International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 7, Issue 5, 2019

Development of Post-coordination Content for the Classification of Endocrine system diseases

Mijung Kim[a],

Department of Health administration, Kwangju Women's University, 201, Yeodai-Gil, Gwangsan-Gu, Gwangju, 62396, Korea
Corresponding Author: *kmj@kwu.ac.kr

Codes of diseases are used for administrative purposes such as hospital statistics and insurance claims. In general, coding specialists read clinical documents and then assign exact codes that meet criteria of related classification. However, due to the introduction of the electronic medical record (EMR), disease codes are automatically classified codes that are pre-mapped to the diagnosis. These auto-classified codes are less accurate than the codes given by a coder. The purpose of this study is to guarantee the detailed diagnosis input by clinicians and to obtain correct administrative codes. We have developed post-coordination content to select diagnoses by combining several words. The development target was chosen as endocrine system diseases have the highest error rate when making insurance claims. Based on the lexical analysis, four main models were defined, and post-coordination content was developed according to the models. A total of 64 models and 494 instances were created. By using the post-coordination content, it was possible to assign the KCD-7 codes correctly according to the detailed level of diagnosis. Few studies have been conducted to improve the accuracy of codes, as compared to the importance of codes, for administrative purposes. It is important that the suggested post-coordination content can ensure consistency and accuracy of classification of diseases, as well as the utilization of information such as semantic retrieval of attribute units and knowledge recycling.

Key words: *Clinical data model, Interface terminology, Post-coordination, KCD-7.*

Introduction

Diagnosis is an essential piece of clinical data that is used as criteria for the therapeutic process and is necessary for insurance claims and health statistics (Statistics Korea, 2016). Therefore, clinical documents must contain both, correct diagnoses and codes for data processing.

Most hospitals have a clinical information system and EMR system. Diagnosis codes are given at the same time as diagnoses are entered because of the mapping of diagnosis onto codes (Mi-Jung, 2016). However, issues have steadily been arising about the validity of codes and the granularity of diagnosis obtained in this way (Lee and Shim, 2009; Pakhomov et al., 2006; Bae et al., 2015, Aldulaimi, 2018, Ali, 2013, Alice 2017, Al-Mulali, 2019, Altunkaya and Ates, 2018, Sharif et al., 2019).

The choice of diagnosis by clinicians under the EHR system is closely related to the accuracy of the codes. Regardless of the purpose of classification of diseases, the difference of granularity of codes depends on the doctors who input diagnoses. The same patient may have a different granularity of diagnosis according to a doctor's tendency or the convenience of an input screen. Classification of diseases is the task of assigning KCD-7 codes to a final diagnosis for administrative purposes and is the job of a professional coder. as specified in the national medical law. This work can only be performed by a person who has obtained a national license as a health information manager in Korea (Ministry of Health and Welfare). Even if clinicians enter diagnosis and assign codes, a professional coder reviews all the clinical documents and reassigns the most detailed codes about the diagnosis. Despite the increasing computerization of clinical data, professional coders are still reclassifying administrative codes.

There are limited studies on matching the codes used by doctors with the administrative codes. Domestically, there was a study researching knowledge-based models for diseases of the circulatory system for accurate disease code acquisition by using rules for precise code classification (Mi-Jung, 2016).

In the United States, studies on the detail content model and its' application in the clinical field have produced many papers (Oniki et al., 2014). However, there are limited studies that promote the accuracy of the administrative codes chosen by clinicians. As ICD-9-CM (International Classification of Diseases, Ninth Revision, and Clinical Modification), was changed to ICD-10-CM, which has five times more content than previous, there was one study about integration of postcoordination content into clinical interface terminology to support administrative coding for easy input (Eric et al., 2019). Since KCD-7 is used as administrative codes and diagnosis in clinical documentation should be written in Korean under the Medical Law in Korea, the post-coordination of ICD-10-CM cannot be applied to Korea. KCD-7 is based on ICD-10, with an extension in accordance with domestic reality (Statistics Korea, 2015).

