Developing and Measuring the Reliability and Validity of the Factors Influencing the Implementation of ICD-10-AM and Clinical Coding in Saudi Public Hospitals

Musaed Ali Alharbi[1,3], Barry Tolchard[1,2] & Godfrey Isouard[1,4]

[1] University of New of New England, Armidale, Australia

[2] University of Huddersfield, Queensgate, Huddersfield, UK

[3] Ministry of Health in Kingdom of Saudi Arabia

[4] Australasian College of Health Service Management (ACHSM), Sydney, Australia

Correspondence: Musaed Ali Alharbi, School of Health, University of New England, Elm Avenue, Armidale NSW 2351, Australia. Tel: 61-26-773-3569. E-mail: malharb2@myune.edu.au; musaed52@gmail.com

Abstract

Background: The introduction of a mandatory health insurance system contributing towards the funding of national healthcare in Saudi Arabia necessitates the implementation of clinical coding and a unified health classification system, which has previously not been a feature of Saudi healthcare. As the Ministry of Health (MOH) moves to introduce ICD-10-AM, the Australian modification of the WHO ICD-10, in the Kingdom's public hospitals, it is important to understand the factors that will influence its successful implementation.

Objective: The purpose of this article is to develop and evaluate the internal consistency reliability and validity of a questionnaire establishing the factors influencing the the implementation ICD-10-AM and clinical coding in Saudi public hospitals.

Method: The content validity method was initiated by sending the whole draft questionnaire to a panel of experts to indicate values for each item based on a scale of content validity created by the researchers and, subsequently, using the internal consistency reliability and factorial validity methods to estimate the internal reliability of clusters of items, which were assumed to measure the same factors, grouped in this study into three factorial categories, health information (clinical documentation, classification, and coding requirements), organization (the implementation preparation in individual organizations), and national (institutional support through the national hierarchical structure).

Results: The content validity identified all items of the proposed questionnaire to be valid. Based on the content validity test, several items were removed as they did not meet the proposed model and the final questionnaire was created in accord with the pilot study result. The pilot study utilized Cronbach's α and factor analysis to examine the reliability and validity of Part 2 of the questionnaire and the findings indicated high internal consistency reliability and factorial validity.

Keywords: content validity, Cronbach's α, ICD-10, pilot study, reliability and validity, Saudi Arabia

1. Introduction

The new millennium has seen an acceleration in the development of information technology, enhancing the power and speed of data processing. Its application to healthcare has led to the emergence of a distinct new field, Health Information Technology (HIT) and it related human resources category, Health Information Management (HIM). HIT improved methods of disseminating and sharing health information data between organizations and professionals, ultimately contributing to greater efficiency in healthcare. HIT has also improved data security, privacy, and integrity and is considered the dominant factor in the global improvement in healthcare systems (Al Kiyumi, Walker, Tariq, & FitzGerald, 2017; Narayanan & Rose, 2017). Hersh (2009, p. 3) offered the following definition of this new field of healthcare: "HIM is the discipline that has historically focused on the management of medical records. As the medical record has become electronic, this field has been in transition and increasingly overlaps with informatics."

HIM has a central role in the modern healthcare system and refers to the control and management of healthcare data, which includes the coding of diagnoses and procedures, the storage of medical records, and other individual patient data, as well as the billing process (Adeleke et al., 2015). Fiorito and Edens (2016, p. 2) proposed the following definition: "HIM is the practice of acquiring, analyzing, and protecting digital and traditional medical information vital to providing quality patient care."

Recent trends in the HIM revolution, especially in 'clinical coding' which has become an integral HIM practice, have led to a proliferation of studies highlighting its role in promoting efficiency in healthcare delivery systems. Wager, Lee, and Glaser (2009) emphasized that clinical coding is a primary function of HIM, which plays a vital role in bringing about the best health information and, ultimately, contributes to improving healthcare delivery services. Consequently, implementing clinical coding in health organizations is an essential tool to improve health information management data systems by reducing medical errors and potential costs, managing health insurance, improving tracking of health services, incorporating research, and other purposes (Sanders et al., 2012).

In everyday practice, the term 'clinical coding' refers to the process of reading the patient file to identify the diagnosis and procedures carried out during a period of admission to a hospital or visit to an outpatients clinic; thereafter, assigning the respective codes to the identified diseases and interventions, according to the classification of diseases and procedures used in the institution (Heywood et al., 2016). The stages of the process are described by Global Medical Administration (2019): "Clinical coding is a health administration function that involves the translation of written clinical statements into a code format. A clinical coder will analyse information about an episode of patient care and assign standardized codes using a classification system."

As the World Health Organisation (WHO) publishes only the diagnosis classification and codes, several countries, namely Australia, the United States, Canada, Germany, Thailand, and Korea, have produced separate classifications of the procedures or interventions used in hospitals, as well as modifying the WHO diagnosis classification to suit their national healthcare systems (De Coster, Li, & Quan, 2008; Jetté et al., 2010; Latimer, 2010; Walker et al., 2012).

The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM), the Australian Classification of Health Interventions (ACHI) and the Australian Coding Standards (ACS) represent the Australian modification of the WHO ICD-10. ICD-10-AM is used to assign the appropriate alphanumerical codes to inpatient medical records in Australia and a growing body of licensed countries (Shepheard, 2009), including Bahrain, Bosnia & Herzegovina, Bulgaria, Ireland, Lithuania, Macedonia, Moldova, New Zealand, Montenegro, Qatar, Saudi Arabia, Serbia, Singapore, Slovenia, Turkey, and Ukraine (Independent Hospital Pricing Authority, 2019). In practice, itemized billing is no longer used as a basis for costing and reimbursement and the codes are grouped according to one of several grouping systems. Australia has developed the Australian Refined Diagnosis-related Groups (AR-DRGs). For example, a diagnosis of appendicitis, followed by a standard appendectomy, involves an ultrasound confirmation of the diagnosis executed by a sonographer and assessed by a radiologist, laparoscopic surgery performed by a general surgeon, a general anesthetic administered by an anesthetist, five to seven days recovery in a ward, and a check-up two weeks later. All of these stages have been grouped based on averaging over a considerable period by a grouper algorithm. After the whole hospital sequence has been translated into a coding sequence by the coder, the standard cost for the group will be applied.

