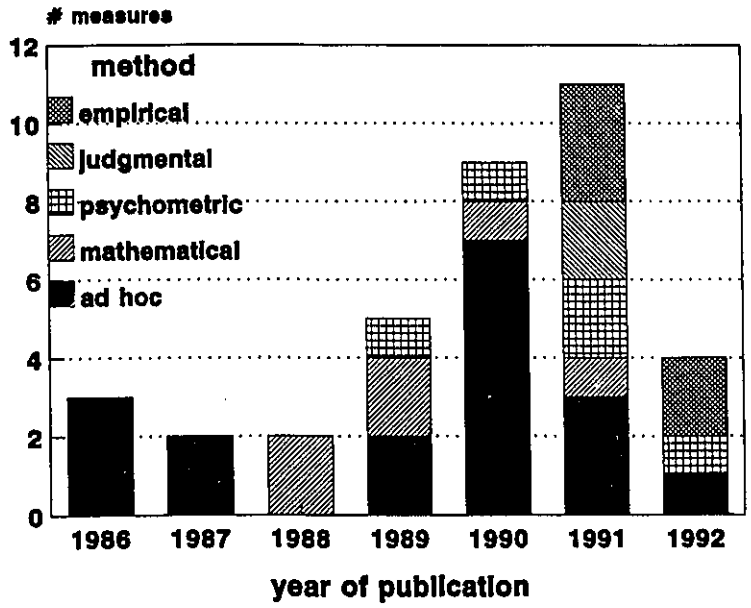


Figure 8. Methods employed by year

Each method is preferential for developing a certain type of index. Five of the 6 measures developed through mathematical method are in the form of probability equations. This type of instrument is also produced by empirical method (1 measure), but not by judgemental and psychometric methods. Conversely, the most frequent type of index among those psychometrically constructed is the scoring system (4 measures out of 5).

Figure 9 shows that the empirical method was used only for measures assessing HIV infection. This is due to the need of a direct measurable outcome, in this case serostatus for HIV, to be used as a dependent variable in regression models employed by the empirical method.

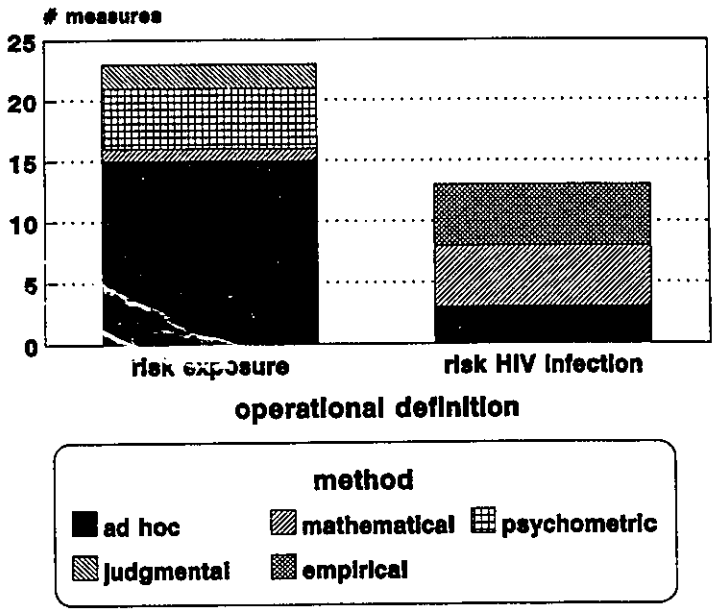


Figure 9. Methods by operational definition

In conclusion, although the number of indexes assessing the risk of HIV is quite small, some characteristics and trends can be identified. First, a growing interest in developing these kinds of instruments. Second, the presence of a small number of measures developed for predicting the future risk of acquiring HIV infection. Third, a change of interest from exposure to HIV to the actual risk of having or acquiring the infection. Fourth, a transition from measures for homosexuals to instruments for heterosexuals and towards indexes specific for increasingly better defined populations. Fifth, little emphasis on evaluating the reliability and validity of measures constructed. Moreover, the increasing use of standard methods in the last two years as opposed to ad hoc approaches for developing indexes suggests that the field is undergoing positive changes. The discussion provided in the following chapter could possibly contribute to this maturation process.

CHAPTER 5 - DISCUSSION AND CONCLUSIONS

This chapter examines how the classification of risk assessment methods for developing summary measures, proposed in chapter 3, works within the context of the risk of sexual transmission of HIV. It further discusses the adequacy of the existing instruments developed for HIV, and will provide some recommendations for improvements and future development.

Since this work was built mostly on the review and appraisal of papers describing indexes of risk found in medical, psychological, sociological, and biostatistical literature, a concern may arise about the completeness of the search due to the large number of articles which were found from primary reference. To check whether some fundamental key words were omitted, we inspected the MESH terms attached to articles not retrieved with computerized searches. The results did not indicate that any key words were left out, yet it confirmed the lack of specific MESH terms available for risk assessment and summary measures. Moreover, indexes are often not mentioned in the title or in the abstract, because they are not considered to be one of the main focuses of a paper, explaining why even the search of 'text word' in the titles and in the abstracts did not retrieve certain articles. Some papers may have been overlooked; however, the use of four different data-bases, the numerous trials carried out for finding the best strategy to retrieve the articles, the large number of titles initially investigated (approximately 1,200), and the accurate scanning of papers' bibliographies for secondary references, all contribute to the confidence placed in the adequacy of the literature search.

A problem found in reviewing the indexes in order to infer general concepts was the scarce attention paid by the authors to describing the instruments. Frequently the measure is not characterized by a name, the purpose and the application for which the index is intended are obscure, and the conceptual basis to which the developer refers and the operational definition not stated. Norms and indications for a correct use of the measure are described too vaguely

to permit replication by others. Moreover, the development process is rarely illustrated, and information about the assessment of reliability and validity is often lacking, preventing evaluation of the quality of the measure.

5.1. DISCUSSION ON METHODS USED FOR SUMMARY MEASURES FOR RISK OF HIV

The classification of methods proposed in chapter 3 seems to adequately fit the measures reviewed. In fact, it has been useful in classifying the measures for risk of HIV: the categories have been proven to be exhaustive since every index could be catalogued in only one class, and its use is quite simple. However, the large number of instruments for which a standard method has not been employed may create some suspicion, while the observation that in the last two years a small percentage of measures were developed with an ad hoc approach may indicate that the authors are now more careful in constructing indexes and hence utilise a standard method. At least two factors may be implicated in this common use of an ad hoc approach: first, the newness of the use of summary measures in the field of the risk assessment and the absence of an organized body of knowledge on the methods developed by various disciplines, resulting in researchers being unaware of the available methods; and second, the absence of guidelines for developing and evaluating summary measures, which would require following certain rules and criteria and explaining the logic of each step of the construction process, forcing the researcher to adopt a standard method. Moreover, for several measures the absence of a clear description of the development process precludes the appraisal of a method that may have been used.

Organizing the indexes according to their underlying method of development shows that each method is used preferentially for developing certain types of measures and that some factors influence the choice of the method. For example, the empirical method has been employed only when the risk of HIV infection is measured. In this case, HIV seropositivity can

be used as a dependent variable in regression techniques. Conversely, when the object of the measure is the risk of exposure, all methods have been used, with exception of the empirical method.

A further consideration is the distribution of the methods over the considered years. Indexes developed with the mathematical method were developed first (1988), followed by those developed with the psychometric method (1989), the judgemental method (1991), and the empirical method (1991). Two factors influence this pattern: the accessible knowledge and the availability of a sufficient number of cases of disease. At the beginning of the epidemic, only the mathematical method was feasible, as this method has explored the new disease without empirical data, simply by modelling the phenomena. It contributed to the advance of scientific research by indicating the information needed and by suggesting hypotheses about the mechanisms and factors involved in HIV transmission. As some knowledge became available, the new theories were used by the psychometric approach, and empirical data allowed for the developed measures to be tested. However, for the judgemental method, more knowledge was required in order for "experts" on sexual transmission of HIV to exist, while the empirical method required both knowledge for deciding which variables needed to be collected and enough data to perform statistical analysis. Moreover, all the existing measures developed with the empirical method are intended to screen individuals for their present risk of being infected with HIV because they are derived from cross-sectional studies. There has not been enough time to enrol and follow for a sufficient number of years cohorts that could be employed to construct indexes with the intent of predicting the future risk of acquiring the infection.

In the near future, measures developed with the empirical and judgemental methods will become increasingly common. Instruments developed with the mathematical and psychometric approaches will continue to be constructed, though they will likely become more complex and accurate than the existing ones since they will take advantage of the new knowledge and theories

on sexual transmission of HIV and of the information produced by indexes developed through empirical and judgemental approaches.

5.2. DISCUSSION ON EXISTING SUMMARY MEASURES FOR RISK OF HIV

The aim of the following discussion is to summarize and comment on the state of the art of summary measures that assess the risk of sexual transmission of HIV and to interpret some of the trends found. The first indexes found were published in 1986. This is understandable as the first cases of AIDS were only beginning to be identified in 1981, the cause of the disease (the HIV virus) was discovered in 1984[179], and the first test for detecting the antibodies in infected persons became available in 1985[180]. The increasing number of measures developed in the subsequent years parallels the rising number of cases due to sexual transmission, the scepticism concerning a readily available vaccine, and the consequent increasing interest in preventing risky behaviours.

Since 1988, the spread of the epidemic into the general population has produced an increasing number of indexes assessing the risk of having or acquiring the infection. The reason for developing a measure which predicts the presence of infection, when a gold standard (i.e., the serologic test) exists, is related to the possibility of screening a large number of people with a non-invasive and inexpensive instrument and of alerting and advising those at higher risk to seek antibody testing and counselling. Self-administered forms of risk indexes may also permit self-evaluation and ensure the confidentiality of the information. Moreover, the high number of false positive results obtained with the antibodies test, in a population with a low prevalence of infection, can be reduced by pre-screening the people and later examining only high-risk individuals. This can be accomplished by using the risk index and the serological test in series which improves the specificity of the results.

On the other hand, instruments estimating the future risk may be used to educate

people. Presenting the potential risk of infection as being dependent on individual behaviour may incite a person to learn about risky behaviours for sexual transmission of HIV, resulting in possible behavioural change. Some of these predictive measures also intend to clarify the relative importance of risk factors in acquiring HIV infection in the future, by modelling the risk, in order to find priorities for educational messages and for health policy recommendations aimed at controlling the spread of sexually transmitted HIV.

There is still relatively little knowledge on sexual transmission of HIV. The risk involves factors which increase the probability of transmission, factors which elevate the likelihood of meeting an infected partner, and the prevalence of HIV among potential partners. Some of these are related to the person's sexual behaviour, while others are inborn or demographic characteristics, and environmental factors:

a) Sexual behaviours exposing the individual to the HIV virus

- sexual practices and behaviours which expose the individual, but which pose various levels of risk (e.g., oral, anal or vaginal intercourse, duration of the relationship, frequency of intercourse, condom use, sexual intercourse during menstruations, etc.);
- number and choice of partners which influence the probability of meeting an infected partner;

b) Susceptibility of the individual (e.g., genetic factors, age, concomitant STDs, impaired immunity, trauma during sex, etc.);

c) Infectiousness of the infected partner (e.g., stage of HIV infection, concurrent infections or STDs, etc.);

d) Infectivity and virulence of HIV (viral strains or isolates); and

e) Prevalence of HIV in the area.

The existing measures mostly include factors related to sexual behaviours and choice and number of partners and employ some imprecise estimates of prevalence. However, in the last couple of years, some instruments have also considered factors that increase the

susceptibility of the host[69,170] and the infectiousness of the infected partner[70]. This is probably a result of the acquisition of new knowledge in recent years. In fact, epidemiological studies have consistently found an association between a concurrent STD and the risk of acquiring the infection[181-3], supporting the biological hypothesis that STDs could enhance penetration of HIV by causing epithelial disruption or increasing the pool of target cells. Another case is the accumulating information suggesting that HIV infected persons may be more infectious in an advanced stage of disease when there is an increased amount of replicating virus in the individual[184-8]. Immunologic and virologic parameters, such as T4 lymphocyte count, antigenemia, or viraemia, may then be used as markers for increased infectiousness. These examples illustrate how new scientific acquisitions are reflected in the factors included in risk measures and demonstrate the importance of developers' stating the theory or hypothesis on which the index is based.

The change over time, in the choice of populations for which the measures are intended, resembles the growth of the epidemic in Pattern I countries**, where the disease started among homosexuals and then spread to heterosexual women (partly through bisexual men and IVDUs) and only later to heterosexual men. However, one should note the lack of indexes constructed for developing countries, where the extent of the epidemic is much greater and where heterosexual transmission spreads the infection among the general population and not only among limited risk groups.

One third of the measures reviewed was developed for the general population, which includes homosexuals and heterosexuals of both genders. These indexes have a broader application and a wider generalizability, yet the measure is less accurate. The choice of some

---

** Pattern I countries are defined by WHO as those countries where most cases occur among homosexual or bisexual males and intravenous drug users. Heterosexual transmission is responsible for only a small percentage of cases, though it is increasing. This pattern is typical of industrialized countries, including North America, Western Europe, Australia, New Zealand, and many urban areas of Latin America[189].

researchers to develop different indexes specific for heterosexual males and heterosexual females is an attempt to develop more valid instruments.

An important consideration in reviewing existing measures is the inadequate information concerning their evaluation. Validity and reliability may have been examined but not reported in the articles, or the measures may not have been tested at all; in either case, there still exist doubts regarding the quality of the available instruments. Reliability has been tested more frequently in measures developed with the psychometric method, suggesting that developers who use this method place greater importance on verifying reliability. Moreover, even the papers which refer to the reliability of measures did not completely evaluate reliability, and only one measure has been tested for the major forms of reliability[118]. The evidence of testing for validity is also discouraging. Some positive signals can be seen in the increase of validated measures in the most recent years, although the thoroughness of the testing is still an issue. This may be imputed to the lack of clarity of the conceptual basis to which the measures refer. On the other hand, validity is seldom described for measures developed with an ad hoc approach. This may be indicative of the lack of attention paid to all methodological aspects by this approach, both in the development and in the evaluation of the instruments.

5.3. RECOMMENDATIONS FOR IMPROVEMENTS AND FUTURE DEVELOPMENTS

5.3.1. Recommendations for methods for risk assessment

The confusion found in the literature concerning the available methods for risk assessment could explain the presence of a great number of summary measures developed in an ad hoc manner. There is a clear need of enhancing awareness with respect to developing methods, their individual advantages and weaknesses, and the factors that influence the appropriateness of a method. Researchers from various disciplines should contribute to the creation of a body of knowledge by increasing comparison among methods, promoting debate,

and more clearly expressing the theoretical basis.

Moreover, the lack of a clear construction process, revealed by the existing risk indexes, suggests the need for guidelines on the development and evaluation of summary measures for risk assessment. Some clues can be taken from the present work, which illustrates the sequence of operations required by the construction process, according to the different methods. The main steps should roughly include: (1) stating the characteristics of the desired measure (type, purpose, operational definition, and application); (2) identifying a pool of potential variables; (3) selecting the appropriate variables; (4) assessing the adequate scale for each variable; (5) combining the variables into a single global measure (if necessary, assigning weights to the variables); (6) stating norms for the standard use of the measure and reference values for interpreting the results; (7) assessing validity and reliability of the measure; and (8) refining the instrument according to the results of the validation.