To represent clinical content in the EMR, hospitals use unique terminologies for each hospital, including the domain of diseases that are mapped to the disease names onto its codes. Clinical terminologies in health care are distinguished into interface terminologies, reference terminologies, and classification (Rosenbloom et al., 2006). Interface terminologies represent the language used in a domain. Interface terminologies consist of terms designed for clinical representation in EHR and, through mappings onto standardized vocabularies, supports secondary uses of patient data, including clinical decision support and billing claims. In other words, interface terminologies are a set of terms designed for clinical documentation (Eric et al., 2019). In the past, interface terminology, called entry terminology and application terminology, supported various user-oriented expressions in specific computer programs. However, it has been developed into a controlled vocabulary by applying standard vocabularies, to enable accurate meaning and expression of clinical information (Rocha et al., 1994). Reference terminology is designed to give meaning to each other through the same concept code for terms that have the same sense, even though the expression is different. It is characterized by mapping concepts defined in various standard terminologies (Rosenbloom et al., 2006). Classification is the categorization of terms used in a particular area between similar ones for a specific purpose, such as statistics or reporting. ICD-10 and KCD-7 are representative examples.

Post-coordinated expression means it contains two or more concepts. Post-coordination combines concepts and allows more detail to be added to a meaning than what is represented by a single concept (IHTSDO, 2017). SNOMED CT is a representative reference terminology and a compositional terminology. The construction of post-coordinated expressions allows users to specify new meanings by referencing existing concepts. Clinical expressions use concepts of two types, one is a pre-coordinated expression type which uses a single concept identifier, and the second is a post-coordinated expression type which contains more than one SNOMED CT identifier.

Intermountain Healthcare in the US has a long history of using interface terminology and detailed clinical models to govern the storage of clinical data to ultimately facilitate clinical decision support and semantic interoperability. This is called the clinical element model (CEM). CEMs were made to be used for orders, observations, procedures, patients, encounters, referrals, and so on in EHR. CEM supports post-combinations that have been developed in the field, and over 5,000 CEMs have been developed (Oniki et al., 2014).

The purpose of this study is to develop post-combination content that helps the user select correct KCD-7 codes and exact names of diseases.

Methods

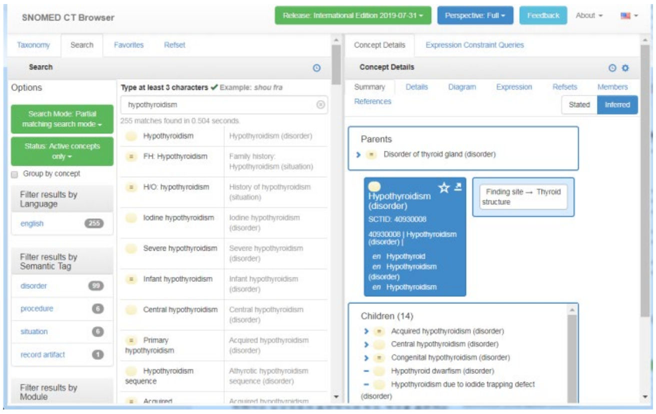
The scope to develop a post-coordination content model was endocrine, nutritional, and metabolic diseases corresponding to chapter IV of KCD-7. The codes and Hangul titles included in them are shown in Table 1.

Table1: Blocks of endocrine, nutritional and metabolic diseases in KCD-10

Block	Title
E00-E07	Disorders of thyroid gland
E10-E14	Diabetes mellitus
E15-E16	Other disorders of glucose regulation and pancreatic internal secretion
E20-E35	Disorders of other endocrine glands
E40-E46	Malnutrition
E50-E64	Other nutritional deficiencies
E65-E68	Obesity and other hyper-alimentation
E70-E90	Metabolic disorders

SNOMED CT, an international standard terminology, was selected to represent words and give a meaning to words. We used SNOMED CT online browser international version (Snomed) (see Fig.1). A link to reference terminologies, such as SNOMED CT, is essential to guarantee semantic interoperability. In this paper, the SNOMED CT code means the SNOMED CT concept code.

Fig.1. SNOMED CT Browser



The content building stage refers to the previous study (Mi-Jung, 2016).

Firstly, we analyzed the knowledge of the classification of diseases and the lexical structure in the endocrine system in KCD-7. The 494 disease Korean names were segmented into meaningful term units. Some of the terms can be divided into several words, but they are handled as one term for the convenience of post-coordination. The minimum unit name can be used alone without being combined with other terms. Words that can be used alone as a diagnosis name were set to the minimum unit called a fundamental concept, and we used them as model names. The modifiers were sorted based on the fundamental concept. Based on the lexical analysis, the attribute categories and attribute values were defined so the attributes that are needed to describe a disease could be combined.