Previously, Saudi Arabia has not had a unified national healthcare classification system. However, the establishment of mandatory health insurance to alleviate the financial burden of providing free healthcare, demands a unified system and the Saudi government has established the mechanisms to implement the Australian national modification, ICD-10-AM/ACHI/ACS, for use in all Saudi healthcare provider hospitals, overseen by the Council of Health Services (Council of Health Services, 2016).

2. Aim of the Study

This paper forms a part of a mixed method study entitled "Factors Influencing the Implementation of Clinical Coding and ICD-10-AM in Saudi Public Hospitals". It is based on the chapter that covers the design and construction of the survey questionnaire, which forms the study quantitative instrument, as well as the process of testing its internal consistency reliability and validity.

3. Methods and Results Used to Design and Develop the Questionnaire

The design and development of the questionnaire used in this study followed the steps outlined in Figure 1. The subsequent sections of this paper provide a detailed description and rationale for Steps 1 to 4, based on the literature covering the validation of survey questionnaires.
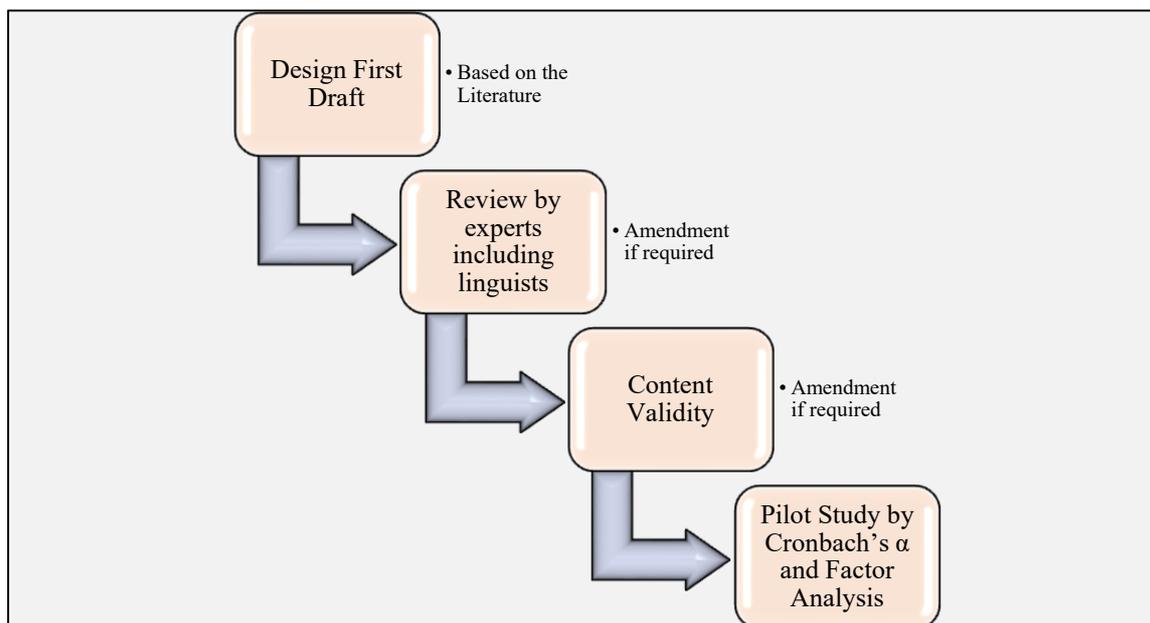
Figure 1. Flowchart of Questionnaire Design and Development

3.1 First Draft

The design of the questionnaire was preceded by a general review of the healthcare literature on clinical coding and ICD-10 implementation, grouping Australia, Canada, and the United States as developed nations and Saudi Arabia and Thailand as developing nations. The review evaluated peer-reviewed articles, reports, articles, and databases and was broadened by examining the bibliographies of certain articles to probe more deeply into the topic. A distinction was noted in the influencing factors applicable to developed and developing nations. The Saudi Arabian literature anticipating the introduction of clinical coding points to many inherent problems in Saudi healthcare that included inadequate and incompatible technology and a dearth of national data standards, a shortage of trained administrative staff, collective resistance to innovation from the body of physicians, and an antiquated attitude in the MOH which ignores modern standards of consultative management. In contrast, the ICD-10 literature from Thailand produced a far more positive attitude, emphasizing that health reform in developing countries is a longer-term process compared to the developed nations but must still move forward, despite its slower pace.

A comparative overview of the ICD-10 literature indicated that Australia emphasizes training the clinical coder workforce and organizational preparedness as the keys to a successful ICD-10 implementation. National data standards are barely referred to, as the Australian healthcare system has long been unified. Canada, however, due to the historical division into French and English-speaking sectors, previously had two classification systems and different healthcare data standards, which have finally been integrated. This has facilitated the gathering of national healthcare statistical data. Alternatively, the US literature reflected a split into two camps; one for and the other against an ICD-10 implementation. Whereas in Australia, health information is clearly left in the hands of health information professionals, in the US the health information professionals were concerned about the importance of training to deal with the expansion of classification in ICD-10 and the greater benefit this implied for global healthcare, while the physician outlook was based on the burden of costs and resources this expansion would entail.