The review of the literature also has shown the problem of jargon and terminology in the field of summary measures for risk assessment. Since this is a new field, building on the experience of various disciplines and depending on the exchange of knowledge and information among researchers of different backgrounds, it is necessary to standardize the terminology or at least to create a common language that will not preclude future development, updating it as new terms appear. Parallel to this, other recommendations regard the creation of specific MESH terms for risk assessment in order to easily search the literature and the establishment of university courses in methods for risk assessment for spreading the knowledge also among students of health sciences.

The temporal progression seen among measures for HIV developed with a standard method and the influence of accessible knowledge and availability of a sufficient number of cases on the possible use of each method in an iterative process over time should be taken into account. However, integration of these methods is strongly recommended. Researchers of

different disciplines have faced the problem of combining risk factors into a single summary measure for global assessment using only tools provided by their scientific areas. A multidisciplinary approach could help to strengthen the methods developed in different environments. For example, the theoretical bases of the mathematical and psychometric methods could be integrated with empirical data, and the expertise of a panel of judges could support decisions implied in the empirical approach, while data could be used to strengthen experts' decisions. The mutual support of different methods could enrich the process of selecting, weighting and aggregating the risk factors.

5.3.2. Recommendations for summary measures for risk

Among the 36 measures for HIV that have been reviewed, there are only two notable cases where researchers have employed existing indexes developed by another group (measures 11 and 17), while most others, although quite similar, have been newly produced without any reference to previous work. Before deciding to develop a new index, the researcher has to be aware of the existing instruments, some of which may suit the purpose of the research. The suggestion of not reinventing similar instruments leads to the recommendation that instruments be clearly described, so that others may use them. This can be accomplished in two ways. First, establishing criteria for the acceptance of manuscripts by the journals' editors, as done with epidemiological studies, and second, setting guidelines for the presentation of measures. The journals editors' criteria should probably be even more stringent than those for papers describing a study, since the presented instrument would be used, modified, or further validated by others, hence requiring that all the important elements be described. For example, the article should contain the identifying name for the instrument, a statement about the purpose and the intended application of the measure, a description of the conceptual basis together with supporting references, an explanation of the logic process used to construct the measure from

its theoretical approach, and a justification of the choice of the particular development method used. The methods section should include a clear description of the construction process and the strategy for assessing reliability and validity. For most of the indexes reviewed, these aspects are lacking or are very insufficient, preventing both the judgement and the replication of the process. Finally, papers without information on testing for validity and reliability should not be accepted. Since journals usually have limited space, the papers should at least give a brief description of the measure and refer to another accessible source for a detailed presentation.

Standard guidelines for presenting risk measures would compel investigators to include all the elements, so that other researchers could employ the instrument. Beyond the basic characteristics (type, purpose, operational definition, application), clear and objective decision rules for each step involved in deriving the final score should be included: indications to identify the variables and the associated scale, the mechanism of aggregation of the basic components, and the output categories that emerge. The description should also supply the actual questionnaire or the other instruments for collecting the information, as well as details on the method of its administration. Finally, reference values or standard cut-offs should be provided for helping with the interpretation of the results.

Investigators that wish to select an existing index face the problem of judging the quality of the instrument. It would be helpful to identify a set of criteria for evaluating and comparing risk assessment indexes. One example is the Trauma Severity Index conference held in Woodstock, Illinois in 1980, which established seven criteria, with different weights indicating their relative importance, for assessing and comparing severity indices[21]. These criteria should at least include the thoroughness of reliability and validity testing, and the quality of evidence for reliability, content validity, and criterion and/or construct validity[17].

5.3.3. Recommendations for measures for risk of HIV

The review of instruments for sexual transmission of HIV shows that two main areas seem to be in particular need of improvement. The first regards the lack of reference to a conceptual basis, which is critical for developing, interpreting, and evaluating a summary measure. A clear description of the underlying theoretical reference should be followed by an explanation of the logic process that led to the operational definition, the selection, and the combination of factors. The second problematic area is the validation of the developed measures: less than half of the indexes for HIV have been tested for validity, and not even one fourth have been tested for reliability. The importance of instrument reproducibility and validity should be stressed, the practice of evaluating should be extended to all methods, and the quality and completeness of testing should be encouraged for every new measure. There are ethical reasons for which measures that have not been tested should not be used in either research or clinical fields: they may provide incorrect information to the scientific community, may result in useless preventive measures, and may mistakenly define at-risk individuals, causing psychological and emotional distress among those wrongly classified. Moreover, a young field of research such as this, may also suffer from lack of credibility, being accused of producing inaccurate instruments. The type of testing could be suggested by the conceptual basis and should involve all major aspects of reliability and both internal and external forms of validity.

Researchers in this field should carefully watch the new scientific acquisitions and knowledge on sexual transmission of HIV and be ready to include new factors into summary measures. The trend already seen towards indexes specific for better defined populations should be encouraged because, as the epidemic progresses, there is a clear need to develop more accurate instruments, while the absence of measures for developing countries should be filled.

The frequent use of an intuitive or ad hoc approach, seen among the existing measures for sexual transmission of HIV, should be discouraged because it precludes the replication of

results and the evaluation of the development process. The last recommendation to the researchers thus regards the use of standard methods for the development of new measures for assessing the risk of HIV, resulting in an enhanced quality of the instruments.

5.4. CONCLUSIONS

In achieving the goal of studying procedures used to appraise or to develop summary measures for assessing the risk involved in sexual transmission of HIV, the objectives of this research have been attained.

Moreover, some conclusions can be drawn from this work. First, principles and concepts of summary measures are not common among researchers in health sciences. Medical literature, textbooks, and courses do not pay much attention to this field. Second, methods for risk assessment have been developed by various disciplines, but a comprehensive and structured body of knowledge that organises, describes, and compares the available methods is still not available. Third, the review of measures for risk of HIV demonstrates the increasing need and interest in these type of instruments. However, researchers are individually attempting to deal with the confusion present in the literature and the jungle of jargon produced by the different disciplines.

The proposed classification of risk assessment methods organises the existing materials into generalizable categories, helps to clarify and expand upon the knowledge on single strategies, and illustrates potential procedures for developing summary measures of risk. The assessment of the strengths and weaknesses of each method facilitates the understanding of the problems and advantages of indexes developed with a certain method. The comparison of different methods stresses the importance of the factors which influence the choice of the methods beyond the intrinsic strengths and weaknesses and emphasizes the recognition that the usefulness of a method changes over time and according to the specific situation. Accessible

knowledge and availability of a sufficient number of cases influence the possible use of each method in an iterative process over time.

Though imperfect, at times based on contradictory references, and in need of refinements and revisions, the classification has proven to be useful in its application to the existing measures for risk of HIV. In fact, with this classification, it has been possible to arrange the indexes meaningfully and to describe the current status of this field, recognizing its limitations and identifying positive trends that should be encouraged. This interesting application can be seen as preliminary, rough evidence of the importance of this type of work. Moreover, it may help researchers in developing new measures of risk, being aware of the available procedures and their implications. However, the efforts have been worthwhile, resulting in important and useful work in the fields of risk assessment in general and HIV in particular.

Three areas for future improvements and developments can be recognized: first, the creation of a well-defined body of knowledge on available methods for risk assessment and the refinement of the proposed classification and its underlying conceptualization through more applications to other fields; second, the standardization of principles, strategies, and terminology for summary measures of risk with the establishment of guidelines for developing, testing, presenting, and evaluating instruments; and finally, the integration of the available methods in a multidisciplinary approach, which could benefit from the strengths of different procedures.

This work is an attempt to organize, describe, and evaluate the available risk assessment methods. A broad perspective was taken to find methods that could be used for risk assessment, to clearly illustrate the factors influencing the choice of a method, and to organize them in a framework for a better understanding of risk index development and for increasing the number of high quality measures. The appraisal of risk for other multifactorial phenomena, such as the progression from HIV infection to AIDS or events related to other diseases, will improve the efforts aimed at establishing standard methodologies for risk assessment.

REFERENCES

1. Kannel WB, McGee DL. Composite scoring - Methods and predictive validity: insights from the Framingham study. Health Serv Res 1987;22:499-535.

2. Chin J. Global estimates of HIV infections and AIDS cases: 1991. AIDS 1991;5(suppl 2):S57-S61.

3. World Health Organization. Current and future dimensions of the HIV/AIDS pandemic. A capsule summary. Geneva: WHO, 1991; publication no. WHO/GPA/RES/SFE/91.4.

4. Quinn TC. Epidemiologic trends in HIV infection. Curr Opin Infect Dis 1992;5:189-200.

5. Nunnally JC. Psychometric theory. New York, NY: McGraw-Hill Book Company, 1967.

6. Bombardier C, Tugwell P. A methodological framework to develop and select indices for clinical trials: statistical and judgmental approaches. J Rheumatol 1982;9:753-7.

7. Kipke MD, Drucker E. Reliability and construct validity of the Needle Sharing Inventory. Int J Addict 1989;24:515-26.

8. Campostrini S, McQueen D. Estimating exposure to HIV infection from self-reported sexual behavior in a population survey: constructing an index of behavioral risk. Am J Public Health 1993 (in press).

9. Schmidt B, Kirpalani H, Rosenbaum P, Cadman D. Strengths and limitations of the Apgar score: a critical appraisal. J Clin Epidemiol 1988;41:843-50.

10. Wilson PWF, Castelli WP, Kannel WB. Coronary risk prediction in adults. (The Framingham heart study). Am J Cardiol 1987;59:91G-94G.

11. Lemeshow S, Teres D, Avrunin JS, Pastides H. A comparison of methods to predict mortality of intensive care unit patients. Crit Care Med 1987;15:715-22.

12. Feinstein AR. Clinimetrics. New Haven, NY: Yale University Press, 1987.

13. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27-36.

14. Cadman D, Goldsmith C. Construction of social value or utility-based health indices: the usefulness of factorial experimental design plans. J Chronic Dis 1986;39:643-51.

15. Alemi F, Stokes III J, Rice J, Karim E, LaCorte W, Saligman L, Nau R. Appraisal of modifiable hospitalization risks. Med Care 1987;25:582-91.

16. Halvorsen JG. Designing self-report instruments for family assessment. Fam Med 1990;22:478-84.

17. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York, NY: Oxford University Press, 1987.

18. Iezzoni LI. Severity of illness measures. Comments and caveats. Med Care 1990;28:757-61.

19. Pearson TA, LaCroix AZ, Mead LA, Liang K-Y. The prediction of midlife coronary heart disease and hypertension in young adults: the Johns Hopkins multiple risk equations. Am J Prev Cardiol 1990;6(suppl 2):23-8.

20. Kruse JA, Thill-Baharozian MC, Carlson RW. Comparison of clinical assessment with APACHE II for predicting mortality risk in patients admitted to a medical intensive care unit. JAMA 1988;260:1739-42.

21. Gustafson DH, Fryback DG, Rose JH, et al. An evaluation of multiple trauma severity indices created by different index development strategies. Med Care 1983;21:674-91.

22. Keirse MJNC. An evaluation of formal risk scoring for preterm birth. Am J Perinatol 1989;6:226-33.

23. Gustafson DH, Fryback DG, Rose JH, Yick V, Prokop CT, Detmer DE, Moore J. A decision theoretic methodology for severity index development. Med Decis Making 1986;6:27-35.

24. Weiss RA, Jaffe HW. Duesberg, HIV and AIDS. Nature 1990;345:659-60.

25. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press, 1989.

26. Boyle MH, Torrance GW. Developing multiattribute health indexes. Med Care 1984;22:1045-57.

27. Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. Crit Care Med 1985;13:519-25.

28. Keene AR, Cullen DJ. Therapeutic intervention scoring system: update 1983. Crit Care Med 1983;11:1-3.

29. Levy PS, Mullner R, Goldberg J, Gelfand H. The estimated survival probability index of trauma severity. Health Serv Res 1978;13:28-35.

30. Last JM, ed. A dictionary of epidemiology. New York, NY: Oxford University Press, 1988.

31. Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills, CA: Sage Publications, 1979.

32. Bartko JJ. Measurement and reliability: statistical thinking considerations. Schizophr Bull 1991;17:483-9.

33. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York, NY: John Wiley and Sons, 1981.

34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

35. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the Index of Well-Being. Health Serv Res 1976;11:478-507.

36. Sackett DL, Chambers LW, MacPherson AS, Psych D, Goldsmith CH, Mcauley RG. The development and application of indices of health: general methods and a summary of results. Am J Public Health 1977;67:423-8.

37. Tugwell P, Bombardier C. A methodological framework for developing and selecting endpoints in clinical trials. J Rheumatol 1982;9:758-62.

38. American Psychological Association. Standards for educational and psychological tests. Washington, DC: American Psychological Association, 1974.

39. Guildford JP. Psychometric methods. 2nd ed. New York, NY: McGraw-Hill Book Company, 1954.

40. Bellamy N. Critical review of clinical assessment techniques for rheumatoid arthritis trials: new developments. Scand J Rheumatol 1989;80(suppl):3-16.

41. Miller JE. Guidelines for selecting a health status index: suggested criteria. In: Berg RL, ed. Health status indexes. Chicago, IL: Hospital Research and Educational Trust, 1973:243-51.

42. Moriyama IM. Problems in the measurement of health status. In: Sheldon E, Moore W, eds. Indicators of social change. New York, NY: Russel Sage Foundation, 1968.

43. Goldsmith SB. The status of health status indicators. Health Serv Rep 1972;87:212-20.

44. Alemi F, Gustafson DH, Johnson M. How to construct a subjective index. Eval Health Prof 1986;9:42-52.

45. Kannel WB. Some lessons in cardiovascular epidemiology from Framingham. Am J Cardiol 1976;37:269-82.

46. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham study. Am J Cardiol 1976;38:46-51.

47. Orrego H, Israel Y, Blake JE, Medline A. Assessment of prognostic factors in alcoholic liver disease: toward a global quantitative expression of severity. Hepatology 1983;3:896-905.

48. Fedrick J. Antenatal identification of women at high risk of spontaneous pre-term birth. Br J Obstet Gynaecol 1976;83:351-4.

49. Brians LK, Alexander K, Grota P, Chen RWH, Dumas V. The development of the RISK tool for fall prevention. Rehabilitation Nurs 1991;16:67-9.

50. Ross MG, Hobel CJ, Bragonier JR, Bear MB, Bemis RL. A simplified risk-scoring system for prematurity. Am J Perinatol 1986;3:339-44.

51. DesHarnais SI, Chesney JD, Wroblewski RT, Fleming ST, McMahon LF. The risk-adjusted mortality index. A new measure of hospital performance. Med Care 1988;26:1129-45.

52. DesHarnais SI, McMahon LF, Wroblewski RT, Hogan AJ. Measuring hospital performance. The development and validation of risk-adjusted indexes of mortality, readmissions, and complications. Med Care 1990;28:1127-41.

53. De Geest AFE, Schoolmeesters I, Willems JL, De Geest H. An analysis of the level of dental care in cardiac patients at risk for infective endocarditis. Acta Stomatol Belg 1990;87:95-105.

54. Champion HR, Sacco WJ, Hannan DS, et al. Assessment of injury severity: the triage index. Crit Care Med 1980;8:201-8.

55. Lichtenstein MJ, Bolton A, Wade G. Derivation and validation of a decision rule for predicting seat belt utilization. J Fam Pract 1989;28:289-92.