Secondly, the post-combination model was created. The expression method was complemented by adopting the EAV (Entity-Attribute-Value) model, which is the most widely used model for storing data relating to clinical records (Dinu, 2007). KCD-7 codes were assigned to combinations that could be expressed in an EAV format.

The third step was to create an instance. KCD-7 codes and SNOMED CT codes for each instance are mapped, and SNOMED CT codes were assigned to each term for modifiers and modifier values that can be combined.

Results and Discussion

Lexical analysis

The 494 disease names of the endocrine system listed in KCD-7 were segmented into meaningful term units. The lexical analysis was done manually using MS Excel. Some of the words could be divided into several words, but they were dealt with as one word for the convenience of post combination.

Words that identified a diagnosis on their own were set as the minimum unit of the model. The minimum unit can be used as a diagnosis without being combined with other terms. Phrases treated as a single word are ‘not elsewhere classified’, ‘other and unspecified’, ‘with complications’, ‘without’, ‘without signs’, ‘other specified’, ‘subcutaneous tissue’ and ‘in diseases classified elsewhere’. Among the segmented terms, the number of ‘diabetes mellitus’ was the highest, followed by ‘with’, ‘unspecified’, and ‘diabetic’ (see Table 2). When used alone, the meaning of the word is ambiguous, but used together with other words to refer to a particular disease, the word is handled as a term combined with the related words.

The most representative disease of the endocrine system is diabetes mellitus. It is classified into various codes by the type of diabetes and the complications caused by diabetes.

Table 2: Top 10 of the most showed terms

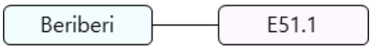
Terms	Count
당뇨병 (diabetes mellitus)	190
동반한 (with)	141
상세불명의 (unspecified)	87
당뇨병성 (diabetic)	64
기타 명시된 (Other specified)	56
기타 (other)	53
및 (and)	48

Post-coordination content model

From segmented terms, disease names that can be used alone were extracted as a minimum unit. The model was created by adding a modifier around the fundamental concept. Each segmented term was given a SNOMED CT code. There were three types of main models:

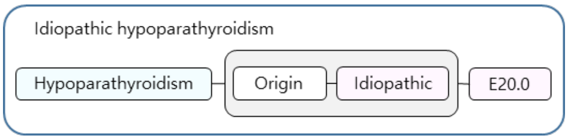
1) Single concept type: In this case, the disease name, which is the minimum unit, is used as it is, such as Beriberi.

Fig.2. Single concept type



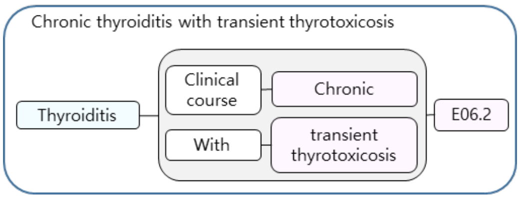
2) Single modifier type: This is the case when there is one modifier that modifies the fundamental concept.

Fig.3. Single modifier type



3) Multiple modifier type: This applies when two or more modifiers modify the fundamental concept. This can represent a disease caused by another disease or disorder.

Fig.4. Multiple modifier type (hypoparathyroidism)



The core components of the content model were fundamental concepts, modifiers, and KCD-7 codes (see Fig.5).

Fig.5. Core components of the content model

<table>
<tr><th>Component</th><th colspan="2">Description</th></tr>
<tr><td>Precoordination</td><td colspan="2">Pre-coordination diagnosis name</td></tr>
<tr><td>Fundamental concept</td><td colspan="2">Minimum words available for diagnosis</td></tr>
<tr><td rowspan="2">Modifiers</td><td>Name</td><td>Values</td></tr>
<tr><td>Group name of modifier values</td><td>Specific modifier value of fundamental concept</td></tr>
<tr><td>KCD Codes</td><td colspan="2">KCD-7 codes assigned to diagnosis</td></tr>
</table>

Build Content

Post-combination content was developed based on the four models mentioned above. The KCD-7 code depends on the unique attributes of the disease or the accompanying conditions and symptoms. 84 fundamental concepts were derived as minimum units, 21 of them were diseases associated with mineral deficiencies. We made a mineral deficiency model and 21 mineral-related instances with them. For other diseases caused by a mineral deficiency, a separate model was created with a multiple modifier type. A diabetes content model was designed and it built 186 instances using multiple modifiers. Through this, we made 64 models and 494 instances. Fig.7 shows the creation of a case of content using the diabetes mellitus model. A SCTID means SNOMED CT concept code.