Alternatively, the Saudi researchers point to a country ill-prepared in both human and technological resources to cope with clinical coding and ICD-10, while researchers in Thailand which had already experienced the benefits of the new classification, argued that no matter how imperfect the healthcare environment appeared, it was essential to make a start.

The first draft of the questionnaire was then designed with the content based on factors considered to influence the implementation of clinical coding and ICD-10 classification drawn from the literature reviewed.

After reviewing the literature, which revealed a marked distinction between developed and developing nations, as well as several factors specific to Saudi Arabia, the major factors influencing implementation were integrated and

classified into three categories, (1) health information: factors relating to HIM, clinical documentation and coding, ICD-10 classification and coding standards (2) organization: factors relating to implementation responsibilities in the individual institutions, including planning, training, and technology and manpower impact and (3) national: factors relating to national data standards, interoperability, and the responsibilities of an overall supervisory body. Hence, an acceptable level of health information is dependent on information-based activities and processes within the organization, which in turn, are founded on the policy and support at national level.

3.2 Review by Experts Including Linguists

Liamputtong (2013) proposed that after thorough revision, the first draft of a questionnaire should be reviewed and evaluated by experts in the specific research topic and in survey development. The researchers may then modify and/or reorder the questions according to the critical feedback provided by acknowledged experts. The researchers were fortunate to engage three health information experts to review the first draft of the English version of the questionnaire. Two evaluators suggested removing all items referring to the AR-DRG system, based on the rationale that AR-DRGs had not yet been introduced in Saudi public hospitals and were likely to be misunderstood by many respondents. Subsequent to the health information experts' feedback, the researchers followed the recommendations of Sousa and Rojjanasrirat (2011) for questionnaires requiring translation from the original language in which it was drafted into the language in which it will be presented to the study respondents. The following procedure was adhered to:

1) The English (original) draft was reviewed by language experts in Saudi Arabia;

2) The questionnaire was translated into Arabic by sending it to a certified translation office in Saudi Arabia;

3) The Arabic version was translated back into English by sending it to a certified translation office in Australia;

4) The two English versions, the original and the version translated back from Arabic, were compared.

3.3 Amendment of Questionnaire

The researchers amended the questionnaire according to the feedback obtained from the expert assessment from the previous step. The initial draft contained 26 statements in the second section, which was amended to 25, as the evaluators considered one item was invalid. The questionnaire was structured in three sections:

1) Respondent demographics (10 items);

2) Factors influencing the implementation of ICD-10 and clinical coding, comprising 25 items, divided into three categories; namely Health Information, Organizational, and National, rated according to a five-point Likert scale of response (Strongly Agree = 1, Agree = 2, Neutral = 3, Disagree = 4 and Strongly Disagree = 5);

3) Open-Ended Question (1 item).

3.4 Content Validity

In their analysis of content validity, Rubio, Berg-Weger, Tebb, Lee, and Rauch (2003, p.94) drew a distinction between face validity and logical validity: "Face validity indicates that a measure appears to be valid based on a subjective evaluation. Logical validity indicates a more rigorous process, such as using a panel of experts to evaluate the content validity of a measure." Content validity is evaluated to ensure that all items in the questionnaire contribute to the purpose of the study. Content validity is established when all items in the study instrument measure the constructs they have been designed to measure (Babbie, 2013). Heale and Twycross (2015) asserted that an essential step in the study instrument validation process is the estimation of a scale of content validity.

The draft questionnaire was sent to a panel of experts to indicate values for each item based on a scale of content validity created by the researchers (Polit & Beck, 2006). Scholars differ regarding the ideal number of experts for testing content validity. Yaghmaei (2003) recommended a minimum of five experts and a maximum of ten as being acceptable, while Lynn (1986) suggested a minimum of three and a maximum of ten.

To validate the content of this study, an additional three health information experts reviewed all items of the revised questionnaire and evaluated the content validity as being consistent with the objective of the study. For this assessment step, the Content Validity Index (CVI) created by the researchers was based on the following three-point scale: (1 = Amendment, 2 = Agree, 3 = Not agree). Where 1 or 3 was chosen, the experts supplied explanatory comments or suggestions on the line beneath the item.

3.4.1 Content Validity Findings

One item in the first draft of the questionnaire was found to be invalid in the CVI analysis, as it received a CVI of

2 = Agree from two experts. The item had been worded: "MOH hospitals have sufficient health informatics and health information management specialists to implement ICD-10." This item was redrafted to make a distinction between the two groups of healthcare professionals:

1) MOH hospitals have sufficient health information management specialists.

2) MOH hospitals have sufficient health informatics specialists.

One expert with a superior understanding of the difficulties facing the healthcare profession in Saudi Arabia, particularly in HIM, suggested adding a question focusing on the absence of a HIM supervisory organization. He pointed to the existence of a conflict of interests between the Council of Cooperative Health Insurance, which is responsible for monitoring the implementation of ICD-10 in the private sector, and the Saudi Health Council (SHC), which is responsible for monitoring the implementation across all Saudi Arabia's health sectors. Another item received a 3 = Not Agree from both experts and was, thus, rendered invalid. The item was worded: "Which of the following factors have caused a delay and/or lack of progress in ICD-10-AM planning and implementation in your hospital. Check all that apply." One of the experts explained that suggesting a delay or lack of progress was an assumption and therefore the item was deleted. Three further items from the first draft questionnaire were also deleted as they did not fall within the parameters of the purpose of the research and were rated 3 = Not Agree by two experts. These items were: "There is resistance to implementing ICD-10 among some decision-makers in Saudi hospitals.", "The Saudi Health Council offers clinical coding implementation support to all Saudi health sectors'", and "My hospital does not use the latest technologies in Health Information Management."