56. Feinstein AR, Wells CK, Walter SD. A comparison of multivariable mathematical methods for predicting survival - I. Introduction, rationale, and general strategy. J Clin Epidemiol 1990;43:339-47.

57. Baisch H, Otto U, Kloppel G. Malignancy index based on flow cytometry and histology for renal cell

carcinomas and its correlation to prognosis. Cytometry 1986;7:200-4.

58. Pepe PE, Thomas RG, Stager MA, Hudson LD, Carrico CJ. Early prediction of the adult respiratory distress syndrome by a simple scoring method. Ann Emerg Med 1983;12:749-55.

59. Baldescu R, Steinbach M. Use of discriminant analysis in the calculation of a prognostic index for the primary prevention of cardiovascular diseases. Med Interne 1984;22:49-54.

60. Iguchi MY, Rosen M, Musikoff H, et al. Predictors of HIV seropositivity in Newark and Jersey City i.v. drug users not currently enrolled in treatment. NIDA Res Monogr 1991;105:473-4.

61. Fortney JA, Whitehorne EW. The development of an index of high-risk pregnancy. Am J Obstet Gynecol 1982;143:501-8.

62. Molfese VJ, Thomson BK, Beadnell B, Bricker MC, Manion LG. Perinatal risk screening and infant outcome. J Reprod Med 1987;32:569-76.

63. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York, NY: Oxford University Press, 1986:118.

64. Shapiro E, Tate R. Who is really at risk of institutionalization? Gerontologist 1988;28:237-45.

65. Roi LD, Flora JD, Davis TM, Wolfe RA. Two new burn severity indices. J Trauma 1983;23:1023-9.

66. Fleming ST. Toward the development of integrative risk-adjusted measures of quality using large clinical data bases. Eval Health Prof 1992;15:43-58.

67. Shaper AG, Pocock SJ, Phillips AN, Walker M. Identifying men at high risk of heart attacks: strategy for use in general practice. Br Med J 1986;293:474-9.

68. Smith MA, Stratton WC, Roy L. Prospective labor risk assessment in a rural community hospital. Am J Perinatol 1988;5:113-20.

69. Crowe SM, Elbeik T, Ulrich PP, Mills J, Moss A. Lack of evidence of occult Human Immunodeficiency Virus in seronegative individuals at very high risk of infection. J Med Virol 1991;35:160-4.

70. European Study Group on heterosexual transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. Br Med J 1992;304:809-13.

71. Walter SD, Feinstein AR, Wells CK. A comparison of multivariable mathematical methods for predicting survival - II. Statistical selection of prognostic variables. J Clin Epidemiol 1990;43:349-59.

72. Tsuji Y, Koga S, Ibayashi H, Nose Y, Akazawa K. Prediction of the prognosis of liver cirrhosis in Japanese using Cox's proportional hazard model. Gastroenterol Jpn 1987;22:599-606.

73. Pedhazur EJ. Multiple regression in behavioral research. 2nd ed. New York, NY: Holt, Rinehart and Wiston, 1982.

74. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. Am J Public Health 1991;81:1630-5.

75. Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. Stat Med 1991;10:1213-26.

76. Wells CK, Feinstein AR, Walter SD. A comparison of multivariable mathematical methods for predicting survival - III. Accuracy of predictions in generating and challenge sets. J Clin Epidemiol 1990;43:361-72.

77. Fos PJ, McLin CL. The risk of falling in the elderly: a subjective approach. Med Decis Making 1990;10:195-200.

78. Gustafson DH, Sainfort FC, Van Konigsveld R, Zimmerman DR. The quality assessment index (QAI) for measuring nursing home quality. Health Serv Res 1990;25:97-127.

79. Gustafson DH, Holloway DC. A decision theory approach to measuring severity in illness. Health Serv Res 1975;10:97-106.

80. Delbeq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning. A guide to nominal group and Delphi processes. Glenview, IL: Scott, Foresman and Company, 1975.

81. Gustafson DH, Shukla RK, Delbecq A, Walster GW. A comparative study of differences in subjective likelihood estimates made by individuals interacting groups, Delphi groups, and nominal groups. Organizational Behav Hum Perform 1973;9:280-91.

82. Alemi F, Turner BJ, Markson LE, Szorady R, McCarron T, Prognosis after AIDS: a severity index based on experts' judgements. Interfaces 1991;21:105-16.

83. Gustafson DH, Kestly JJ, Greist JH, Jensen NM. Initial evaluation of a subjective Bayesian diagnostic system. Health Serv Res 1971;6:204-13.

84. Decker JL. Summary. Conference on outcome measurement in rheumatology clinical trials. J Rheumatol 1982;9:802-6.

85. Nemiroff PM, King DC. Group decision-making performance as influenced by consensus and self-orientation. Hum Relat 1975;28:1-21.

86. Bombardier C, Tugwell P, Sinclair A, Dok C, Anderson G, Buchanan WW. Preference for endpoint measures in clinical trials: results of structured workshops. J Rheumatol 1982;9:798-801.

87. Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE - acute physiology and chronic health evaluation. A physiologically based classification system. Crit Care Med 1981;9:591-7.

88. Wagner DP, Knaus WA, Draper EA. Statistical validation of a severity of illness measure. Am J Public Health 1983;73:878-84.

89. Knaus WA, Le Gall JR, Wagner DP, et al. A comparison of intensive care in the USA and France. Lancet 1982;ii:642-6.

90. Le Gall J-R, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984;12:975-7.

91. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. Crit Care Med 1974;2:57-60.

92. Gonnella JS, Hornbrook MC, Louis DZ. Staging of disease. A case-mix measurement. JAMA 1984;251:637-44.

93. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease

classification system. Crit Care Med 1985;13:818-29.

94. Torrance GW. Toward a utility theory foundation for health status index models. Health Serv Res 1976;11:349-69.

95. Fisher GW. Utility models for multiple objective decisions: do they accurately represent human preferences? Decis Sci 1979;10:451-79.

96. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York, NY: John Wiley and Sons, 1976.

97. Dawes RM. The robust beauty of improper linear models in decision making. Am Psychol 1979;34:571-6.

98. Fryback DG, Keeney RL. Constructing a complex judgmental model: an index of trauma severity. Manage Sci 1983;29:869-83.

99. Thornton JG. Analysis of probability and measurement of values. Bailliere Clin Obstet Gynaecol 1990;4:867-84.

100. Elmslie T, McDowell I, Wells G, Hollingworth G. Sexual behaviour related to transmission of HIV infection among women in the Canadian primary care setting. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;1:409 (abstract M.D.4078).

101. Elmslie T, Wells G, Hollingworth G, McDowell I, Graham J. Study of sexual behaviour associated with transmission of HIV infection among women in the primary care setting. Final report. April 1992 (unpublished materials).

102. Campostrini S. Un indice di rischio di infezione da HIV per la popolazione. Epidemiol Prev 1993 (in press).

103. McQueen DV, Campostrini S, Robertson BJ, Uitenbroek D. A study of lifestyle and health. Interim report # 3. Research Unit in Health and Behavioral Change, University of Edinburgh, April 1991.

104. Robertson BJ, McQueen DV. Continuous collection of data on AIDS-related behaviour in the UK. Proceedings of the 1991 Public Health Conference on Records and Statistics. Washington, DC: National Center for Health Statistics, 1991:247-51.

105. Mulley AG. Assessing patients' utilities. Can the ends justify the means? Med Care 1989;27(suppl 3):S269-S281.

106. Kaplan RM, Ernst JA. Do category rating scales produce biased preference weights for a health index? Med Care 1983;21:193-207.

107. Wilson C, Heath DI, Imrie CW. Prediction of outcome in acute pancreatitis: a comparative study of APACHE II, clinical assessment and multiple factor scoring systems. Br J Surg 1990;77:1260-4.

108. Slovic P, Lichtenstein S. Comparison of Bayesian and regression approach to the study of information processing in judgement. Organizational Behav Hum Perform 1971;6:649-744.

109. Golden RR, Teresi JA, Gurland BJ. Development of indicators scales for the Comprehensive Assessment and Referral Evaluation (CARE) interview schedule. J Gerontol 1984;39:138-46.

110. Baird JC, Noma E. Fundamentals of scaling and psychophysics. New York, NY: John Wiley and

Sons, 1978.

111. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52:281-302.

112. Speltz ML, Gonzales N, Sulzbachers S, Quan L. Assessment of injury risk in young children: a preliminary study of the Injury Behavior Checklist. J Pediatr Psychol 1990;15:373-83.

113. Jackson DN. The dynamics of structured personality tests: 1971. Psychol Bull 1971;78:229-48.

114. Angleitner A, John OP, Lohr F-J. It's what you ask and how you ask it: an itemmetric analysis of personality questionnaires. In: Angleitner A, Wiggins JS, eds. Personality assessment via questionnaires. Berlin: Springer-Verlag, 1986:61-108.

115. Hase HD, Goldberg LR. Comparative validity of differential strategies of constructing personality inventory scales. Psychol Bull 1967;67:231-48.

116. Nowack KM. Initial development of an inventory to assess stress and health risk. Am J Health Promotion 1990;4:173-80.

117. Bassman LE. Object relations and self-reported AIDS self-care behavior. Psychol Rep 1991; 68:915-23.

118. Darke S, Hall W, Heather N, Heather N, Ward J, Wodak A. The reliability and validity of a scale to measure HIV risk-taking behaviour among intravenous drug users. AIDS 1991;5:181-5.

119. Curran JW. Foreword. In: Castillo-Chavez C, ed. Mathematical and statistical approaches to AIDS epidemiology. Berlin: Springer-Verlag 1989:VII-VIII. (Levin S, ed. Lectures Notes in Biomathematics; vol 83).

120. Anderson RM. The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS. J Acquir Immune Defic Syndr 1988;1:241-56.

121. Allard R. A family of mathematical models to describe the risk of infection by a sexually transmitted agent. Epidemiology 1990;1:30-3.

122. Blower SM, Hartel D, Dowlatabadi H, Anderson RM, May RM. Drugs, sex and HIV: a mathematical model for New York City. Philos Trans R Soc Lond Biol 1991;321:171-187.

123. Van Druten JAM, Reintjes AGM, Jager JC et al. HIV infection dynamics and intervention experiments in linked risk groups. Stat Med 1990;9:721-36.

124. Felton WF. Contrasting views on the infectivity of gonorrhoea. Brit J Vener Dis 1973;49:151-4.

125. Holmes KK, Johnson DW, Trostle HJ. An estimate of the risk of men acquiring gonorrhea by sexual contact with infected females. Am J Epidemiol 1970;91:170-4.

126. Pedersen AHB, Harrah WD. Follow up of male and female contacts of patients with gonorrhea. Public Health Rep 1970;85:997-1000.

127. Hooper RR, Reynolds GH, Jones OG, et al. Cohort study of venereal disease. I: the risk of gonorrhea transmission from infected women to men. Am J Epidemiol 1978;108:136-44.

128. De Gruttola V, Mayer KH. Assessing and modeling heterosexual spread of the Human Immunodeficiency Virus in the United States. Rev Infect Dis 1988;10:138-50.

129. Padian NS, Shiboski SC, Jewell NP. The effect of number of exposures on the risk of heterosexual HIV transmission. J Infect Dis 1990;161:883-7.

130. Grant RM, Wiley JA, Winkelstein W. Infectivity of the Human Immunodeficiency Virus: estimates from a prospective study of homosexual men. J Infect Dis 1987;156:189-93.

131. Eisenberg B. The number of partners and the probability of HIV infection. Stat Med 1989;8:83-92.

132. Sandberg S, Awerbuch TA. Mathematical formulation and studies of the risk parameters involved in HIV transmission. Bull Math Biol 1989;51:467-74.

133. Wiley JA, Herschkorn SJ, Padian NS. Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse. Stat Med 1989;8:93-102.

134. Eisenberg B. The effect of variable infectivity on the risk of HIV infection. Stat Med 1991;10:131-9.

135. Peterman TA, Stoneburner RL, Allen JR, Jaffe HW, Curran JW. Risk of Human Immunodeficiency Virus transmission from heterosexual adults with transfusion-associate infections. JAMA 1988;259:55-8.

136. Kaplan EH. Modeling HIV infectivity: must sex acts be counted? J Acquir Immune Defic Syndr 1990;3:55-61.

137. Watts CH, May RM. The influence of concurrent partnerships on the dynamics of HIV/AIDS. Math Biosci 1992;108:89-104.

138. De Gruttola V, Seage III GR, Mayer KH, Horsburgh jr CR. Infectiousness of HIV between male homosexual partners. J Clin Epidemiol 1989;42:849-56.

139. Jewell NP. Some statistical issues in studies of the epidemiology of AIDS. Stat Med 1990;9:1387-416.

140. Biglan A, Metzler CW, Wirt R, et al. Social and behavioral factors associated with high-risk sexual behaviour among adolescents. J Behav Med 1990;13:245-61.

141. Calzavara L, Coates R, Raboud J, et al. Recreational drug use and high-risk sexual behaviour in the Toronto sexual contact study. Proceeding of the VII International Conference on AIDS. Florence, Italy, 1991;1:407 (abstract MD 4071).

142. Calzavara L, Coates R, Raboud J, et al. Recreational drug use and high-risk sexual behaviour in the Toronto sexual contact study (poster made available by the author).

143. Anonymous. Current trends: self-reported changes in sexual behaviors among homosexual and bisexual men from the San Francisco City Clinic Cohort. MMWR Morb Mortal Wkly Rep 1987;36:187-9.

144. Doll LS, Judson FN, Ostrow DG, et al. Sexual behaviour before AIDS: the hepatitis B studies of homosexual and bisexual men. AIDS 1990;4:1067-73.

145. Doll LS, O'Malley PM, Pershing LA, Darrow WW, Hessol NA, Lifson AR. High-risk sexual behaviour and knowledge of HIV antibody status in the San Francisco City Clinic Cohort. Health Psychol 1990;9:253-65.

146. Fineberg HV. Education to prevent AIDS: prospects and obstacles. Science 1988;239:592-6.

147. Del-Rio A, Izazola JA, Basanez R, Palacios M, Valdespino JL, Sepulveda J. AIDS related knowledge, attitude, and infection-risk index numbers in low and high-risk practice groups. Proceeding of the VI International Conference on AIDS. S. Francisco, CA, 1990;2:269 (abstract FC 754).

148. Schneider DJ, Taylor EL, Prater LM, Wright MP. Risk assessment for HIV infection: validation study of a computer-assisted preliminary screen. AIDS Educ Prev 1991;3:215-29.

149. Gladis M, Michela J, Walter H, et al. Psychological factors contributing to adolescents perceptions of AIDS risk. Proceeding of the VI International Conference on AIDS. S. Francisco, CA, 1990;2:405 (abstract 3012).

150. Walter HJ, Vaughan RD, Gladis MM, Ragin DF, Kasen S, Cohall AT. Factors associated with AIDS risk behaviours among high school students in an AIDS epicentre. Am J Public Health 1992;82:528-32.

151. Hanson M, Kramer TH, Gross W, Quintana J, Ping-Wi L, Asher R. AIDS awareness and risk behaviours among dually disordered adults. AIDS Educ Prev 1992;4:41-51.

152. Hearst N, Hulley SB. Preventing the heterosexual spread of AIDS. Are we giving our patients the best advice? JAMA 1988;259:2428-32.

153. Hooykaas C, van der Pligt J, van Doornum GJJ, van der Linden MMD, Coutinho RA. Heterosexuals at risk for HIV: difference between private and commercial partners in sexual behaviour and condom use. AIDS 1989;3:525-532.