Fig.7. Case of content

<table>
<tr><th>Component</th><th colspan="2">Instance</th></tr>
<tr><td>Precoordination</td><td colspan="2">Diabetes mellitus, type 2 with retinopathy
(SCTID: 422034002)</td></tr>
<tr><td>Fundamental concept</td><td colspan="2">Diabetic mellitus
(SCTID: 73211009)</td></tr>
<tr><td rowspan="2">Modifier</td><td>DM Type</td><td>Type 2
(SCTID: 258195006)</td></tr>
<tr><td>Complication</td><td>Diabetic retinopathy
(SCTID: 4855003)</td></tr>
<tr><td>KCD Code</td><td colspan="2">E11.33, H36.0</td></tr>
</table>

International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 7, Issue 5, 2019

Proposed user interface

Most hospitals in Korea often use the name of the diseases of KCD-7 when writing clinical documents. When a part of the name is searched in the existing diagnosis entry, the results are listed and one can be selected.

Since KCD-7 is pre-coordinated terminology, if a diagnosis has different modifiers such as anatomical sites and types, the names of all cases that can be combined are listed in the book. In some cases, too many results are displayed on the screen when a search word is entered on the diagnosis entry. Therefore, the user may not find the desired result on the screen.

The list of names in Fig.8, has names that do not include the type of diabetes, such as types 1 or type 2. So those names have codes beginning with E14, for unspecified diabetes.

A patient with Type 2 Diabetes mellitus and diabetic nephropathy may be given only one name of 'Diabetes mellitus, type 2' or given two names 'diabetes mellitus' and 'nephropathy'. In both cases, the correct codes cannot be obtained.

Fig.8. Existing user interface for diagnosis

Diagnosis list
Diabetes mellitus with proteinuria (fixed)(persistent)(N08.3*)
Diabetes mellitus With end-stage renal disease[ESRD](N08.3*)
Diabetes mellitus with chronic renal failure(N08.3*)
Diabetes mellitus with renal impairment NOS(N08.3*)
Diabetes mellitus With other and unspecified renal complication(N08.3*)
Diabetes mellitus with acute renal failure(N08.3*)
Diabetes mellitus with acute renal impairment(N08.3*)
Diabetes mellitus with medullary (papillary) necrosis(N08.3*)
Diabetes mellitus With ophthalmic complications
Diabetes mellitusWith diabetic nonproliferative retinopathy(H36.0*)
Diabetes mellitus with intraetinal microvascular abnormality(IRMA) , cotton-wool spot(H36.0*)
Diabetes mellitus with intraetinal microvascular abnormality(IRMA), hemorrhage(blotchy)(round)(sma
Diabetes mellitus with intraetinal microvascular abnormality(IRMA), ischemia(H36.0*)
Diabetes mellitus with intraetinal microvascular abnormality(IRMA), venous beading, looping, reduplic

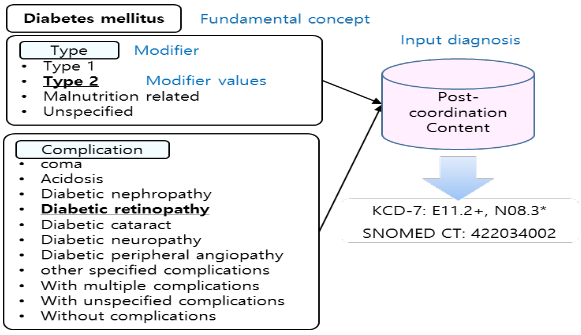
Since clinicians are end users and not professional coders, they may not know the level of classifiable details.

There may also be differences between the level of detail that clinicians use to describe a patient's condition clinically and the level of detail that a coding expert needs to obtain the administrative code. An interface terminology with postcoordination content can solve some of these problems.

We proposed a user interface using post-coordination content. Interface terminology with post-coordination content could be able to prevent missed modifiers of diagnosis, even if the clinician is not a professional coder.