3.5 Internal Consistency Reliability and Factorial Validity

Internal consistency reliability refers to the degree of consistency of the measurement of concepts or factors extracted from clusters of items in a questionnaire or test instrument. It may be defined as 'How well the items measuring a concept hang together as a set" (Cavana, Delahaye, & Sekeran, 2001, p. 320). Cronbach's α was applied to estimate the internal reliability of clusters of items which were assumed to measure the same factor. The unstandardized values of Cronbach's α were estimated for the reliability analysis because the numerical item scales (ranging from 1 to 5) were the same for each item. Cronbach's α was used because it is considered the best technique to test the internal consistency reliability of variables in the evaluation of questionnaires (Cronbach & Shavelson, 2004; Hogan, Benjamin, & Brezinski, 2000).

Cronbach's α values increase in relation to the number of correlations among a cluster of items, such that a high rate of alpha is created by a homogeneous cluster of items that have a similar magnitude of correlations. If the item responses consist only of random errors and the correlations are missing between the items, then the variance is the same between the item scales and the individual items so that alpha equals zero. In such a case, the responses are considered unreliable and an inconsistent measure of the construct. Values of Cronbach's α increase when the correlations between the item scores increase, so that when alpha = 1, a group of items is a perfectly reliable and consistent measure of a construct. However, alpha = 1 is rarely, if ever, encountered in reliability analyses, due to the influence of random error. Statisticians generally recommend that the threshold or cut-off value of Cronbach's alpha should be a minimum of .6 for reliability to be considered 'adequate' and .8 to be considered 'good' (Allen & Yen, 2001). Fan and Lê (2011, p. 368) confirm this in stating "All Cronbach's alpha coefficient values above 0.6 are considered acceptable."

The practical applications of Cronbach's alpha are limited and the use of this statistic is subject to the following assumptions:

1) The calculation of alpha assumes parallelity (i.e. the factor loadings for each item in a scale are constrained to be equal, all the error variances are constrained to be equal, and the error variances are assumed to be uncorrelated). Failure to meet any of these assumptions frequently results in coefficient alpha's underestimation of the reliability of a scale (Peterson & Kim, 2013; Raykov, 2001).

2) The value of Cronbach's alpha is valid only when the item scales measuring several aspects of a single construct move in the same logical direction. The direction of the scales is a critical issue, since Cronbach's alpha, unlike factor analysis, cannot be used to assess the reliability of clusters of items based on scales that move in both forward and reverse directions. Negative values of alpha, which are invalid, and cannot easily be interpreted, are extracted from such opposing scales (Field, 2009). For this reason, factor analysis is more useful to provide evidence for the internal consistency reliability of heterogeneous item scores than Cronbach's alpha.

3) The values of alpha are generally found to increase in direct proportion to the total number of items in the factor (Bonett, 2002). According to the relative values of alpha, a factor containing many variables will

automatically be more reliably measured than a construct containing only a few items. The strong influence of sample size implies that the numerical comparison of individual levels of alpha between factors comprising different numbers of items is not appropriate.

4) If a factor is reported to be reliably measured using Cronbach's alpha in one study, this does not imply that the same factor will be reliably measured in another study using the same questionnaire. Estimates of reliability (and, by extension, validity) are study specific. If the study population (as well as the sample size, sample characteristics, context, and administration of the instrument) differ, then the estimates of scale reliability will axiomatically also be different (Thompson, 2003). For the above reasons, it is not surprising that in a review of reliability issues in psychological measurement, Sijtsma and Emons (2011, p. 565) concluded that "as a reliability estimator, alpha should be replaced by better and readily available methods."

In view of the limitations of Cronbach's alpha, factor analysis was also performed to explore the validity and internal consistency reliability in terms of the structure, pattern, and consistency of the items that constituted each factor. Factor analysis condenses the matrix of item scores classified into 50 cases (in the rows) and 20 questionnaire item scores (in the columns) into a smaller number of hypothetical dimensions or constructs, conventionally called factors (Gorsuch, 1983). The factors are subsets of inter-correlated item scores that explain specified proportions of the variance in the correlation matrix. There are two types of factor analysis commonly used by social scientists (Gorsuch, 1983). The first, Exploratory Factor Analysis (EFA), assumes no preconceptions of what the factors might be. The second, Confirmatory Factor Analysis (CFA), confirms the existence of preconceived factors. This study used CFA because the questionnaire contained three predetermined factors, each with a specified number of items, as outlined in the questionnaire.

Several statistical methods exist to conduct CFA, examples being Principal Components Analysis (PCA) and Maximum Likelihood (ML). After the initial extraction of factors, there are many types of rotations that can be done. Rotation is required because the unrotated solution is often indefinite and difficult to interpret (Gorsuch, 1983). In this study, the factors were extracted using PCA and Varimax rotation. Varimax is the most common rotation tool used to increase and decrease the loading on each variable (Gorsuch, 1983). The results of CFA are sensitive to the sample size. According to Hair, Anderson, Babin, Tatman, and Black (2010, p. 102): "A researcher would generally not factor analyse a sample with fewer than 50 observations." The number of respondents in this study was 50 and, therefore, the sample size was the minimum for factor analysis.

The outputs of CFA are eigenvalues, commonalities, and factor matrices which report the loadings of sequentially numbered factors onto specified clusters of variables. The interpretation of these statistics requires subjective judgments and so the conclusions of factor analysis are more tentative than when using other types of statistical tests (Gorsuch, 1983). The convention adopted in this study was that of extracting factors with eigenvalues greater than 1 and that the cumulative variance explained by the factors should be greater than 50% (Hair et al., 2010; Murtagh & Heck, 2012).

Commonalities are the proportions of variance of each variable that explain the extracted factors. High commonalities (between .5 and 1) indicate closely-related variables, while communalities of less than .25 have little in common with one another (Hair et al., 2010). The convention used in this study was that factor loadings should be high ($\geq$ .5) to contribute significantly to the reliability and validity of a factor, whilst variables with low loadings (< approx. .25) were considered to contribute little or nothing to a factor and could potentially be eliminated (Hair et al., 2010).