154. Hooykaas C, van der Linden MMD, van Doornum GJJ, van der Velde FW, van der Pligt J, Coutinho RA. Limited changes in sexual behaviour of heterosexual men and women with multiple partners in the Netherlands. AIDS Care 1991;3:21-30.

155. Iguchi MY, Rosen M, Musikoff H, et al. An index of risk factors predicting HIV seropositivity in Newark and Jersey City intravenous drug abusers not currently enrolled in treatment. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;2:221 (abstract F.C. 561).

156. Joseph JG, Montgomery S, Kessler RC, Ostrow DG, Emmons CA, Phair JP. Two-year longitudinal study of behavioral risk reduction in a cohort of homosexual men. Proceeding of the III International Conference on AIDS. Washington, DC, 1987:60 (abstract T.10.6).

157. Stall R, Ekstrand M, Pollack L, McKusick L, Coates TJ. Relapse from safer sex: the next challenge for AIDS prevention efforts. J Acquir Immune Defic Syndr 1990;3:1181-7.

158. McCusker J, Stoddard AM, McDonald M, Zapka JG, Mayer KH. Maintenance of behavioral change in a cohort of homosexually active men. AIDS 1992;6:861-8.

159. Trocki K. Preliminary results on sexual risk-taking in a general population sample. Prog Clin Biol Res 1990;325:21-5.

160. Windle M. High-risk behaviors for AIDS among heterosexual alcoholics: a pilot study. J Stud Alcohol 1989;50:503-7.

161. Joseph JG, Adib SM, Koopman JS, Ostrow DG. Behavioral change in longitudinal studies: adoption of condom use by homosexual/bisexual men. Am J Public Health 1990;80:1513-4.

162. Adib SM, Joseph JG, Ostrow DG, Tal M, Schwartz SA. Relapse in sexual behaviour among homosexual men: a 2-year follow-up from the Chicago MACS/CCS. AIDS 1991;5:757-60.

163. Adib SM, Joseph JG, Ostrow DG, James SA. Predictors of relapse in sexual practices among homosexual men. AIDS Educ Prev 1991;3:293-304.

164. Kelly JA, St. Lawrence JS, Brasfield TL, et al. Psychological factors that predict AIDS high-risk versus AIDS precautionary behavior. J Consult Clin Psychol 1990;58:117-20.

165. Kelly JA, St Lawrence JS, Betts R, Brasfield TL, Hood HV. A skills-training group intervention model to assist persons in reducing risk behaviors for HIV infection. AIDS Educ Prev 1990;2:24-35.

166. Martin JL. AIDS risk reduction recommendations and sexual behaviour patterns among gay men: a multifactorial categorical approach to assessing change. Health Educ Q 1986;13:347-58.

167. Meyer-Bahlburg HFL, Exner TM, Dellenbaugh C, El-Sadr W, Gorman JM, Ehrhardt AA. Sexual identity and sexual risk behavior in intravenous drug using men. Proceeding of the VII International Conference on AIDS. Florence, Italy, 1991;1:407 (abstract MD 4035).

168. Meyer-Bahlburg HFL, Ehrhardt AA, Exner TM, Gruen RS. Sexual risk behavior assessment schedule-adult (material made available by the author).

169. Montgomery K, Lewis CE, Kirchgraber P. Telephone screening for HIV risk of infection. Med Care 1991;29:399-407.

170. Nyamathi A. Comparative study of factors relating to HIV risk level of Black homeless women. J Acquir Immune Defic Syndr 1992;5:222-8.

171. Ostrow D, Beltran E, Wesch J, Joseph J. Recreational drug use and homosexual behaviour: the role of volatile nitrites ("poppers") in explaining the association. Proceeding of the VI International Conference on AIDS. San Francisco, CA, 1990;2:262 (abstract F.C 726).

172. Ostrow DG, VanRaden MJ, Fox R, et al. Recreational drug use and sexual behaviour change in a cohort of homosexual men. AIDS 1990;4:759-65.

173. Rapkin AJ, Erickson PI. Difference in knowledge of and risk factors for AIDS between Hispanic and non-Hispanic women attending an urban family planning clinic. AIDS 1990;4:889-99.

174. Shaw NS. Wessell LG. Risk assessment form. In: Keeling RP, ed. AIDS on the college campus. ACHA special report. Rockville, MD: American College Health Association, 1986:55-7.

175. Leapley P, Kasselman MJ. AIDS risk behaviour of entering college students. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;1:398 (abstract M.D.4034).

176. Stall R, McKusick L, Wiley J, Coates TJ, Ostrow DG. Alcohol and drug use during sexual activity and compliance with safe sex guidelines for AIDS: the AIDS behavioral research project. Health Educ Q 1986;13:359-71.

177. Leigh BC. The relationship of substance use during sex to high-risk sexual behaviour. J Sex Res 1990;27:199-213.

178. Holmes KK, Kreiss J. Heterosexual transmission of Human Immunodeficiency Virus: overview of a neglected aspect of the AIDS epidemic. J Acquir Immune Defic Syndr 1988;1:602-10.

179. Berkelman RL, Heyward WL, Stehr-Green JK, Curran JW. Epidemiology of Human Immunodeficiency Virus infection and Acquired Immunodeficiency Syndrome. Am J Med 1989;86:761-70.

180. Sloand EM, Pitt E, Chiarello RJ, Nemo GJ. HIV testing. State of the art. JAMA 1991;266:2861-6.

181. Plummer FA, Simonsen JN, Cameron DW, et al. Cofactors in male-female transmission of Human Immunodeficiency Virus. J Infect Dis 1991;163:233-9.

182. Lazzarin A, Saracco A, Musicco M, Nicolosi A, Italian Study, Group on HIV Heterosexual Transmission. Man-to-women sexual transmission of the Human Immunodeficiency Virus. Arch Intern Med 1991;151:2411-6.

183. Quinn TC, Cannon RO, Glasser D, et al. The association of syphilis with risk of Human Immunodeficiency Virus infection in patients attending sexually transmitted disease clinics. Arch Intern Med 1990;150:1297-302.

184. Holmberg SD, Horsburgh CR Jr, Ward JW, Jaffe HW. Biological factors in the sexual transmission of Human Immunodeficiency Virus. J Infect Dis 1989;160:116-25.

185. Goedert JJ, Eyster ME, Biggar RJ, Blattner WA. Heterosexual transmission of Human Immunodeficiency Virus: association with severe depletion of T helper lymphocytes in men with haemophilia. AIDS Res Hum Retroviruses 1987;3:355-61.

186. Fischl MA, Dickinson GM, Scott GB, Klimas N, Fletcher MA, Parks W. Evaluation of heterosexual partners, children and household contacts of adults with AIDS. JAMA 1987;257:640-4.

187. Laga M, Taelman H, Van der Stuyft P, Bonneux L, Vercauteren G, Piot P. Advanced immunodeficiency as a risk factor for heterosexual transmission of HIV. AIDS 1989;3:361-6.

188. European Study Group on Heterosexual Transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. Br Med J 1992;304:809-13.

189. Chin J, Mann J. Global surveillance and forecasting of AIDS. Bull WHO 1989;67:1-7.

APPENDIX - PRESENTATION OF THE EXISTING MEASURES FOR RISK OF HIV

The following format has been used for reviewing the existing measures

General

* Measure identifier number
  Number assigned to measures, listed in alphabetical order by author's name (except for alternative forms which follow the primary measure with a decimal number)
* Title
  The title of the measure used by the author. For those indexes with no exact title, a sentence found in the papers has been used
* Reference
  One or more reference
* Year
  Year of the first published paper
* Presentation
  A short description of the measure
* Factors
  The number and description of factors included in the measure
* Method of administering
  The method of administering (self-administered questionnaire, interview, computer assisted administration, telephone interview)

Development process

* Underlying method
  The method for developing the measure, according to the proposed classification (empirical, judgemental, psychometric, mathematical, ad hoc approach)
* Development process
  Few notes about the development process, the data or material used, and the assumptions made

Measure characteristics

* Type of measure
  Form of the summary measure (checklist, classification, score, probability equation)
* Purpose
  A brief description of what the measurement is meant to accomplish. The purpose of the measure is provided both in a categorical form (discriminative, evaluative, predictive) and in a conversational style, with sentences used in the paper
* Operational definition
  Definition of the risk the measure is intended to assess (risk of exposure, risk of HIV infection). This also includes whether only sexual or also drug-injecting behaviours are considered, and the time frame
* Application
  - Population to which the measure has been applied or for which it is intended, according to sexual preference (general population, homosexual, heterosexual), gender (female, male), age (youth, adults, elderly), country, and other special characteristics
  - Setting for which the measure has been developed or used (community, hospital, outpatients)

Measure quality

* Reliability
  Information available on reliability (internal consistency, test-retest, inter-observer agreement, intra-observer agreement)
* Validity
  Information available on validity (content validity, construct validity, criterion validity)

Alternative forms

Variants that cannot be considered as a separate instrument. The alternative form follows the primary measure (not in alphabetical order) and a decimal number is assigned to them (e.g. primary measure 17.1, alternative form 17.2)

General
* Measure identifier number: 1
* Title: HIV Infection Prevention Scale
* Reference:
- **Bassman LE. Object relations and self-reported AIDS self-care behavior. Psychol Rep 1991; 68:915-23.**
* Year: 1991
* Presentation
  - the sub-scale for sexual behaviour contains the following 10 items (response choices are: none, 1-19 times, 20+ times)
  a. deep kissing
  b. mutual masturbation
  c. vaginal sex with condom with spermicide
  d. vaginal sex with condom without spermicide
  e. vaginal sex without condom
  f. oral sex performed on you
  g. performing oral sex on a man without ejaculation in the mouth
  h. performing oral sex on a man with ejaculation in the mouth
  i. withdrawal (vaginal or oral sex without condom and without ejaculation)
  l. anal sex performed on you without a condom
  - another sub-scale for social behaviour with 3 items is included
  a. ask about the potential partner's sexual history
  b. delay having sex in order to get to know the partner better
  c. negotiate acceptable sexual practices to reduce the danger of getting AIDS
* Factors: three
  a. type and frequency of sexual practice
  b. condom and spermicide use
  c. type of partner (anonymous, known)
* Method of administering: self-administered questionnaire

Development process
* Underlying method: psychometric method
* Development process
  - initially, 35 items were chosen based on review of the literature on safer sex practices
  - the scale has been pretested on 150 students and items without full range response, that did not contribute to reliability, or did not reach 0.25 item-total correlation, were eliminated
  - the final scale is the sum of the score for the 13 items

Measure characteristics
* Type of measure: score
* Purpose: discriminative
  - to study the relation of risky sexual behaviour with demographic variables, knowledge, and attitudes toward AIDS
  - to study predictors (based on object relation theory) of risky sexual behaviours
* Operational definition
  - risk of exposure to HIV
  - due to sexual behaviour
  - time frame not stated
* Application
  * Population
  - heterosexual
  - males and females
  - youth (college students)
  - United States
  * Setting: college

Measure quality
* Reliability
  - Internal consistency: Cronbach alpha for the sexual behaviour sub-scale was 0.84 and for the social behaviour sub-scale 0.74
* Validity: not stated

General

* Measure identifier number: 2
* Title: Index of high-risk sexual behaviour
* Reference:

- Biglan A, Metzler CW, Wirt R, et al. Social and behavioral factors associated with high-risk sexual behavior among adolescents. J Behav Med 1990;13:245-61.

* Year: 1990
* Presentation
  - six items are considered:
  a. number of partners in the past year
  b. number of times the subject had sex with partners not well known
  c. number of times the subject had sex with an IVDU
  d. number of times the subject had sex with a partner who was having sex with other people
  e. frequency of intercourse without the use of condom
  f. engagement in anal sex
  - the z scores of these items are averaged to form a composite index
* Factors: five
  a. type of sexual practice: anal intercourse
  b. condom use
  c. frequency of intercourse
  d. type of partners (risk group or not well known)
  e. number of partners
* Method of administering: self-administered questionnaire

Development process

* Underlying method: psychometric method
* Development process
  - five factors were identified mainly from the literature as risk factors
  - the items were moderately correlated (internal consistency)
  - the means of the z scores for six items are averaged to form a composite score

Measure characteristics

* Type of measure: score
* Purpose: discriminative
  - to study the interrelationship among high-risk heterosexual behaviour to see if forming a single index was appropriate
  - to study the relationship with other problem behaviours and the social context
* Operational definition
  - risk of exposure to HIV
  - due to sexual behaviour
  - in the past year
* Application
  * Population
  - heterosexual
  - males and females
  - adolescents (grades 8 - 12)
  - United States
  * Setting: community

Measure quality

* Reliability
  - Internal consistency: the items were moderately correlated (coefficient alpha 0.77 and 0.73 for the two samples)
* Validity: not stated

General

* Measure identifier number: 3
* Title: Summary sexual activity risk score (SARS)
* Reference.

- Calzavara L, Coates R, Raboud J, et al. Recreational drug use and high-risk sexual behaviour in the Toronto sexual contact study. Proceeding of the VII International Conference on AIDS. Florence, Italy, 1991;1:407 (abstract MD 4071).

- Calzavara L, Coates R, Raboud J, et al. Recreational drug use and high-risk sexual behaviour in the Toronto sexual contact study (poster made available by the author).

* Year: 1991
* Presentation

SARS = (risk level of activity 1 x number of men with whom activity 1 has been performed) + (risk level of activity 2 x number of men with whom activity 2 has been performed) + etc...