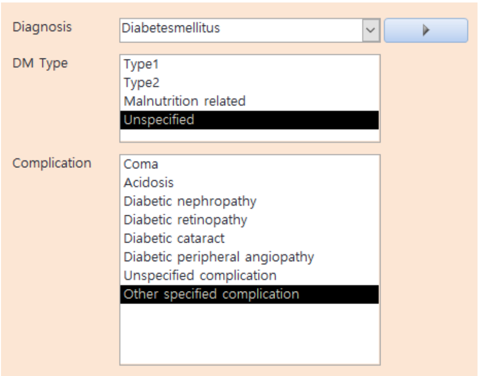
When selecting a diagnosis, the modifiers that can be combined with it are listed. In the case of diabetes mellitus, KCD-7 codes reduced according to the diabetic type, presence and complications, with related codes given together. A set code indicating the cause and condition will also be available (see Fig.9).

Fig.9. Illustration of code assignment using post-coordination content



It is possible to select multiple modifier values for the same modifier of a diagnosis. For example, when writing the diagnosis of a patient with type 2 diabetes who has both diabetic nephropathy and diabetic cataract, it is possible to select two values for a diabetic complication modifier.

Fig.10: Suggested user interface for diagnosis by the post-coordinated approach



The advantages of an interface that can input a diagnosis by combining several words are as follows: first, it minimizes the burden on clinicians to look through exhaustive lists of diseases and helps users avoid missing modifiers. Secondly, when the clinician writes a diagnosis on the clinical document, codes for administrative purposes can be obtained, no additional code classification work is required. Thirdly, coding experts may no longer be required for accurate classification of diseases.

Challenges

As content build work proceeded, several challenges emerged.

The first was how to define the fundamental concepts. The type and number of models depends on the level of the fundamental concept. For example, we had to decide whether the fundamental concept of vitamin deficiency was a set deficiency, or we had to decide whether a mineral type was used as a modifier or vitamin deficiency itself was set a fundamental concept. We defined mineral types as a modifier group and made 21 cases covered by the deficiency model, but we could have made 21 separate models by pre-combining deficiency and mineral types. To consistently collect detailed levels of clinical content, we needed a consensus on how the clinical information represented attributes and values for expressing characteristics (Oniki et al., 2014), (Goossen et al., 2010). The post-coordination model was used to obtain detailed diagnosis and codes, but we should consider both the convenience of end-users and the usability of the data.

Another issue was that ambiguous situations need to be expressed as clear facts. We had to figure out ways to specify ambiguous modifiers such as 'other', 'other specified', 'other and unspecified', and 'not elsewhere classification'. Also, KCD-7 may or may not provide codes for some modifiers like ‘unilateral’ and ‘bilateral’ of a paired body part. Someone may also want to use ‘right’ or ‘left’ as a modifier, but these are not in KCD-7. As such, it was necessary to collect detailed diagnosis names for each clinician and make instances so that they could use clear terms not vague ones.

The completeness of mapping between SNOMED CT and interface terminology depends on the scope, level of detail provided by the two schemes, and the precision of mapping required to safely meet the intended mapping use case.

We identified a problem, whereby users could select more than one diagnosis separately, which needs to be done in a post coordinated diagnosis despite the use of post-coordination content. Correct codes can be obtained by selecting a corporal name that combines two corporal names, but an incorrect code is obtained by selecting two corporal names separately. To achieve our end goal, where correct codes are selected by clinicians without professional coders oversight, we added a knowledge item where users can leave a message when creating post-coordination content. A knowledge item can be appropriately used in the implementation layer.

Conclusion

We have developed post-combination content to classify diseases of the endocrine system with correct disease codes. Post-combination content is extensible and recyclable because it can refine various diagnostic names by adding new modifiers. Applying the post-combination content suggested in this study, to the diagnosis entry interface, helps doctors write the exact diagnosis, solves the problem of poor quality of medical information due to incorrect codes, and can fundamentally limit the causes of insurance claim rejection.

Accurate classification of diseases will ultimately contribute to improving the quality of care and clinical information, increasing the reliability of statistical data at the national level, and enhancing the value of health information.