3.6 Pilot Study for the Second Part of Study Questionnaire

A pilot study was conducted after the content validity process, by using Internal Consistency Reliability and Factorial Validity for proposed factors, which is a methodological approach to evaluate the validity and reliability of a questionnaire in a small study before the questionnaire is administered in a further major study (De Vaus, 2013). Riazi (1999, p. 198) defines a pilot study as 'a small-scale replica and a rehearsal of the main study.' Polit-O'Hara, Beck, and Hungler (2001) defined a pilot study as the process of testing the validity and reliability of an instrument before collecting the data for the main study. De Vaus (2013, p. 54) warned: "Do not take the risk. Pilot test first." Van Teijlingen and Hundley (2002) noted that the pilot study is an essential step to test the validity and reliability of the research instruments. Baker and Risley (1994) advised that social science studies should always include pre-testing of the research instruments before collecting the main study data, to validate and test the reliability of the factors in the second section of the questionnaire. Christensen, Johnson, and Turner (2011) noted that a pilot study affords the researcher an opportunity to validate the research tools and test the feasibility of the study design. It is evident, therefore, that a pilot study was essential to ensure that the questionnaire was valid and reliable for the data collection of the main study.

It has been recommended that the sample size of a pilot study should be approximately 10% of the main study (Hertzog, 2008; Johanson & Brooks, 2009; Lackey & Wingate, 1998). As the researchers aimed to distribute the questionnaire to 500 participants in the targeted hospitals in the final study, the sample size of the pilot study was fifty. Thus, fifty participants comprising healthcare professionals from six specified groups of the study hospitals, representing 10% of the final study sample size were recruited for the pilot study using a random sample technique and answered all items. The participants were drawn from clinical coders and medical records staff, physicians, nurses, and other health professionals such as pharmacists and technicians, as well as from non-health professionals such as administrators.

3.6.1 Findings Related to Organizational Factors

The analysis of the eight items designed to measure the organizational factors impacting on implementation resulted in one factor accounting for 38.64% of the variance. High communalities (.634 to .845) indicated that the items were related. The eight items had moderate loadings (.381 to .748) and, therefore, all contributed significantly to the reliability and validity of the questionnaire. Cronbach's alpha (.763) reflected excellent internal consistency reliability. All factors related to Organization were established as valid and reliable.

3.6.2 Findings Related to Health Information Factors

The analysis of the six items designed to measure the health information factors impacting on implementation resulted in one factor accounting for 37.33% of the variance. Moderate communalities (.363 to .864) indicated that the six items were related. The six items had strong loadings (.434 to .785) and, therefore, contributed significantly to the reliability and validity of the questionnaire. Cronbach's alpha (.651) reflected adequate internal consistency reliability for the six items. All factors related to Health Information were established as valid and reliable.

3.6.3 Findings Related to National Factors

The analysis of the remaining nine items, designed to measure National factors impacting on implementation, resulted in one factor accounting for 26.35 % of the variance. High communalities (.451 to .891) indicated that the items were related. The nine items had moderate loadings (.242 to .682) and therefore barely adequate consistency reliability and validity of the questionnaire. Cronbach's alpha (.604) reflected moderate internal consistency reliability. The National Factors impacting on implementation were, therefore, on the borderline of validity and reliability.

Table 1. Factor Analysis and Cronbach Alpha Table Results

	Mean	Standard Deviation	Communality	Loading	Eigenvalue	% of Variance	Cronbach's alpha
ORGANIZATIONAL							
ORG 1	4.26	0.944	0.706	0.693			
ORG 2	4.28	0.858	0.721	0.674			
ORG 3	4.56	0.675	0.620	0.748			
ORG 4	4.54	0.579	0.845	0.637	3.91	38.64	0.763
ORG 5	4.36	0.563	0.643	0.381			
ORG 6	4.44	0.541	0.634	0.644			
ORG 7	4.26	0.853	0.517	0.626			
ORG 8	4.48	0.544	0.723	0.490			
HEALTH INFORMATION							
INFO 1	4.26	0.751	0.675	0.434			
INFO 2	4.22	0.840	0.419	0.423			
INFO 3	4.14	0.833	0.864	0.785	2.24	37.33	0.651
INFO 4	4.02	0.845	0.801	0.745			
INFO 5	4.02	0.979	0.363	0.553			
INFO 6	4.08	0.877	0.430	0.629			

NATIONAL

NAT 1	4.00	0.782	0.736	0.594			
NAT 2	4.06	1.077	0.501	0.288			
NAT 3	4.04	0.699	0.775	0.242			
NAT 4	4.48	0.544	0.891	0.304			
NAT 5	4.26	0.853	0.752	0.516	2.38	26.53	0.604
NAT 6	4.40	0.535	0.538	0.636			
NAT 7	4.12	0.746	0.649	0.682			
NAT 8	4.04	0.880	0.451	0.626			
NAT 9	4.14	0.857	0.760	0.522			

Note. NAT 6 and NAT 9 data includes reverse-scored questions.

4. Discussion

The transition from the ninth to the tenth revision of the International Classification of Diseases (ICD-9 to ICD-10) has been an issue in international healthcare since 1994, when the first national implementations of the new version occurred. In the developed nations with existing qualified clinical coder workforces, preparations included the training of HIM staff and physicians in the enormously expanded classification, which required more specific clinical documentation from physicians and greater depth in anatomical and physiological knowledge from clinical coders. While some Australian states began using ICD-10-AM in 1998, the United States encountered strong opposition from some powerful physician bodies and the implementation of the US national modification was delayed several times before officially taking place in 2015. In the case of the ongoing Saudi implementation, it is an implementation from scratch in Saudi hospitals, which have not previously utilized the clinical coding process.