- sexual activities and risk level assigned:

a. no risk = 0: mutual masturbation

b. minimal risk = 1: insertive fellatio (with/without condom); receptive fellatio (no ejaculate swallowed); digital-anal (insertive or receptive); anilingus (insertive or receptive); object (receptive).

c. low risk = 2: receptive fellatio (swallow ejaculate); manual-anal (receptive); anal receptive (with condom); anal insertive (with condom)

d. high risk = 10: anal receptive (without condom); anal insertive (without condom)

* Factors: three

a. type of sexual practice

b. condom use

c. number of partners

* Method of administering: interview

Development process

* Underlying method: ad hoc approach
* Development process

- the sexual activities and the risk levels are based on safer sex guidelines in wide use during the study period (reference given) and on research on risk factors for HIV infection (reference given)

Measure characteristics

* Type of measure: score
* Purpose: discriminative + evaluative

- to study the relationship between the use of recreational drugs at the time of sexual activity and high-risk sexual behaviour
- to study the change over time in behaviour

* Operational definition

- risk of exposure to HIV
- due to sexual behaviour
- in the previous 3 months

* Application

* Population
- homosexuals
- males
- adults
- individuals having a sexual partner with AIDS or ARC
- Canada
* Setting: Toronto Sexual Contact Study Cohort

Measure quality

* Reliability

- the questionnaire has been tested for reliability of self-reported data

* Validity

- the questionnaire has been tested for validity of self-reported data
* Construct validity.
- Group difference. Those individuals who were HIV positive had higher risk scores than those who were HIV negative

General

* Measure identifier number: 4
* Title: Risk indexes for HIV exposure due to sexual activities
* Reference

- **Anonymous. Current trends: self-reported changes in sexual behaviors among homosexual and bisexual men from the San Francisco City Clinic Cohort. MMWR Morb Mortal Wkly Rep 1987;36:187-9.**
- **Doll LS, Judson FN, Ostrow DG, et al. Sexual behaviour before AIDS: the hepatitis B studies of homosexual and bisexual men. AIDS 1990;4:1067-73.**
- **Doll LS, O'Malley PM, Pershing LA, Darrow WW, Hessol NA, Lifson AR. High-risk sexual behaviour and knowledge of HIV antibody status in the San Francisco City Clinic Cohort. Health Psychol 1990;9:253-65.**

* Year: 1987
* Presentation
    - risk indexes are calculated for any combination of type of partner and sexual activities.
    - the number of partners the individual had in the past 4 months is multiplied by the average percentage of time the participant engaged in the sexual behaviour.
    a. sexual activities: insertive anal intercourse, receptive anal intercourse with ejaculation, receptive anal intercourse without ejaculation, receptive oro-genital, insertive oro-genital contact
    b. type of partner: steady (one or two contacts) and non-steady (three or more contacts)
* Factors: two
    a. number of partners
    b. type and frequency of sexual practice
* Method of administering: interview

Development Process

* Underlying method: ad hoc approach
* Development process: not stated

Measure characteristics

* Type of measure: score
* Purpose: discriminative + evaluative
    - to assess frequency of risk behaviour
    - to evaluate changes over time in risk behaviours (CDC)
    - to evaluate changes in relation with knowledge of HIV status
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous 4 months
* Application
    * Population
    - homosexual
    - males
    - adults
    - United States
    * Setting: STDs clinics

Measure quality

* Reliability
    - The authors (Doll, Health Psych) stated that no information is available on the reliability or validity of the self-reported data

* Validity
    * Construct validity
    - Group difference. Seropositive individuals had significantly higher baseline risk indices than did seronegative individuals on receptive anal intercourse with non-steady partners ($p < 0.0001$), on receptive anal intercourse with steady partners ($p < 0.01$) and on insertive anal intercourse with steady partners ($p < 0.004$)

General
* Measure identifier number: 5
* Title: Expected probability of infection
* Reference.
- **Crowe SM, Elbeik T, Ulrich PP, Mills J, Moss A. Lack of evidence of occult Human Immunodeficiency Virus in seronegative individuals at very high risk of infection. J Med Virol 1991;35:160-4.**
* Year: 1991
* Presentation

$$\text{- expected probability of being positive } p = \frac{1}{1 + \exp^{-A}}$$

- where $A = -3.30 + 0.79x_1 + 1.78x_2 + 2.32x_3 + 1.07x_4 + 0.93x_5 + 1.00x_6 + 1.41x_7 + 0.82x_8 + 0.98x_9$
- variables:
  - x_1-x_3 = # of receptive anal partners (1-3, 4-19, 20-39, 40+)
  - x_4 = past syphilis
  - x_5 = past giardiasis
  - x_6-x_7 = use of nitrates (ever, <65, >65 times per month)
  - x_8 = use of enemas
  - x_9 = partners with AIDS

* Factors: six
  a. number of partners
  b. type of partner (AIDS case)
  c. history of syphilis
  d. history of giardiasis
  e. use of nitrates
  f. use of enemas
* Method of administering: not stated

Development process
* Underlying method: empirical method
* Development process
  - cohort study of [illegible] sexual men
  - through logistic regression the risk factors for seropositivity were identified and the coefficients estimated

Measure characteristics
* Type of measure: probability equation
* Purpose: discriminative
  - to calculate the expected probability of being positive
* Operational definition
  - risk of HIV infection
  - due to sexual behaviour
  - not stated the time frame
* Application
  * Population
  - homosexual
  - males
  - adults
  - United States
  * Setting: San Francisco General Hospital Cohort

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 6
* Title: HIV risk-taking behaviour scale (HRBS)
* Reference
- **Darke S, Hall W, Heather N, Heather M, Ward J, Wodak A. The reliability and validity of a scale to measure HIV risk-taking behaviour among intravenous drug users. AIDS 1991;5:181-5.**
* Year: 1991
* Presentation
- the sub-scale for sexual behaviour contains 5 items for current sexual behaviour:
a. number of sexual partners in the previous month
b. condom use with casual partners
c. condom use with regular partners
d. condom use in paid sex
e. participation in anal intercourse
- each item is scored (0-5 scale) and the scores are added
- another sub-scale with 6 items for injecting behaviour is available
* Factors: four
a. number of partners
b. type of partners
c. condom use
d. type of sexual practice: anal sex
* Method of administering: interview

Development process
* Underlying method: psychometric approach
* Development process
- each item has been chosen to address a specific current HIV risk-taking behaviour
- the authors state they will apply weight when knowledge about relative risk is available

Measure characteristics
* Type of measure: score
* Purpose: discriminative + evaluative
- to measure and screen for the overall current HIV-related risky behaviours
- to measure behavioural change
* Operational definition
- risk of exposure to HIV
- due to sexual and injecting behaviours
- in the past month
* Application
* Population
- heterosexual
- males and females
- adults
- IVDUs
- Australia
* Setting: clinics

Measure quality
* Reliability
- Internal consistency: Cronbach's coefficient alpha 0.70
- Test-retest: Pearson product-moment correlation coefficient between the two interviews is 0.86
- Inter-observer agreement: Test-retest with the same or different interviewers, correlation coefficient r = 0.87 and 0.85
* Validity
- validity of the self-reported data. Percentage agreement between the subjects and their collaterals ranges for different items between 81.8-100%
* Content validity
- Factorial validity. Explored by submitting scores on the items to principal components analysis with varimax rotation. The general scale reflects two ways of transmission (due to sexual and injecting behaviours)
* Criterion validity
- Predictive validity. A longitudinal cohort is planned to validate whether the score predicts seropositivity

General
* Measure identifier number: 7
* Title: Probability of HIV infection
* Reference
- **Elsenberg B. The number of partners and the probability of HIV infection. Stat Med 1989;8:83-92.**
* Year: 1989
* Presentation
  * Probability of becoming infected for a given number of contacts = $1 - \prod_{j=1}^{m} [1 - P + P(1 - b)^{nj}]$
  - where:
    (P) probability that a randomly selected partner is infected
    (b) probability of infection per sexual contact
    (n) number of contacts
    (m) number of partners
    (Π) product over each partner
  * In the case of n contacts is found the following ranking from least risk to greatest risk
    a. a monogamous relationship with a non-infected partner
    b. a monogamous relationship with a randomly selected partner
    c. m randomly selected partners (with m less or equal to n)
    d. n randomly selected partners (one partner per contact)
    e. a monogamous relationship with an infected partner
    - for small n, P has great importance, while m is not relevant
    - for large n, the number of partner has more importance
  * Probability of becoming infected over a period (t) of time
    - for one randomly selected partner = $1 - (1-P) + P \exp^{-bl}$
    - for different partners each time = $1 - \exp^{-Pbl}$
  - where:
    (P) probability that a randomly selected partner is infected
    (b) probability of infection per sexual contact
    (t) time period
    (l) expected number of contacts per unit time
* Factors: five
  a. frequency of intercourse
  b. number of partners
  c. type of partner (one time partner, negative or positive or unknown HIV serostatus partner)
  d. sexual practice: one at the time (infectivity)
  e. prevalence of HIV infection
* Method of administering: not stated

Development process
* Underlying method: mathematical method
* Development process
  * Assumptions for the model for a given number of contacts:
  a. binomial model
  b. partners (m) are independent trial, each with probability P of being infected
  c. contacts (n) with a given partner are independent trials
  d. probability (P) that a randomly selected partner is infected is constant over time
  e. infectivity (b) is constant for different partners
  * Assumptions for the model for a given period of time:
  a. Poisson model
  b. contacts are independent trials
  c. probability (P) that a randomly selected partner is infected is assumed constant over time
  d. probability of infection (b) per sexual contact is assumed constant

Measure characteristics
* Type of measure: probability equation + classification
* Purpose: predictive
  - to clarify the relative importance of the number of sexual contacts versus the number of different sexual partners.
  - to provide a formula to assess the risk of sexual behaviour with regard to AIDS
* Operational definition:
  - risk of HIV infection
  - due to sexual behaviour
  - after n contacts or over a period of time
* Application
  * Population
  - general population
  * Setting: not stated

Measure characteristics
* Reliability: not stated
* Validity
  - author states that the model needs data for goodness-of-fit test
  - the assumptions are discussed in light of what is known and of the empirical data

General
* Measure identifier number: 8
* Title: Behavioral risk score
* Reference
- Elmslie T, McDowell I, Wells G, Hollingworth G. Sexual behaviour related to transmission of HIV infection among women in the Canadian primary care setting. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;1:409 (abstract M.D.4078).
- Elmslie T, Wells G, Hollingworth G, McDowell I, Graham J. Study of sexual behaviour associated with transmission of HIV infection among women in the primary care setting. Final report. April 1992 (unpublished materials).
* Year: 1991
* Presentation
    * one partner model
    - factors and levels in the final model are:
    a. protection (0= always condom, 1= never/sometimes)
    b. practice (0= vaginal, 1= anal)
    c. partner (0= no partner, 1=heterosexuals/no IVDU, 2= hetero/unknown IVDU, 3=hetero/IVDU, 4= haemophiliac or IVDU or bisexual, 5= HIV positive)
    - final score = 1.4 + 1.3 a + 0.4 b + 0.6 c_1 (or 1.7 c_2; or 2.1 c_3; or 2.9 c_4; or 3.9 c_5) + a x 0.7 c_1 (or 0.9 c_2; or 1.4 c_3; or 1.7 c_4; or 2.0 c_5)
    * multiple partners model
    - score = 0.9 + 0.5 (highest single risk score) + 0.4 (highest risk partners) + 0.8 (overall protection) + 0.2 (second highest risk score)
* Factors: four
    a. condom use
    b. type of sexual practice
    c. type of partner (risk groups)
    d. number of partners
* Method of administering: self-administered questionnaire

Development process
* Underlying method: judgemental method (derived approach, factorial experimental design)
* Development process
    * One partner model
    - five factors (with levels for each) were selected for one partner situation by the investigators
    - a derived approach was chosen, which require scenarios to be rated by the judges
    - since the possible number of possible scenarios was 216, the investigators excluded some of the factors and levels: three factors were retained for a total of 32 scenarios
    - twenty-five experts from across Canada were asked to rate the scenario
    - the median risk scores attributed by the rates for each scenario were tabulated
    - using ANOVA to weight the factors, three models were produced which were each compared to the raters' mean score to arrive at the best fitting model for the scoring system
    * multiple partner situation
    - since the possible number of scenarios were 725, a sample of 2, 3 and 4 partner scenarios were rated
    - multiple regression analysis was used, with the raters' global risk assessment of the multiple partner scenarios as dependent variables and several independent variables
    - the model which best fit the raters' responses has been chosen

Measure characteristics
* Type of measure: score
* Purpose: discriminative
    - to assess the prevalence of sexual behaviour associated with the risk of HIV transmission
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous year and in the previous 10 years
* Application
    * Population
    - heterosexual
    - females
    - adults (16-35 years)
    - Canada
    * Setting: primary care setting

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 9
* Title: Level of HIV risk for heterosexual males
* Reference.
- **European Study Group on heterosexual transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. Br Med J 1992;304:809-13.**
* Year: 1992
* Presentation
    - two risk factors are considered: advanced stage of HIV infection in the female infected partner (T4 more/less than 200); sexual contact during menses (never/at least once).
    - people are classified by levels of risk:
    level 1. none of the risk factors
    level 2. one of the risk factors
    level 3. both risk factors
* Factors: two
    a. infectiousness of the index cases
    b. sexual contact during menses
* Method of administering: interview

Development process
* Underlying method: empirical method
* Development process
    - cohort study of heterosexual couples (female index cases and male stable partners)
    - through logistic regression analysis, risk factors for HIV transmission were identified

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - to classify according to level of risk
* Operational definition
    - risk of HIV infection (transmission with a positive partner)
    - due to sexual behaviour
    - not stated the time frame
* Application
    * Population
    - heterosexual
    - males
    - adults
    - stable partner of an infected women
    - Europe
    * Setting: hospital wards, outpatients clinics, health department

Measure quality
* Reliability: not stated
* Validity
    * Criterion validity
    - Concurrent validity (HIV positive status as gold standard): estimate of HIV prevalence for each level of risk were:
    level 1 = 1% prevalence
    level 2 = 16% prevalence
    level 3 = 57% prevalence

General
* Measure identifier number: 10
* Title: Level of HIV risk for heterosexual females
* Reference.
- **European Study Group on heterosexual transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. Br Med J 1992;304:809-13.**
* Year: 1992
* Presentation
    - three risk factors are considered: advanced stage of HIV infection in the male infected partner (T4 more/less than 200); anal sex (never/at least once); advanced age of female susceptible partner (more/less than 45 years).
    - people are classified by levels of risk:
    level 1. none of the three risk factors
    level 2. one of the risk factors
    level 3. two out of three risk factors
* Factors: three
    a. infectiousness of the index case
    b. sexual practice: anal intercourse
    c. age of the partner
* Method of administering: interview

Development process
* Underlying method: empirical method
* Development process
    - cohort study of heterosexual couples (male index cases and female stable partners)
    - through logistic regression analysis, risk factors for HIV transmission were identified

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - to classify according to level of risk
* Operational definition
    - risk of HIV infection (transmission with positive partner)
    - due to sexual behaviour
    - time frame not stated
* Application
    * Population
    - heterosexual
    - females
    - adults
    - stable partner of an infected male
    - Europe
    * Setting: hospital wards, outpatients clinics, health department

Measure quality
* Reliability: not stated
* Validity
    * Criterion validity
    - Concurrent validity (HIV positive status as gold standard): estimated HIV prevalence for each level of risk were:
    level 1 = 10% prevalence
    level 2 = 31% prevalence
    level 3 = 54% prevalence

General
* Measure identifier number: 11.1
* Title: Cumulative probability of HIV infection
* Reference.
- **Fineberg HV. Education to prevent AIDS: prospects and obstacles. Science 1988;239:592-6.**
- **Del-Rio A, Izazola JA, Basanez R, Palacios M, Valdespino JL, Sepulveda J. AIDS related knowledge, attitude, and infection-risk index numbers in low and high-risk practice groups. Proceeding of the VI International Conference on AIDS. S. Francisco, CA, 1990;2:269 (abstract FC 754).**
* Year: 1988
* Presentation
  - probability of infection in 1000 sexual exposure = $1 - \{P(1 - b)^{NC} [1 - (1 - c)b]^{C} + (1 - P)\}^{m}$
  - where:
    (b) risk of infection by single exposure (0.01)
    (c) reduction risk per exposure through condom (0.90)
    (P) prevalence of HIV among potential partners
    (C) number of exposures per partner using a condom [total number of exposure (1000) divided by the number of partners (m), the result multiplied by the frequency of condom use (0, 0.5, 1)]
    (NC) number of exposures per partner without a condom [total number of exposure (1000) divided by the number of partners (m), the result multiplied by the frequency of condom use (1, 0.5, 0)]
    (m) number of partners
* Factors: five
    a. sexual practice: one at the time (infectivity)
    b. number of partners
    c. frequency of intercourse
    d. frequency of condom use
    e. prevalence of HIV infection (probability of selecting an infected partner)
* Method of administering: not stated

Development process
* Underlying method: mathematical method
* Development process
    - Assumptions:
    a. binomial theorem
    b. selection of partners (m) is random with respect to their probability of being infected
    c. sexual contacts (C, NC) are assumed independent
    d. prevalence among partners is constant
    e. infectivity is constant

Measure characteristics
* Type of measure: probability equation
* Purpose: predictive
    - to discriminate between groups (Del Rio)
    - to find priorities of educational messages by modelling the risk of various combinations of number of partners, prevalence of infection, and frequency of condom use (Fineberg)
    - to study infection risk, for policy definition, and educational strategy design (Del Rio)
* Operational definition
    - risk of HIV infection
    - due to sexual behaviour
    - after 1000 exposure
* Application
    * Population
    - general population
    - Mexico (Del Rio)
    * Setting: not stated

Measure quality
* Reliability: not stated
* Validity
    * Construct validity
    - Group difference (Del Rio). A modified version of the model (not available) was used with data from a survey on low-risk sexual practice groups (general population, graduate level students, health personnel) and high-risk sexual practices groups (female prostitutes and homo/bisexual men) in Mexico. The model initially pointed out that female prostitutes had a significantly higher value. When the prevalence was modified according to serological studies among homosexuals in Mexico City, the model also discriminated homosexuals.

