An algorithm to identify patients with type 2 diabetes mellitus among undocumented migrants using data on drugs dispensation by charities: a pilot study

G Fiorini [(1)], C Cerri [(2)], S Bini [(3)], AE Rigamonti [(1)], N Marazzi [(4)], A Sartorio [(4)], SG Cella [(1)]

(1) Department of Clinical Sciences and Community Health, University of Milan, Italy;
(2) Department of Medicine and Surgery, University of Milano-Bicocca, Italy;
(3) Osservatorio Donazione Farmaci Banco Farmaceutico Onlus, Milan, Italy; 4IRCCS – Istituto Auxologico Italiano, Experimental Laboratory for Auxo-endocrinological Research, Milan and Verbania, Italy

***CORRESPONDING AUTHOR:** Silvano Cella, Department of Clinical Sciences and Community Health, University of Milan - Via Vanvitelli, 32 – 20129 Milan – Italy - Phone: +39 02 503 17015 - e-mail: silvano.cella@unimi.it*

***DOI:* 10.2427/12700**
Accepted on December 28, 2017

ABSTRACT

Background and objectives: The composition of the Italian population is rapidly changing due to the massive phenomenon of migration. Health data of natives and documented migrants are easily accessible, but this does not happen for the growing population of undocumented migrants. Since type 2 diabetes mellitus is one of the main causes of morbidity and mortality also in developing countries, we propose a method to identify the undocumented diabetics based on the only available data, i.e. the drugs dispensed to them by charities.
Methods: In this pilot study, we analysed the databases of two Italian non-governmental organisations (NGOs) containing the records of all drugs dispensed to 12,386 undocumented migrants from January 1st, 2013 to December 31st, 2016, with the aim to identify patients treated for type 2 diabetes (T2D) on the basis of demographic data and dispensed medicines. Medications were classified according to the Anatomical Chemical Therapeutic (ATC) classification. All the patients with at least one dispensation per year of any A10 (antidiabetic) drug were selected. An algorithm to match this observation with the diagnosis of T2D was developed.
Results: The algorithm identified 660 patients with T2D. An *ex-post* evaluation carried out on 400 of these patients demonstrated a full concordance with the diagnostic records. When our patients were grouped according to ethnicity, we found that all ethnic groups contributed a comparable percentage of patients with T2D. Also, no difference was seen between the group of EU citizens living in poverty cared for by the NGOs and any of the ethnic groups.
Conclusions: This algorithm can be tested in other situations to identify patients treated for T2D when no diagnostic codes are available; if its efficacy and reliability are confirmed, this method could become a useful tool for different aspects of public health.

Key words: undocumented migrants; type 2 diabetes mellitus; algorithm

INTRODUCTION

The International Classification of Diseases (ICD) was introduced for the first time in the seventies as a tool to code the causes of death, and subsequently it was adjusted to classify diseases into categories for administrative and reimbursement purposes [1]. Currently, its ninth and tenth revisions (ICD-9 and ICD-10) are widely used.

Usually, health researchers use the administrative ICD databases in the study of diseases, mainly for epidemiological evaluations and death cause analysis. Thus, these electronic data bases are turned into a disease surveillance tool, but some limitations exist, due to the fact that this is not the original aim for which they were designed. For example, it has been demonstrated that death causes can be more accurately defined using a statistical method able to reduce the impact of misclassification instead of the raw ICD-9 (and ICD-10) codes [2]. For outpatients, problems can raise from the fact that, for every consultation, the physician can arbitrarily choose the code of the presenting complaint or that of an underlying chronic condition. For in-patients, there should be a more accurate classification, since ICD-9 are used for billing purposes by skilled personnel, but still this does not prevent possible medical inaccuracy. Typically, two different codes can sometimes be used for the same clinical problem, while the same code can describe two different clinical entities. This is the case, for example, for the ICD code 410.9, comprising both myocardial infarction with elevation of the ST segment and myocardial infarction without elevation of the ST segment [3].

To increase the accuracy and reliability of electronic databases, specific algorithms are developed, taking into account other available information in adjunct to administrative codes. This is the case, for example, for myocardial infarction [4] and diabetes [5-7]. Getting complete and reliable information is of paramount importance for any public health intervention and especially to implement, in the coming years, adequate and sustainable policies to deal with the increasing and multifaceted epidemic of chronic non-communicable diseases [8]. This problem is made more complicated by the changing composition of the resident population, as a consequence of continuous migration from areas of war, poverty and persecution and by the fact that undocumented migrants, escaping medical surveillance, are a growing part of this migratory flux [9]. Having reliable epidemiologic data on chronic diseases as diabetes and hypertension in such a peculiar population is therefore becoming mandatory but, of course, there are no administrative databases available. For diabetes, a starting point can be the dispensation of insulin and oral hypoglycaemic agents to undocumented migrants cared for by charities. These charities keep a record of the drugs they distribute and they usually collect some other data for each patient: age, gender and ethnicity; only few keep accurate diagnostic records using ICD-9 codes. A diagnostic tool as an algorithm for diabetes screening has therefore to rely on this sparse available information.