The absence of studies on coding in Saudi public hospitals makes the questionnaire the first to explore these factors. The design of the questionnaire was based on the identification of the underlying global factors influencing the implementation of clinical coding and ICD-10, using a literature review.

The Saudi Arabian literature, however, points to several external or indirect factors that are likely to have an impact. Alsadan et al. (2015) argued that technologically, most developing countries including the Arab world, have not managed to keep pace with the developed nations. The failure to keep pace with HIT developments has severely restricted general healthcare improvements. The failure to achieve intended reforms, in a decade of global acceleration of the development of information technology is the result of a combination of poor MOH co-ordination and the opportunism of vendors to exploit this, in the provision of technological components and systems that lack interoperability (Alsadan et al., 2015).

In the most recent Saudi study dedicated to assessing quality improvements in general in Saudi Hospitals by Alaraki (2018), the author states:

Despite the use of different quality improvement models to improve healthcare in Saudi hospitals during the last two decades, consistent improvements have not yet been achieved and the results are still far below expectations. This may reflect the presence of hidden organizational factors in the local contexts that hinder quality improvement efforts. (p.8)

Other recurrent inherent weaknesses that have never been rectified that have the potential to impact on a successful ICD-10 and clinical coding implementation are the lack of national health data standards and protocols (Alkraiji, 2012), the fact that nearly two-thirds of the Saudi healthcare workforce consists of expatriates, making the provision of training more complicated and costly, a continuous lack of consultation with individual institutions, and a perpetual resistance to innovation from physicians (Al-Ahmadi & Roland, 2005; El Mahalli, 2015).

These specific Saudi unresolved problems were combined with the global challenges to create the questionnaire items. The questionnaire was then sent to three health information experts to review the first draft of the questionnaire.

Content validity, internal consistency reliability (Cronbach's alpha) and, factor analysis were used to describe the process of evaluating a questionnaire for the study on "Factors Influencing the Implementation of ICD-10 in Saudi

Public Hospitals". Content validity results indicate that the degree of honesty ratings given by the experts on all items in the questionnaire presented a value of validation. Cronbach's alpha and factor analysis were used to examine the reliability and validity of part 2 of the questionnaire to be used in the main study. Cronbach's alpha for all these factors showed different values between alpha = 0.604 to alpha = 0.763. These study results of internal consistency show that the items in each factor rated between adequate and excellent, in terms of the internal consistency reliability among these items. In addition, factor analysis varied between good and excellent with results indicating different loadings (0.242 to 0.785) and different moderate communalities (0.363 to 0.891) among the items in each factor. In terms of the limitations of the questionnaire, some variables included indicated a low internal consistency loading or commonalities but were well within the moderate range, based on the fundamentals of the test tools used (Allen & Yen, 2001; Hair et al., 2010).

5. Conclusion

The aim of the study was to assess the internal consistency reliability and validity of survey questionnaire designed for the study "Factors Influencing the Implementation of ICD-10 in Saudi Public Hospitals". The results in this paper indicated that the questionnaire was valid and reliable, as illustrated in Appendix.

Of particular relevance is the integration of diverse national literature into the survey questionnaire and the reflection in the results of the survey itself. Thus, the survey results reflected factors covered by all researchers covered in the literature review. As was stated in the introduction, the national literature reviews highlighted specific approaches to the implementation of clinical coding and ICD-10. In that clinical coding has not previously been executed in Saudi public hospitals, the factors focused on by the researchers from each of the countries all have relevance.

As was indicated in the introduction, the researchers undertook the named study in that it has no precedent in Saudi healthcare literature. The results of the survey questionnaire applied to a representative sample of Saudi public healthcare workers indicates near unanimity that while the Saudi MOH intends to implement ICD-10-AM, the purchase of a license is not going to produce the essential level of clinical coding required without consideration of the training needs of workers, the overcoming of staff shortages, the lack of a national supervisory body and the need for national data standards to facilitate interoperability on a national scale.

Competing Interests Statement

The authors declare that there are no competing or potential conflicts of interest.

References

Adeleke, I. T., Ajayi, O. O., Jimoh, A. B., Adebisi, A. A., Omokanye, S. A., & Jegede, M. K. (2015). Current clinical coding practices and implementation of ICD-10 in Africa: A survey of Nigerian hospitals. *American Journal of Health Research, 3*(1-1), 38-46. https://doi.org/10.11648/j.ajhr.s.2015030101.16

Al Kiyumi, R., Walker, S. M., Tariq, A., & FitzGerald, G. (2017). *Health information management professionals* [Present circumstances and future expectations].

Al-Ahmadi, H., & Roland, M. (2005). Quality of primary health care in Saudi Arabia: a comprehensive review. *International Journal for Quality in Health Care, 17*(4), 331. https://doi.org/10.1093/intqhc/mzi046

Alaraki, M. (2018). Assessing the Organizational Characteristics Influencing Quality Improvement Implementation in Saudi Hospitals. *Quality management in health care, 27*(1), 8-16. https://doi.org/10.1097/QMH.0000000000000152

Alkraiji, A. (2012). *Issues of the adoption of HIT related standards at the decision-making stage of six tertiary healthcare organizations in Saudi Arabia* (Doctoral dissertation). Available from Semantic Scholar: Full Text.

Allen, M. J., & Yen, W. M. (2001*). Introduction to measurement theory*. Waveland Press.

Alsadan, M., Elmetwally, A., Anna, A., Jamal, A., Khalifa, M., & Househ, M. (2015). Health Information Technology (HIT) in Arab Countries: A Systematic Review Study on HIT Progress. *Journal of Health Informatics in Developing Countries, 9*(2).