Alternative forms
* Measure identifier number: 11.2
* Title: Risk assessment for HIV infection
* Reference.
- Schneider DJ, Taylor EL, Prater LM, Wright MP. Risk assessment for HIV infection: validation study of a computer-assisted preliminary screen. AIDS Educ Prev 1991;3:215-29.
* Year: 1991
* Presentation
    * Probability of infection
    - variables for sexual history (segmented in three year intervals) are: number of partners, racial characteristics of partners, frequency of sexual events, frequency of receptive anal intercourse, city of residence of the partners, frequency of condom use.
    - information is combined with expert data on:
        - per-event infectivity for vaginal intercourse (0.002) and for anal intercourse (0.005). No other sexual activities are considered at risk
        - local estimates of HIV prevalence by ethnic group category (using the CDC assumption that the distribution of AIDS incidence is an indicator of prevalence).
        - condom use failure 10% (referenced) and 30% failure for teenagers
    - probability of infection = $1 - [P(1 - b)^{NC} (1 - \{1 - c\}b)^{C} + (1 - P)]^{m}$
    - where:
        (b) risk of infection by single exposure (0.002 - 0.005)
        (c) reduction risk per exposure through condom (0.90-0.70)
        (P) prevalence of HIV among potential partners
        (C) number of exposure per partner using a condom
        (NC) number of exposure per partner without a condom
        (m) number of sexual partners
    - Cut-off for high/low risk = 0.01

    * Checklist for risk factors for AIDS (not supported by a method):
    - respondents are also declared at risk if one or more of the following conditions are present: blood transfusions, needle sharing, (for females) sexual relations with men at risk (bisexuals, IVDUs, or haemophiliacs), (for males) sexual relations with women at risk (residents of the 20 largest US cities, prostitutes, IVDUs, haemophiliacs, sexual partners of a male at risk)

* Factors: 5 + 3 = 8
    - factors included in probability estimate are:
    a. sexual practice: anal and vaginal (infectivity)
    b. number of partners
    c. frequency of intercourse
    d. frequency of condom use
    e. prevalence of HIV infection (probability of selecting an infected partner)
    - factors included in the checklist for risk of AIDS
    a. blood transfusions
    b. needle sharing
    c. heterosexual relation with a partner in risk group
* Method of administering: computer assisted administration

Development process
* Underlying method: see the previous measure
* Development process: see the previous measure

Measure characteristics
* Type of measure: probability equation combined with checklist
* Purpose: predictive
    - to screen for HIV risk, voluntary preliminary screening to alert and advice to seek antibody testing, and avoiding false positive tests
    - general education and personal risk assessment
* Operational definition:
    - risk of HIV infection
    - due to sexual behaviour
    - in a certain number of sexual exposures
* Application
    * Population
    - general population
    * Setting: not stated

Measure quality
* Reliability
    - readability and facility: 77 individuals were asked to fill out a short hard-copy questionnaire to evaluate the software.
* Validity
    * Criterion validity: concurrent (antibody status as gold standard)
    - High risk sample: 70 participants, clientele of an AIDS antibody testing and counselling centre completed the computer interview, which was linked to antibody status. Sensibility 96.6%; specificity 19.5%
    - Low risk sample: 74 students completed the interview and were asked to seek antibody testing, plus 17 individuals from US military reserve unit, whose antibody status was known. Sensibility: nobody was HIV+; specificity 87.5%

General

* Measure identifier number: 12
* Title: AIDS behaviour index scores
* Reference.

- Gladis M, Michela J, Walter H, et al. Psychological factors contributing to adolescents perceptions of AIDS risk. Proceeding of the VI International Conference on AIDS. S. Francisco, CA, 1990;2:405 (abstract 3012).

- Walter HJ, Vaughan RD, Gladis MM, Ragin DF, Kasen S, Cohall AT. Factors associated with AIDS risk behaviours among high school students in an AIDS epicentre. Am J Public Health 1992;82:528-32.

* Year: 1990
* Presentation
  - participants are classified with the following seven categories:
  0. sexual abstinence and no IVDU
  1. intercourse consistently using a condom with one low-risk partner and no IVDU
  2. intercourse consistently using a condom with two or more low-risk partners and no IVDU
  3. intercourse with one low-risk partner without or inconsistently using a condom and no IVDU
  4. intercourse with two or more low-risk partners without or inconsistently using a condom and no IVDU
  5. intercourse consistently using a condom with one or more high-risk partners and no IVDU
  6. sexual intercourse with one or more high-risk partners without or inconsistently using a condom, or illicit drug use or history of an STD
* Factors: four sexual + one IVDU = five
  a. type of sexual practice: generic sexual intercourse
  b. condom use
  c. type of partner (high/low risk)
  d. number of partners
  e. (IVDU)
* Method of administering: self-administered questionnaire

Development process

* Underlying method: mathematical method (see Hearst model, measure 14)
* Development process
  - five factors were selected according to the model (Hearst), and 7 categories were created

Measure characteristics

* Type of measure: classification
* Purpose: discriminative
  - to ascertain the prevalence of behaviours
  - to investigate the association between beliefs and behaviours
  - to determine predictors of behaviours
* Operational definition
  - risk of exposure to HIV
  - due to sexual and injecting behaviour
  - in the past year
* Application
  * Population
  - general population
  - males and females
  - adolescents (10th grade students)
  - United States
  * Setting: high school

Measure quality

* Reliability
  - Internal consistency: high agreement (98%) of responses to similarly worded items assessing the same behavioural domains that were dispersed throughout the survey
* Validity
  * Construct validity
  - convergent validity: similarity of the behaviour prevalence rates reported in this study to those reported in other studies

General
* Measure identifier number: 13
* Title. Sexrisk Test
* Reference.
- **Hanson M, Kramer TH, Gross W, Quintana J, Ping-Wi L, Asher R. AIDS awareness and risk behaviours among dually disordered adults. AIDS Educ Prev 1992;4:41-51.**
* Year: 1992
* Presentation
  - seven item scale: respondents are asked to report on the frequency with which they previously or currently engaged in sexual activities such as anal sex, sex with prostitutes, sex with multiple partners, sex with drug users.
  - the responses are scored on a frequency scale
  - the final scale has a potential range of 7 to 42
* Factors: four
  a. type of sexual practice: anal intercourse
  b. frequency of intercourse
  c. type of partner
  d. multiple partners
* Method of administering: interview

Development process
* Underlying method: psychometric method
* Development process: not stated

Measure characteristics
* Type of measure: score
* Purpose: discriminative
  - to describe the prevalence of high-risk behaviours for HIV in the population
  - to study the association with knowledge and disorders
* Operational definition:
  - risk of exposure to HIV
  - due to sexual behaviour
  - time frame not stated
* Application
  * Population
  - heterosexual
  - males and females
  - adults
  - dually disordered for drug abuse and mental disorders.
  - United States
  * Setting: hospital-based outpatient clinic

Measure quality
* Reliability
  * Internal consistency. Standardized alpha coefficient = 0.71
* Validity: not stated

General
* Measure identifier number: 14
* Title: Risk of HIV infection for heterosexual intercourse
* Reference.
- **Hearst N, Hulley SB. Preventing the heterosexual spread of AIDS. Are we giving our patients the best advice? JAMA 1988;259:2428-32.**
* Year: 1988
* Presentation
  - probability of infection (500 contacts = 4 1/2 years) = $P \times [1 - (1 - b \times c)^n]$
  - where:
  (P) prevalence of infection among potential partners (estimates, referenced, are given for people with unknown status not in high risk group and in high risk group; people with negative status with no history of high risk behaviour, and with continuing high-risk behaviour; and for seropositive individuals)
  (b) infectivity from a single vaginal contact (0.002 referenced; it is advise that other activities are less or more infectivity, and that other factors can influence the infectiousness and susceptibility, and that infectivity of the virus may be different)
  (c) reduction of infectivity through condom use by a factor of 10 (referenced)
  (n) number of contacts with the same partner (fixed at 500 as the number of contacts a couple has in four and a half years, which is the median time between infection and development of symptoms)
* Factors: three
  a. prevalence of HIV infection
  b. sexual practice (infectivity)
  c. condom use (reduction in risk)
* Method of administering: not stated

Development process
* Underlying method: mathematical method
* Development process
  - Assumption:
  a. binomial model
  b. contacts (n) are independent trial
  c. infectivity (b) per contact is constant
  d. prevalence (P) is constant within each group
  - the estimates represent approximate group mean

Measure characteristics
* Type of measure: probability equation
* Purpose: predictive
  - to find the priorities in the recommendations for controlling the spread of sexually transmitted HIV by modelling the risk with different estimates of prevalence
* Operational definition:
  - risk of HIV infection
  - due to sexual behaviour
  - after 500 contacts
* Application
  * Population
  - heterosexual
  - males and females
  - adults
  - United States
  * Setting: not stated

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 15
* Title: Score for HIV risk
* Reference.
- **Hooykaas C, van der Pligt J, van Doornum GJJ, van der Linden MMD, Coutinho RA. Heterosexuals at risk for HIV: difference between private and commercial partners in sexual behaviour and condom use. AIDS 1989;3:525-532.**
- **Hooykaas C, van der Linden MMD, van Doornum GJJ, van der Velde FW, van der Pligt J, Coutinho RA. Limited changes in sexual behaviour of heterosexual men and women with multiple partners in the Netherlands. AIDS Care 1991;3:21-30.**
* Year: 1989
* Presentation
    - scores are calculated separately for private and commercial partners
    - the frequency of practising sexual techniques (vaginal intercourse) is measured on a five point scale (1=never, 2=sometimes, 3=half, 4=often, 5=always)
    - condom use is measured on a five point scale (5=never, 4=sometimes, 3=half, 2=often, 1=always)
    - the five point frequency scales are transformed into five-point interval scales ranging from 0 to 1 (0, 0.25, 0.5, 0.75, 1).
    - the transformed frequency score for sexual techniques is multiplied by the transformed score for condom use and by the number of partners in the preceding 4 months.
    - higher score indicates greater risk for HIV infection.
* Factors: three
    a. number of sexual partners
    b. frequency of sexual practice
    c. condom use
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process
    - risk factors are found in the literature
    - three factors are selected

Measure characteristics
* Type of measure: score
* Purpose: discriminative + evaluative
    - to describe the sexual behaviour of a high risk heterosexual group
    - to study changes in sexual behaviours
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the past 4 months
* Application
    * Population
    - heterosexual
    - males and females
    - adults
    - high risk group (more than four partners of the opposite sex in the preceding 6 months: mostly prostitutes and prostitutes' clients)
    - Netherlands
    * Setting: STDs clinics

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 16.1
* Title. Clinical risk index for seropositivity
* Reference.
- **Iguchi MY, Rosen M, Musikoff H, et al. An index of risk factors predicting HIV seropositivity in Newark and Jersey City intravenous drug abusers not currently enrolled in treatment. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;2:221 (abstract F.C. 561).**
* Year: 1991
* Presentation
    - variables identified as risk factors for seropositivity are:
    a. years of intra-venous drug use
    b. frequency of injecting heroin and cocaine together in the past 6 months
    c. frequency of injecting cocaine in the past 6 months
    d. number of sexual partners in the past 6 months (inversely related)
    e. highest grade of schooling (inversely related)
    - each risk factor is dichotomized in high/low risk
    - the risk index value, ranging from 0 to 5, is equal to the number of risk factors in which an individual is categorized as being in the higher risk group level
* Factors: 1 for sexual behaviour
    a. number of sexual partners (inversely related)
* Method of administering: interview

Development process
* Underlying method: empirical method
* Development process
    - 500 variables related to risk behaviour from the NIDA AIDS Initial Assessment v. 8.0 were examined
    - using forward and backward regressions, 5 variables were identified as significantly and independently related to HIV seropositivity

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - to identify IVDUs at highest risk for developing HIV, possibly allowing for effective treatment allocation
* Operational definition:
    - risk of HIV infection
    - due to sexual and injecting behaviour
    - at the present
* Application
    * Population
    - general population
    - males and females
    - adults
    - IVDU not in treatment
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity
    * Criterion Validity (antibody testing as gold standard)
    - concurrent validity: relative risk for testing HIV positive are:
        - 2 factors / 0-1 factor = RR 1.5
        - 3 factors / 0-1 factor = RR 3.4
        - 4 factors / 0-1 factor = RR 6.0
        - 5 factors / 0-1 factor = RR 10.8

Alternative form

* Measure identifier number: 16.2
* Title: Clinical risk index for seropositivity
* Reference.

- Iguchi MY, Rosen M, Musikoff H, et al. Predictors of HIV seropositivity in Newark and Jersey City i.v. drug users not currently enrolled in treatment. NIDA Res Monogr 1991;105:473-4.

* Year: 1991
* Presentation
    * variables with the standardized discriminant function coefficient are:
    - Intra-venous drug use variables:
    a. years of intra-venous drug use (.30)
    b. frequency of injecting heroin and cocaine together in the past 6 months (.24)
    c. frequency of injecting cocaine in the past 6 months (.20)
    - Non intra-venous drug use variable are:
    d. use of crack cocaine in the past 6 months (-.27)
    e. non-injection heroin use in the past 6 moths (-.21)
    - Health-related variables:
    f. individuals assessment of their own risk of developing AIDS (.37)
    g. reporting more than one of eight HIV-risk associated health problems in their lifetime (endocarditis, syphilis, tuberculosis, pneumonia, hepatitis, genital herpes, gonorrhoea, chlamydia)(.14)
    - Other variables:
    h. number of times in jail in the past 5 years (.29)
    i. history of abusing glue/paint (.23)
    j. presence of sexual partners in the past 6 months (-.22)
    - Demographic variables:
    k. completion of high school (-.18)
    l. being non-black (-.15)
    m. being female (.12)
    * employing a discriminant classification analysis with the 13 variables found to be significant, individuals are classified as high or low risk
    * using the discriminant function coefficients, a clinical risk index is calculated for each case, ranging from -9.27 to 28.98
* Factors: 1 for sexual behaviour
    a. presence of sexual partners (inversely related)
* Method of administering: interview

Development process

* Underlying method: empirical method
* Development process
    - 500 variables related to risk behaviour from the NIDA AIDS Initial Assessment v. 8.0 were examined
    - 200 variables were found to be significantly associated with HIV serostatus (univariate)
    - Using both logistic regression and discriminant function analysis with a Bonferroni adjustment, 13 variables were identified as significantly and independently related to HIV seropositivity

Measure characteristics

* Type of measure: score + classification
* Purpose: discriminative
    - to identify IVDUs at highest risk for developing HIV, possibly allowing for effective treatment allocation
* Operational definition:
    - risk of HIV infection
    - due to sexual and injecting behaviour
    - at the present
* Application
    * Population
    - general population
    - males and females
    - adults
    - IVDU not in treatment
    - United States
    * Setting: community

Measure quality

* Reliability: not stated
* Validity
    * Criterion validity
    - concurrent (HIV test as gold standard). The classification analysis applied on the source data, including 1278 subjects, provides a sensibility of 72.9% and a specificity of 68.1%
    - prospective: index effectiveness is currently being studied in a prospective study
    * Construct Validity
    - group difference: the relative odds ratio for testing HIV positive between the 94 cases assigned a clinical risk score equal to or more than 20.0 (80 HIV+, 14 HIV-) and the 100 cases assigned a clinical risk score of less than 1.0 (15 HIV+, 85 HIV-) was 40.82.