Since this study is limited to endocrine diseases, it is necessary to expand the content to other categories in the future. Moreover, development of an application program for managing post-coordination content is required. In addition, research on an interface that can maximize user convenience is needed.

REFERENCES

Aldulaimi, S. H. (2018). The Influence of National Culture on Commitment that Produce Behavioral Support for Change Initiatives. International Journal of Applied Economics, Finance and Accounting, 3(2), 64-73.

Ali, A. (2013). How to Differentiate between 'Leadership'and 'Management'Function in Organization: A Review of Scholarly Thoughts. International Journal of Economics Business and Management Studies, 2(1), 38-44.

Alice, H., 2017. Closing the rhetoric reality gap: Effectively implementing engagement and wellbeing policies in Queensland state secondary schools. International Journal of Innovation, Creativity and Change, 3(3): 124-139.

Al-Mulali, U., Tang, C. F., Tan, B. W., & Ozturk, I. (2019). The nexus of electricity consumption and economic growth in Gulf Cooperation Council economies: evidence from non-stationary panel data methods. Geosystem Engineering, 22(1), 40-47.

Altunkaya, H., & Ates, A. (2018). Sources of Reading Anxiety among the Learners of Turkish as a Foreign Language. Asian Journal of Education and Training, 4(3), 161-169.

Bae, S.O., Kang, K.W., Boo, Y.K., Lee, Y., Cheo, H.S. & Choi, H.Y. (2015). A study on the difference in disease coding of doctors, medical insurance review nurses and medical record administrators based on coding simulation. Journal of Health Informatics and Statistics, 40(3): 161-174.

Dinu, V., Nadkarni, and Prakash, (2007). Guidelines for the effective use of entity-attribute-value modeling for biomedical databases. International Journal of Medical Informatics, 76: 769–779.

Eric, R., Steven, R., Andrew, S. K., Matthew, C. & Frank, N.-R. (2019). Integration of postcoordination content into a clinical interface terminology to support administrative coding, appl clin inform. Applied Clinical Informatics, 10(1): 51-59.

Goossen, W., Goossen-Baremans, A. & van der Zel, M. (2010). Detailed clinical models: A review. Healthcare Informatics Research, 16(4): 201-215.

IHTSDO, (2017). Snomed CT starter guide (2017.7.28 version) [Online]. Available: https://confluence.ihtsdotools.org/download/
attachments/28742871/doc_StarterGuide_Current-en-US_INT_20170728.pdf?version=3&modificationDate=1501254629000&api=v2

Lee, J.H. & Shim, M.S. (2009). The accuracy of the ICD-10 code for trauma patients visiting on emergency department and the error in the ICISS. Journal of Trauma and Injury, 22(1): 108-115.

Mi-Jung, K. (2016). A knowledge-based model for efficient classification of disease. Jeju: Jeju National University, Computer Engineering.

Ministry of Health and Welfare, Medical service technologists, etc. act, 2 of Article 1, Clause 2.

Oniki, T.A., Coyle, J.F., Parker, C.G. & Huff, S.M. (2014). Lessons learned in detailed clinical modeling at Intermountain Healthcare. J Am Med Inform Assoc, 21(6): 1076-1081.

Pakhomov, S.V., Buntrock, J.D. & Chute, C.G. (2006). Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. Journal of the American Medical Informatics Association, 13(5): 516-525.

Rocha, R. A., Huff, S. M., Haug, P. J. & Warner, H. R. (1994). Designing a controlled medical vocabulary server: The VOSER project. Computers and Biomedical Research, 27(6): 472-507.

Rosenbloom, S. T., Miller, R. A., Johnson, K. B., Elkin, P. L. & Brown, S. H. (2006). Interface terminologies: Facilitating direct entry of clinical data into electronic health record systems. Journal of the American Medical Informatics Association, 13(3): 277-288.

Statistics Korea, (2015). Korean standard classification of disease 7th revision, Korean medical record association, Seoul.

Statistics Korea, (2016). Classification of disease manual. Daejeon: Statistics Korea.

Snomed CT Online brower. Available: https://browser.ihtsdotools.org

Sharif, A., Raza, S. A., Ozturk, I., Afshan, S. (2019). The Dynamic Relationship of Renewable and Nonrenewable Energy Consumption with Carbon Emission: A global study with the application of heterogeneous panel estimations. Renewable Energy, 133(2019), 685-691.