Babbie. (2013). *The practice of social research*. Nelson Education.

Baker, & Risley, A. J. (1994). *Doing social research* (2nd ed.). New York: McGraw-Hill Inc.

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of educational and behavioral statistics, 27*(4), 335-340. https://doi.org/10.3102/10769986027004335

Cavana, R., Delahaye, B. L., & Sekeran, U. (2001). *Applied business research: Qualitative and quantitative methods*. John Wiley & Sons Australia.

Christensen, L. B., Johnson, B., & Turner, L. A. (2011). *Research methods, design, and analysis*. Allyn & Bacon.

Council of Health Services. (2016). *Medical coding*. Retrieved from http://chs.gov.sa/En/MedicalCoding/Pages/default.aspx

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391-418. https://doi.org/10.1177/0013164404266386

De Coster, C., Li, B., & Quan, H. (2008). Comparison and validity of procedures coded With ICD-9-CM and ICD-10-CA/CCI. *Medical care, 46*(6), 627-634. https://doi.org/10.1097/MLR.0b013e3181649439

De Vaus, D. (2013). *Surveys in social research*: Routledge.

El Mahalli, A. A. (2015). Electronic health records: Use and barriers among physicians in eastern province of Saudi Arabia. *Saudi Journal for Health Sciences, 4*(1), 32. https://doi.org/10.4103/2278-0521.151407

Fan, S., & Lê, Q. (2011). Developing a valid and reliable instrument to evaluate the Web-Based Learning Environment in an Australian University Context. *Journal of Online Learning and teaching, 7*(3), 366.

Field, A. (2009*). Discovering statistics using SPSS*. Sage publications.

Fiorito, M., & Edens, T. (2016). *Health Information Management (HIM) White Paper*.

Global Medical Administration. (2019). *What is Clinical Coding*? Retrieved from https://synapsemedical.com.au/news/2019/06/04/what-is-clinical-coding/

Gorsuch, R. L. (1983). *Factor Analysis. Hillsdale, New Jersey*: Lawrence Erlbaum.

Hair, J. F., Anderson, R. E., Babin, B. J., Tatman, R. L., & Black, W. C. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Prentice Hall.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based nursing, 18*(3), 66-67. https://doi.org/10.1136/eb-2015-102129

Hersh, W. (2009). A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making, 9*(1), 24. https://doi.org/10.1186/1472-6947-9-24

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing & health*, *31*(2), 180-191. https://doi.org/10.1002/nur.20247

Heywood, N. A., Gill, M. D., Charlwood, N., Brindle, R., Kirwan, C. C., & Collaborative, N. R. (2016). Improving accuracy of clinical coding in surgery: collaboration is key. *Journal of Surgical Research*. https://doi.org/10.1016/j.jss.2016.05.023

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*(4), 523-531. https://doi.org/10.1177/00131640021970691

Independent Hospital Pricing Authority. (2019). Country licence agreement. Retrieved from https://www.ihpa.gov.au/what-we-do/products/AR-DRG-classification-system/country-licence-agreement

Jetté, N., Quan, H., Hemmelgarn, B., Drosler, S., Maass, C., Moskal, L., . . . Jakob, R. (2010). The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Medical care, 48*(12), 1105-1110. https://doi.org/10.1097/MLR.0b013e3181ef9d3e

Johanson, G. A., & Brooks, G. P. (2009). Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164409355692

Lackey, N. R., & Wingate, A. L. (1998). The pilot study: One key to research success. *Advanced design in nursing research*, 375-384. https://doi.org/10.4135/9781452204840.n15

Latimer, J. (2010). *Meeting of the who collaborating centres for the family of international classifications*.

Liamputtong, P. (2013). *Research methods in health: foundations for evidence-based practice*: Oxford University Press.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-386. https://doi.org/10.1097/00006199-198611000-00017

Murtagh, F., & Heck, A. (2012). *Multivariate data analysis* (Vol. 131): Springer Science & Business Media.

Narayanan, S., & Rose, R. J. (2017). A Survey of Health Information Management (HIM). *ERES International Journal of Computer Networks, 5*(2), 1-3. Retrieved from http://www.eresjournals.org/journals/index.php/ijcn/article/view/69.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194. https://doi.org/10.1037/a0030767

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health, 29*(5), 489-497. https://doi.org/10.1002/nur.20147

Polit-O'Hara, D., Beck, C. T., & Hungler, B. (2001). *Essentials of nursing research: Methods, appraisal, and utilization* (5th ed. Vol. 1). United States: Lippincott Williams & Wilkins.

Raykov, T. (2001). Bias of Coefficient afor Fixed Congeneric Measures with Correlated Errors. *Applied psychological measurement, 25*(1), 69-76. https://doi.org/10.1177/01466216010251005

Riazi, A. M. (1999). *A dictionary of research methods: Quantitative and qualitative*: Rahnama Publications.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social work research, 27*(2), 94-104. https://doi.org/10.1093/swr/27.2.94

Sanders, T. B., Bowens, F. M., Pierce, W., Stasher-Booker, B., Thompson, E. Q., & Jones, W. A. (2012). The road to ICD-10-CM/PCS implementation: forecasting the transition for providers, payers, and other healthcare organizations. *Perspectives in health information management/AHIMA, American Health Information Management Association*, 9(Winter).

Shepheard, J. (2009). Health Classification-a Complex World. *Health Information Management Journal, 38*(1), 4. https://doi.org/10.1177/183335830903800101

Sijtsma, K., & Emons, W. H. (2011). Advice on total-score reliability issues in psychosomatic measurement. *Journal of psychosomatic research, 70*(6), 565-572 https://doi.org/10.1016/j.jpsychores.2010.11.002

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of evaluation in clinical practice, 17*(2), 268-274. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Thompson, B., (2003). *Score reliability: Contemporary thinking on reliability issues*: Thousand Oaks, CA: Sage.