General
* Measure identifier number: 17.1
* Title: Four level objective risk index
* Reference.
- **Joseph JG, Montgomery S, Kessler RC, Ostrow DG, Emmons CA, Phair JP. Two-year longitudinal study of behavioral risk reduction in a cohort of homosexual men. Proceeding of the III International Conference on AIDS. Washington, DC, 1987:60 (abstract T.10.6).**
- **Stall R, Ekstrand M, Pollack L, McKusick L, Coates TJ. Relapse from safer sex: the next challenge for AIDS prevention efforts. J Acquir Immune Defic Syndr 1990;3:1181-7.**
- **McCusker J, Stoddard AM, McDonald M, Zapka JG, Mayer KH. Maintenance of behavioral change in a cohort of homosexually active men. AIDS 1992;6:861-8.**
* Year: 1987
* Presentation
    - people are classified by one of the four categories:
    1. no risk = celibate
    2. low risk = only anal sex with a condom within a monogamous relationship or, if not monogamous, no anal sex at all
    3. modified high-risk = only anal sex without a condom within a monogamous relationship or, if not monogamous, have anal sex only while using condoms
    4. high risk = anal sex without condoms outside of monogamous relationships.
* Factors: three
    a. type of sexual practice: only anal intercourse
    b. condom use
    c. number of partners (monogamous/polygamous)
* Method of administering: self-administered questionnaire (McCusker)

Development process
* Underlying method: ad hoc approach
* Development process
    - the main reference, Joseph et al, is an abstract and does not contain enough information. It is said that the index was constructed and validated using HIV serological data from the Chicago Multicenter AIDS Cohort Study Group. The secondary reference used a modified version of the scale.

Measure characteristics
* Type of measure: classification
* Purpose: discriminative + evaluative
    - to evaluate change in the patterns of sexual risk taking (Stall)
    - to study predictors of risk reduction and relapse from safer sex (Stall)
    - to study association between risk perception and actual risk (McCusker)
    - to study risk perception as predictors of behavioural change (McCusker)
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the past 30 days
* Application
    * Population
    - homosexual
    - males
    - adults
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated other than that included in the abstract

Alternative forms
* Measure identifier number: 17.2
* Title: Sexual risk-taking categories
* Reference.
- **Trocki K. Preliminary results on sexual risk-taking in a general population sample. Prog Clin Biol Res 1990;325:21-5.**
* Year: 1990
* Presentation
    - individuals are classified by one of the following categories:
    1. low risk = celibate
    2. modified low risk = monogamous relationship and always use condoms;
    3. modified high 1 = monogamous but not always use condoms
    4. modified high 2 = non monogamous but always use condoms;
    5. high risk = not always use condoms and have multiple partners or have a partner who has had other sexual partners in the last year.
* Factors: two
    a. condom use
    b. number of partners (monogamous/polygamous)
* Method of administering: interview

Development process
* Underlying method: see previous measure
* Development process
    - categories are roughly equivalent to categories created in studies of homosexual populations

Measure characteristics
* Type of measure: classification
* Purpose: discriminative
    - to study the relationship to alcohol use
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous 12 months
* Application
    * Population
    - heterosexuals
    - males and females
    - adults
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

Alternative forms
* Measure identifier number: 17.3
* Title: Sexual risk status
* Reference
- **Windle M. High-risk behaviors for AIDS among heterosexual alcoholics: a pilot study. J Stud Alcohol 1989;50:503-7.**
* Year: 1989
* Presentation
    - individuals are classified by one of the following categories:
    1. celibate = no sexual activities
    2. low-risk = only protected sex
    3. moderate-risk = unprotected sex within a monogamous relationship
    4. high-risk = unprotected sex with no primary partners
* Factors: two
    a. condom use
    b. number of partners (monogamous/ polygamous)
* Method of administering: interview

Development process
* Underlying method: see previous measure
* Development process
    - modification of the categorical scheme

Measure characteristics
* Type of measure: classification
* Purpose: discriminative
    - to investigate differences in drug use for individuals categorized according to sexual-risk status
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous 6 months
* Application
    * Population
    - heterosexual
    - males and females
    - adults
    - United States
    - alcoholic inpatients
    * Setting: hospital

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 18
* Title: Index for Receptive Anal Sex (RAS) and Insertive Anal Sex (IAS)
* Reference.
- Joseph JG, Adib SM, Koopman JS, Ostrow DG. Behavioral change in longitudinal studies: adoption of condom use by homosexual/bisexual men. Am J Public Health 1990;80:1513-4.
- Adib SM, Joseph JG, Ostrow DG, Tal M, Schwartz SA. Relapse in sexual behaviour among homosexual men: a 2-year follow-up from the Chicago MACS/CCS. AIDS 1991;5:757-60.
- Adib SM, Joseph JG, Ostrow DG, James SA. Predictors of relapse in sexual practices among homosexual men. AIDS Educ Prev 1991;3:293-304.
* Year: 1990
* Presentation
- index for Receptive Anal Sex (RAS) and Insertive Anal Sex (IAS) are computed separately
- the classification categories are (in order of probable safety)
0. celibate
1. not engaging in anal sexual practices
2. consistently using condoms in anal sex practices
3. inconsistently using condoms
4. never using condoms
- categories 1 and 2 are considered safe sexual practices;
- categories 3 and 4 are considered unsafe sexual practices;
* Factors: two
a. type of sexual practice: receptive or insertive anal sex
b. condom use
* Method of administering: self-administered questionnaire

Development process
* Underlying method: ad hoc approach
* Development process: not stated

Measure characteristics
* Type of measure: classification
* Purpose: discriminative + evaluative
- to describe prevalence of safer sexual practices
- to evaluate pattern of change in sexual practices
- to identify psychological predictors of relapse in sexual practices
* Operational definition
- risk of exposure to HIV
- due to sexual behaviour
- in the past 30 days
* Application
* Population
- homosexual
- males
- adults
- United States
* Setting: community

Measure quality
* Reliability: not stated
* Validity
* Construct validity
- convergent validity. HIV seroconversion was assumed to occur within six months of exposure. The rates of HIV seroconversion found are:
a. for celibate = 0;
b. safer sexual practices (classes 1 and 2) = 1.05 per 100 person-year;
c. unsafe sexual practices (classes 3 and 4) = 2.83 per 100 person-year.

General
* Measure identifier number: 19.1
* Title: High risk practices index and condom-use safety index for homosexuals
* Reference
- **Kelly JA, St. Lawrence JS, Brasfield TL, et al. Psychological factors that predict AIDS high-risk versus AIDS precautionary behavior. J Consult Clin Psychol 1990;58:117-20.**
* Year: 1990
* Presentation
    * high risk practice index:
    - computed by multiplying number of unprotected anal intercourse occasions in the past 3 months by number of different partners.
    * condom-use safety index:
    - computed as above for protected anal intercourses

Factors: two
    a. type of sexual practice: anal intercourse
    b. number of partners
* Method of administering: self-administered questionnaire

Development process
* Underlying method: ad hoc approach
* Development process
    - the two risk factors included are based on the literature
    - the authors assume that, under most circumstances, the level of risk for AIDS due to a risky practice is a multiplicative function of both number of contacts and number of partners

Measure characteristics
* Type of measure: score
* Purpose: discriminative
    - to study the relationship between the AIDS risk sexual practice level and psychological and behavioural variables
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous 3 months
* Application
    * Population
    - homosexual
    - males
    - adults
    - United States
    * Setting: gay bars

Measure quality
* Reliability: not stated
* Validity: not stated

Alternative form
* Measure identifier number: 19.2
* Title: Risk index score
* Reference
- **Kelly JA, St Lawrence JS, Betts R, Brasfield TL, Hood HV. A skills-training group intervention model to assist persons in reducing risk behaviors for HIV infection. AIDS Educ Prev 1990;2:24-35.**
* Year: 1990
* Presentation
    - one index is computed for each sexual practice
    - computed by multiplying number of risky sexual practices by the number of different partners with whom these practices occurred
    a. risky sexual practices are: unprotected anal intercourse, oral-genital contact, oral intercourse with fluid exchange
* Factors: two
    a. type of sexual practice (one at the time)
    b. number of partners
* Method of administering: self-administered questionnaire

Development process
* Underlying method: ad hoc approach (see previous measure)
* Development process: not stated
    - the authors state that factors were chosen because they are predictive of the development of HIV infection

Measure characteristics
* Type of measure: score
* Purpose: evaluative
    - to evaluate the changes in AIDS risky sexual practice level
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous 4 months
* Application
    * Population
    - homosexual
    - males
    - adults
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 20
* Title: Multifactorial index of sexual behaviour for HIV risk
* Reference
- **Martin JL. AIDS risk reduction recommendations and sexual behaviour patterns among gay men: a multifactorial categorical approach to assessing change. Health Educ Q 1986;13:347-58.**
* Year: 1986
* Presentation
  - the people are classified in seven classes
  1. celibate
  2. monogamous, at home, low risk acts
  3. monogamous, at home, high risk acts
  4. polygamous, at home, low risk acts
  5. polygamous, out of home, low risk acts
  6. polygamous, at home, high risk acts
  7. polygamous, out of home, high risk acts.
  - risk acts were divided into high risk (receptive anal intercourse and insertive oral anal contact) and low risk (all others)
* Factors: three
  a. type of sexual practice
  b. number of partners (monogamous or polygamous)
  c. location of sexual encounters
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process
  - three factors were selected based on areas of sexual activity most often focused upon in educational risk reduction literature.
  - each factor was scaled in an ordinal scale, based on knowledge available, as follows:
  a. number of different sex partners (0, 1, 2 or more).
  b. location in which sexual contact took place (home-low risk, extra-high risk). Based on the assumption that extra-domestic sex involves strangers.
  c. sexual acts engaged (all-low risk, receptive anal intercourse and insertive oral-anal contact high risk). Based on the increased risk for HIV in literature.
  - the variables were combined in all logically possible ways (5 categories were excluded because impossible), resulting in a seven-category typology of sexual behaviour patterns.

Measure characteristics
* Type of measure: classification
* Purpose: evaluative
  - to evaluate changes in individual patterns and not in single sexual behaviour
* Operational definition
  - risk of exposure to HIV
  - due to sexual behaviour
  - in the previous 12 months
* Application
  * Population
  - homosexual
  - males
  - adults (20 - 65 years)
  - United States
  * Setting: community

Measure quality
* Reliability
  - test-retest was studied to evaluate the degree to which sexual behaviours reported for each time period could be considered reliable. It demonstrated good reliability.
* Validity: not stated

General
* Measure identifier number: 21
* Title: HIV Virus Infection-exposure Risk Index (VIRI)
* Reference.
- McQueen DV, Campostrini S, Robertson BJ, Uitenbroek D. A study of lifestyle and health. Interim report # 3. Research Unit in Health and Behavioral Change, University of Edinburgh, April 1991.
- Robertson BJ, McQueen DV. Continuous collection of data on AIDS-related behaviour in the UK. Proceedings of the 1991 Public Health Conference on Records and Statistics. Washington, DC: National Center for Health Statistics, 1991:247-51.
- Campostrini S, McQueen DV. Estimating exposure to HIV infection from self-reported sexual behavior in a population survey: constructing an index of behavioral risk. Am J Public Health 1993 (in press).
- Campostrini S. Un indice di rischio di infezione da HIV per la popolazione. Epidemiol Prev 1993 (in press).
* Year: 1991
* Presentation
- six variables included are:
a. homosexual activities in past 5 years
b. current homosexual activities
c. steady partner
d. length of present relationship
e. condom use
f. number of sexual partners over the past five years
- these variables are combined in a "tree format"
- there are 31 possible branches, each of which is classified into 5 categories of risk: high risk, medium risk, low risk, very low risk, no risk.
* Factors: five
a. number of partners
b. condom use
c. length of the present relationship
d. sexual orientation (homosexual or heterosexual)
e. type of partner: steady
* Method of administering: interview

Development process
* Underlying method: judgemental method (derived approach, Delphi method)
* Development process
- a panel of eight researchers connected with the study was selected
- eight different sex related variables (contained in the data-base of the RUHBC-LAH study) were given to the panellists to rank in a tree format
- first Delphi round: ranking the first tree of sexual behaviours. The combination of the 8 variables gave 49 different scenarios, which were ordered. Low agreement by participants.
- second Delphi round: ranking the second tree. Two variables were dropped. Still low agreement by the participants
- third Delphi round: re-ranking the second tree considering the feedback from the second round. High level of agreement
- the final tree considers 6 variables and the different scenarios were ordered in 5 risk classes

Measure characteristics
* Type of measure: classification
* Purpose: discriminative + evaluative
- to describe the general population in terms of risk
- to evaluate changes in risk in the population over time
* Operational definition
- risk of exposure to HIV
- due to sexual behaviour
- in the previous 5 years and currently
* Application
* Population
- general population
- males and females
- adults (18-60 years)
- United Kingdom
* Setting: community

Measure quality
* Reliability: not stated
* Validity
* Construct validity
- convergent validity: a second panel of national experts repeated the construction of the index through another Delhi exercise; the two panels achieved similar results

General
* Measure identifier number: 22
* Title: Sexual Risk Behaviour Assessment
* Reference.
- Meyer-Bahlburg HFL, Exner TM, Dellenbaugh C, El-Sadr W, Gorman JM, Ehrhardt AA. Sexual identity and sexual risk behavior in intravenous drug using men. Proceeding of the VII International Conference on AIDS. Florence, Italy, 1991;1:407 (abstract MD 4035).
- Meyer-Bahlburg HFL, Ehrhardt AA, Exner TM, Gruen RS. Sexual risk behavior assessment schedule-adult (material made available by the author).
* Year: 1991
* Presentation
- people are interviewed with the Sexual Risk Behaviour Assessment Schedule for Adult Drug-using Men (SERBAS-A-DM) and then classified into seven classes:
1. celibate
2. lower risk acts, monogamous
3. lower risk acts, polygamous
4. moderate risk acts, monogamous
5. moderate risk acts, polygamous
6. higher risk acts, monogamous
7. higher risk acts, polygamous
- sexual acts are divided into:
a. lower risk = physical contact without exchange of body fluid and without contacts between mucous membranes; this includes sexual contacts with consistent condom use.
b. moderate risk = unprotected insertive-anal intercourse, (unprotected insertive-vaginal intercourse); unprotected active oral contact; unprotected receptive oral contact; unprotected receptive oral-anal contact
c. higher risk = unprotected insertive oral-anal contact (unprotected receptive anal intercourse
* Factors: three
a. number of partners
b. type of sexual practice
c. condom use
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process: not described

Measure characteristics
* Type of measure: classification
* Purpose: discriminative
- to describe sexual risk behaviour
* Operational definition
- risk of exposure to HIV
- due to sexual behaviour
- in the previous 6 months
* Application
* Population
- general population
- males
- adults
- IVDUs
- United States
* Setting: hospital

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 23
* Title: Screening for HIV-related risk
* Reference.
- **Montgomery K, Lewis CE, Kirchgraber P. Telephone screening for HIV risk of infection. Med Care 1991;29:399-407.**
* Year: 1991
* Presentation
    - individuals are classified as follows:
    1. no or low risk (none of the criteria)
    2. increased risk (one or more criteria)
    - the criteria are:
    a. multiple sexual partners
    b. contact with sex partner who is at risk (has other partner, engages in prostitution, is IVDUs)
    c. homosexual contact that included anal intercourse
* Factors: three
    a. number of partners: multiple
    b. type of partner: high risk
    c. type of sexual practice: anal intercourse (for homosexuals only)
* Method of administering: telephone interview

Development process
* Underlying method: ad hoc approach
* Development process
    - the criteria are identified in the literature as the sexual behaviours consistently associated with seropositivity

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - to interview individuals to identify those whose behaviours place them at risk for HIV and to encourage them to be tested for HIV antibody
* Operational definition
    - risk of HIV infection
    - due to sexual behaviour
    - in the previous 2 years
* Application
    * Population
    - general population
    - males
    - adults (18 - 60 years)
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 24
* Title: HIV risk level for black homeless women
* Reference.
- **Nyamathi A. Comparative study of factors relating to HIV risk level of Black homeless women. J Acquir Immune Defic Syndr 1992;5:222-8.**
* Year: 1992
* Presentation:
    - individuals are classified into the following categories:
    1. high HIV risk: history of IVDU
    2. moderate risk: engaging in unprotected sex with two or more partners, or having sex with an IVDU, or using non-intravenous drugs
    3. low risk: engaging in unprotected sex with one partner, having a history of STDs or being homeless
* Factors: five
    a. number of partners
    b. type of partner: IVDU
    c. history of STDs
    d. intravenous drugs
    e. non intravenous drugs
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process
    - based on data showing that 50% of all adult female cases of AIDS are a result of parenteral drug abuse, whereas another 30% of cases in women are a result of heterosexual exposure

Measure characteristics
* Type of measure: classification
* Purpose: discriminative
    - to study predictors of HIV risk (environmental, demographic, personal, resources, knowledge, attitudes, health)
* Operational definition
    - risk of HIV infection
    - due to sexual and injecting behaviours
    - in the previous 2 years
* Application
    * Population
    - heterosexual
    - black females
    - adults (18-69 years)
    - homeless
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 25
* Title: Four-level sexual risk behaviour index
* Reference.
- Ostrow D, Beltran E, Wesch J, Joseph J. Recreational drug use and homosexual behaviour: the role of volatile nitrites ("poppers") in explaining the association. Proceeding of the VI International Conference on AIDS. San Francisco, CA, 1990;2:262 (abstract F.C 726).
- Ostrow DG, VanRaden MJ, Fox R, et al. Recreational drug use and sexual behaviour change in a cohort of homosexual men. AIDS 1990;4:759-65.
* Year: 1990
* Presentation
- individuals are classified into four classes:
1. no risk: no anal insertive or receptive intercourse, no oral receptive, no faecal-oral (rimming), or manual-anal (fisting) activity
2. low risk: only receptive oral and/or insertive anal intercourse (with or without condom)
3. modified high risk: receptive anal intercourse limited to one partner or the consistent use of condoms with all partners and/or faecal-oral or manual-anal intercourse limited to one partner
4. high risk: receptive anal intercourse, faecal-oral, or manual-intercourse with multiple partners and without the consistent use of condoms
* Factors: three
a. type of sexual activities
b. condom use
c. number of partner (monogamous, polygamous)
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process:
- four levels of sexual risk behaviour were created which reflected "a priori" conception about AIDS risk generally shared by investigators and the gay community at that time.
- pretesting of the questionnaire for reliability (see below)

Measure characteristics
* Type of measure: classification
* Purpose: discriminative + evaluative
- to study the relationship between drug use and persistence of high-risk behaviours
- to evaluate change in sexual behaviour
* Operational definition
- risk of exposure to HIV
- due to sexual behaviour
- in the previous 6 months
* Application
* Population
- homosexuals
- males
- adults
- United States
* Setting: not stated

Measure quality
* Reliability
- reliability of the questionnaire was evaluated in pretesting and questions considered unreliable or ambiguous by respondents were eliminated or revised
* Validity
* construct validity:
- convergent: the empirical utility of the categories in measuring the risk of HIV exposure was assessed by computing HIV seroconversion rates by category during the 6 months of follow-up period. The rates of seroconversion are:
a. none/low levels = 0.7 %
b. modified high levels = 2.7%
c. high levels = 7%
- potential modifications of the categorization were also examined in relation to seroconversion (e.g. incorporates insertive oral intercourse)

General
* Measure identifier number: 26
* Title: Risk assessment for HIV/AIDS
* Reference.
- **Rapkin AJ, Erickson PI. Difference in knowledge of and risk factors for AIDS between Hispanic and non-Hispanic women attending an urban family planning clinic. AIDS 1990;4:889-99.**
* Year: 1990
* Presentation
    - the risk factors are counted and added separately for traditional risk factor and life-style, yielding a subtotal or a grand total.
    a. traditional risk factors
    - self risk factors: history of transfusion; of IVDU; of prostitution (0, 1, 2)
    - partner risk factors: partner with history of transfusion; with haemophilia; with history of IVDU; bisexual partners; partner HIV-positive or with AIDS (0, 1, 2-5).
    - sub-total for traditional risk factors: 0, 1, 2, 3-6.
    b. life-style: sexual behaviour and substance use
    - life-style/self: more than 1 current partner; five or more life-time partners; history of STDs; frequent alcohol use; frequent marijuana use; frequent cocaine use (0, 1, 2, 3, 4, 5, 6)
    - life-style/partner: partner has sexual contacts with prostitutes; uses alcohol; uses marijuana (0, 1, 2, 3)
    - sub-total for life-style risk factors: 0, 1, 2, 3-6.
    - total possible risk factors: 0-17
    - women with any identified traditional risk factor are referred for HIV testing, and women with life-style factors are encouraged to consider being tested
* Factors: four + three
    - Factors for sexual behaviours
    a. type of partner (high risk group)
    b. number of partners
    c. history of STDs
    d. history of prostitution
    - Other factor are:
    e. history of transfusion
    f. history of IVDU
    g. use of alcohol or non intravenous drugs
* Method of administering: interview

Development process
* Underlying method: ad hoc approach
* Development process: not described

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - case finding and referring for testing
    - to study the prevalence of risk factors for HIV
* Operational definition
    - risk of HIV infection
    - due to sexual behaviours and others risk factors
    - in the life-time
* Application
    * Population
    - heterosexuals
    - females
    - adults
    - United States
    * Setting: family planning clinic

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 27
* Title: Probability of becoming infected with HIV
* Reference.
- **Sandberg S, Awerbuch TE. Mathematical formulation and studies of the risk parameters involved in HIV transmission. Bull Math Biol 1989;51:467-74.**
* Year: 1989
* Presentation
  - probability of becoming infected after N contacts = $1 - \{1 - P [1 - (1 - b)^n]\}^m$
  - where:
    (n) number of sexual contacts
    (m) number of partners
    (b) probability of infection per sexual contact
    (P) probability that a partner is infected

    * In the case of $N = m \times n$ = constant
    - results on reducing each risk factor. For b small, N, P and b have an equal influence on the transmission probability, so that halving P or b or N leads to an equivalent reduction in risk.
    - results on transmission probabilities as the epidemic progress. If the prevalence is constant, having multiple partners versus monogamous relationship is unimportant; but if the prevalence is increasing and multiple partners are chosen over time, a monogamous relationship where that partner is chosen early in the epidemic is less risky.
* Factors: four
    a. number of contacts
    b. number of partners
    c. sexual practice: one at a time (infectivity)
    d. prevalence of HIV infection among partners
* Method of administering: not stated

Development process
* Underlying method: mathematical method
* Development process
    - Assumptions
    a. binomial model
    b. partners (m) are independent trials, each with probability P of being infected
    c. contacts (n) with a given partner are independent trials
    d. probability (P) that a random selected partner is infected is constant over time
    e. infectivity (b) is constant for different partner

Measure characteristics
* Type of measure: probability equation
* Purpose: predictive
    - to assess the individual risk
    - to study, with computer simulations, the effect of reducing each of the risk factors considered (number of sexual contacts, number of partners and safer sexual habits) on lowering the probability of infection and to give indication on what behavioural changes should be made
    - to study the transmission probability as the epidemic progress
* Operational definition
    - risk of HIV infection
    - due to sexual behaviour
    - after n contacts
* Application
    * Population: general population
    * Setting: not stated

Measure quality
* Reliability: not stated
* Validity: not stated
    - some discussion about the assumption and results

General
* Measure identifier number: 28
* Title: Behavioural risk assessment form
* Reference.
- Shaw NS. Wessell LG. Risk assessment form. In: Keeling RP, ed. AIDS on the college campus. ACHA special report. Rockville, MD: American College Health Association, 1986:55-7.
- Leapley P, Kasselman MJ. AIDS risk behaviour of entering college students. Proceeding of the VII International Conference on AIDS. Florence, Italy 1991;1:398 (abstract M.D.4034).
* Year: 1986
* Presentation
    - eight items are included:
    a. partners per month last year,
    b. partners per months the year before,
    c. types of relationship (stable or casual)
    d. place of sexual contact,
    e. drug use during sexual encounters,
    f. IV drug use,
    g. geographical area of sexual encounters,
    h. sexual practices
    - each factors is scaled (scored 1,2 or 3)
    - the scores are summed up to obtained the final score (range 0-24)
    - two cut-off points are chosen, dividing the final score into three categories: high (17-20), medium (12-16), low risk (0-11)
* Factors: six
    a. number of partners
    b. type of relationship
    c. place of encounters
    d. drugs use
    e. geographical area
    f. sexual practice.
* Method of administering: self-administered questionnaire or interview

Development process
* Underlying method: ad hoc approach
* Development process: based on guidelines

Measure characteristics
* Type of measure: score
* Purpose: discriminative
    - to assist clients and health care providers in understanding the behavioural factors that influence an individual's risk of AIDS
    - to educate about behaviour that may lead to exposure to the virus and to teach risk reduction
    - to reach the decision to undergo HIV antibody test
* Operational definition
    - risk of exposure to HIV
    - due to sexual and injecting behaviours
    - time frame not stated
* Application
    * Population
    - general population
    - males and females
    - adolescents
    - United States
    * Setting: college

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 29.1
* Title: Summary risk score for HIV related sexual activity in homosexuals (Stall)
* Reference
- Stall R, McKusick L, Wiley J, Coates TJ, Ostrow DG. Alcohol and drug use during sexual activity and compliance with safe sex guidelines for AIDS: the AIDS behavioral research project. Health Educ Q 1986;13:359-71.
* Year: 1986
* Presentation
    - the frequency of each sexual behaviour during the previous month is multiplied by its weight; these products are then added.
    - each sexual behaviour has been given the following weight:
    0. completely safe = massage, cuddling, frottage,
    1. probably safe = wet kissing, anal intercourse with condoms, oral sex, water sports, digital penetration,
    2. probably risky = rimming, swallowing semen during oral sex, penetration with dildo,
    3. risky = anal intercourse without a condom, fisting.
    - the total score is classified into three categories for risk: no risk (0), medium risk (1-9), high risk (10 or more)
* Factors: two
    a. type of sexual practice
    b. frequency of contacts
* Method of administering: self-administered questionnaire

Development process
* Underlying method: ad hoc approach
* Development process
    - the choice of sexual practice and their relative weight are based on information disseminated by health education campaigns in San Francisco during the study period

Measure characteristics
* Type of measure: score
* Purpose: discriminative + evaluative
    - to evaluate change in sexual behaviour over time
    - to determine whether an association exists between alcohol and/or drug use and level of risk in sexual activity
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous month
* Application
    * Population
    - homosexual
    - males
    - adults
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

Alternative form:
* Measure identifier number: 29.2
* Title: Summary risk score
* Reference
- **Leigh BC. The relationship of substance use during sex to high-risk sexual behaviour. J Sex Res 1990;27:199-213.**
* Year: 1990
* Presentation
    - each sexual behaviour is multiplied by its weight and by the number of times the respondent had engaged in these behaviours in the past month.
    - weights for sexual practices: safe (0 = mutual masturbation), probably safe (1 = vaginal or anal intercourse with condom, vaginal oral sex, oral sex without swallowing), probably risky (2 = oral sex with swallowing semen), risky (3 = vaginal or anal sex without a condom).
    - the scores are summed up to yield the final score
    - modified categories for risk: low risk, moderate risk, high risk
* Factors: two
    a. type of sexual practice
    b. number of contacts
* Method of administering: self-administered questionnaire

Development process
* Underlying method: see previous measure with adaptation for heterosexuals
* Development process
    - the choice of sexual practice and their relative weight are based on public health guidelines

Measure characteristics
* Type of measure: score
* Purpose: discriminative
    - to study the relationship of substance use and unsafe sex
* Operational definition
    - risk of exposure to HIV
    - due to sexual behaviour
    - in the previous month
* Application
    * Population
    - general population
    - males and females
    - adults
    - United States
    * Setting: community

Measure quality
* Reliability: not stated
* Validity: not stated

General
* Measure identifier number: 30
* Title: Composite high-risk index
* Reference
- **Windle M. High-risk behaviors for AIDS among heterosexual alcoholics: a pilot study. J Stud Alcohol 1989;50:503-7.**
* Year: 1989
* Presentation
    - the seven items on sexual activities (plus 2 on IVDU) are:
    a. three or more sexual partners;
    b. unprotected vaginal intercourse with primary partner;
    c. unprotected anal intercourse with primary partner;
    d. unprotected vaginal intercourse with non-primary partner;
    e. unprotected anal intercourse with non-primary partner;
    f. oral sex with swallowing with primary partner (only women);
    g. oral sex with swallowing with non-primary partner (only women);
    h. IVDU;
    i. IVDU in the previous 7 years.
    - the risk factors are summed (0-7 for men, 0-9 for women)
    - high risk defined as the self-reported engagement in each of the risk factors identified at least once in the time frame
* Factors: three + one
    a. number of partners
    b. type of sexual practice
    c. type of partner (primary / non-primary)
    d. IVDU
* Method of administering: interview

Development process
* Underlying method: psychometric method
* Development process
    - risk factors were identified from existing surveys on homosexuals
    - the internal consistency of the measure has been studied (see reliability)

Measure characteristics
* Type of measure: checklist
* Purpose: discriminative
    - to measure incidence and prevalence of level of high-risk behaviours for AIDS
* Operational definition
    - risk of exposure to HIV
    - due to sexual and injecting behaviours
    - in previous 6 months
* Application
    * Population
    - heterosexual
    - males and females
    - adults
    - alcoholic inpatient
    - United States
    * Setting: hospital

Measure quality
* Reliability
    - Internal consistency: KR-20 for seven risk factors common to both men and women was 0.70
* Validity: not stated