Van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing standard, 16*(40), 33-36. https://doi.org/10.7748/ns.16.40.33.s1

Wager, K. A., Lee, F. W., & Glaser, J. P. (2009). *Health care information systems: a practical approach for health care management*: John Wiley & Sons.

Walker, R. L., Hennessy, D. A., Johansen, H., Sambell, C., Lix, L., & Quan, H. (2012). Implementation of ICD-10 in Canada: how has it impacted coded hospital discharge data? *BMC health services research, 12*(1), 1. https://doi.org/10.1186/1472-6963-12-149

Yaghmaei, F. (2003). Content validity and its estimation. *Journal of Medical Education, 3*(1), 25-27. https://doi.org/10.22037/jme.v3i1.870

Appendix

Questionnaire -Final Version

Part one. Demographics

Please specify name of your hospital:

- ☐ King Fahad Medical City
- ☐ Imam Abdulrahman Alfaisal Hospital
- ☐ Al-Iman Hospital
- ☐ Al-Yamamah Hospital
- ☐ King Salman Hospital
- ☐ Buraidah Central Hospital
- ☐ Maternity and Children Hospital

Gender

- ☐ Female   ☐ Male

Age

- ☐ Less than 30 years   ☐ Between 30 to 40
- ☐ Between 41 to 50   ☐ Between 51 to 60
- ☐ More than 60 years

Nationality

- ☐ Saudi   ☐ Other

Occupational category

- ☐ Health informatics professional
- ☐ HIM or Medical records professional
- ☐ Physician
- ☐ Nurse
- ☐ Other health professional (pharmacist, health technician)
- ☐ Non-health professional (administrator)

How many years of professional experience in the health sector do you have ?

- ☐ Less than 5 years   ☐ 6-10 years
- ☐ 11-15 years   ☐ More than 15 years

Do you have a registered certificate from any recognized organization in the clinical coding system ?

- ☐ Yes   ☐ No

Have you participated in any clinical coding training courses ?

- ☐ Yes   ☐ No

Highest level of education achieved ?

- ☐ Doctoral degree   ☐ Master
- ☐ Bachelor   ☐ Diploma

What is the percentage completion of the implementation of the clinical coding in your hospital ?

- ☐ Partially implemented
- ☐ Fully implemented
- ☐ I don't know

Part Two. Factors Impacting at the National, Organizational and Health Information-level

This part contains three sections comprising statements that describe factors that may impact on the quality of implementation of clinical coding, according to health literature sources from Developed and Developing nations, including Saudi Arabia. For each statement below, please circle the response that best characterizes how you feel about the statement, according to the following criteria: 1 = Strongly Agree, 2 = Agree, 3 = Neutral or undecided, 4 = Disagree, and 5 = Strongly Disagree.

Organizational Factors		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
ORG 1	MOH hospital employees receive education and training to improve their performance and further their careers	1	2	3	4	5
ORG 2	MOH hospitals evaluate the effectiveness of staff training programs offered	1	2	3	4	5
ORG 3	MOH hospitals offer specific clinical coding training programs	1	2	3	4	5
ORG 4	MOH hospitals have sufficient health coders trained in the clinical coding	1	2	3	4	5
ORG 5	MOH hospitals have sufficient health information management specialists	1	2	3	4	5
ORG 6	MOH hospitals have sufficient health informatics specialists	1	2	3	4	5
ORG 7	My hospital uses clinical coding and is equipped with a networking connection to exchange of health information with health insurance and claims management companies	1	2	3	4	5
ORG 8	MOH hospitals are using well-qualified vendors to provide technology installations and upgrades to support the implementation of clinical coding	1	2	3	4	5
Health Information Factors						
INFO 1	MOH hospital staff understand the usefulness of clinical coding in the storage and retrieval of medical data.	1	2	3	4	5
INFO 2	MOH hospital staff understand the usefulness of clinical coding in health information management and sharing.	1	2	3	4	5

INFO 3	MOH hospital staff understand the positive impact of clinical coding on healthcare quality, through its facilitation of statistics and research.	1	2	3	4	5
INFO 4	MOH hospital staff understand the usefulness of clinical coding in processing health insurance claims and hospital funding	1	2	3	4	5
INFO 5	In MOH hospitals, physicians and coders interact in order to decide on the correct clinical codes to match the patient diagnoses and procedures	1	2	3	4	5
INFO 6	In MOH hospitals, the level of clinical diagnosis by physicians as existing in current manual records is sufficient to enable coders to apply clinical coding codes	1	2	3	4	5
National Factors						
NAT 1	A Saudi health information management supervisory organization monitors the implementation of ICD-10 in all health sectors	1	2	3	4	5
NAT 2	The MOH will fund the installation of a national network to link MOH hospitals to support ICD-10 and the management and sharing of health information	1	2	3	4	5
NAT 3	My hospital is part of an integrated and compatible electronic network established for the purpose of exchanging health information with other hospitals	1	2	3	4	5
NAT 4	The MOH will fund the cost of maintaining and upgrading health information management software at hospitals to support ICD-10	1	2	3	4	5
NAT 5	The MOH has sufficient ICD-10 course trainers	1	2	3	4	5
NAT 6	The MOH needs to provide hospitals with additional coders at the commencement of ICD-10 implementation	1	2	3	4	5
NAT 7	MOH hospitals are provided with funding specifically dedicated to the implementation of HIM and electronic health project development	1	2	3	4	5
NAT 8	The MOH funding of information technology infrastructure upgrades support the implementation of ICD-10 in hospitals	1	2	3	4	5
NAT 9	The application of standardized electronic health records (EHR) in MOH hospitals will facilitate the implementation of ICD-10	1	2	3	4	5

Copyrights