In this study we describe an algorithm that we developed to identify patients with T2D in a large population of undocumented migrants and the evaluation of its reliability.

METHODS

Setting and data source

In Milan, Italy, there are several NGOs providing medical assistance to the steadily growing population of undocumented migrants. All of them have to meet the standards required by our national health authority for medical activity.

Among them, two are bigger, have excellent standards and keep electronic records of all the persons seen in their outpatient clinics; these charities are the "Opera San Francesco" (OSF) and the "Fratelli di San Francesco" (FSF). In their data bases all the patients are identified with a unique alphanumeric code and their demographic data are recorded together with their ethnicity and medical history. A record is generated for every consultation, detailing the drugs dispensation, if any. In some cases (but not always) the records also contain the ICD-9 code related to the main diagnosis. For this study, we obtained both from OSF and from FSF the data of patients seen from January 1st 2013 to December 31st 2016. Demographic data of this population, for a total of 12,386 patients with 60,325 drug dispensations are shown in table 1. No significant differences in age distribution were found among the different ethnic group (P=NS).

Design of the study

This study is composed of three parts: 1) generation of an algorithm to identify diabetic subjects and to differentiate those with T2D from those with Type 1 diabetes (T1D) among undocumented migrants; 2) application of the algorithm to the population of the study and *ex-post* control of its reliability in all the patients for whom diagnostic data were available; 3) assessment of the burden of T2D in different ethnic groups. In the two databases all prescriptions were recorded using the ATC classification [10]. We therefore analysed both of them to search for any dispensation with the A10 code (comprising all medicines used in diabetes treatment). We then looked for all the subjects on insulin aged less than 35 years as potential patients with T1D (7). From the remaining patients we initially excluded women of childbearing age on metformin, to avoid subjects possibly treated for polycystic ovary syndrome (POS), as suggested by others

TABLE 1. Overall population demographics

Ethnic group	Number of patients		Percentage of patients		Mean age±SD (years)	
	Males	Females	Males	Females	Males	Females
Italy and EU	173	401	3.60%	5.30%	41±18	50±18
Eastern Europe	1644	1124	34.17%	14.84%	45±15	41±16
Northern Africa	507	2297	10.54%	30.34%	38±15	38±11
Sub-Saharan Africa	277	1011	5.76%	13.35%	45±17	35±12
Asia and Middle East	439	1646	9.12%	21.74%	44±14	36±11
Latin America	1771	1093	36.81%	14.43%	41±15	41±15

[7], but this not being the case, they were reintroduced. The remaining patients were considered to have T2D. The flowchart for the final version of the algorithm is shown in fig 1. We also analysed the number of dispensations of anti-diabetic drugs for each patient since, as suggested by other authors [6], a further "selectivity filter" consisting of at least 2 drug dispensations may be useful to avoid diagnostic, prescription or dispensation errors. Finally, diabetic patients were grouped according to their ethnicity in five groups as previously described [11] and the relative impact of T2D was calculated for each group of patients.

Statistical analysis

Statistical calculations were performed with the SPSS statistical package. ANOVA with the Bonferroni correction was used for examining differences of mean age among ethnic groups and to evaluate differences of age among T2D patients in the different ethnic groups. To assess the relative impact of T2D among the different ethnic groups the Chi-square statistics were used. The same was done to assess the concordance between the diagnosis of T2D obtained with our algorithm and the ICD-9 codes. All data are expressed as mean ± SD, or as percentage, as appropriate.

RESULTS

Using our algorithm we could identify 683 patients, in a population of 12,386 undocumented migrants, with at least one dispensation of antidiabetic drugs. Among them, 23 were aged less than 35 years and were treated only with insulin. So, we considered them as potential T1D patients. The remaining 660 patients were therefore possibly affected by T2D.

At first we excluded from the total count of patients on antidiabetic drugs women of childbearing age on metformin (N = 50). However, when we checked their ICD-9 codes, we found that all of them had a clinical diagnosis of T2D. Therefore this step was skipped in the final version of the algorithm.

To assess the reliability of our algorithm, the diagnosis of T2D obtained with its use was compared with all the ICD 9 codes actually available in the databases (i.e. 400 subjects). We could not check the remaining 260 subjects identified as diabetics by the algorithm, since no diagnostic informations were available for them.

As a control group we have randomly selected a sample of patients of equal size (400 subjects), categorized as non-diabetics by the algorithm.

In all the patients *ex-post* analysed (400 diabetics and 400 non-diabetics) we found a 100% concordance for the diagnosis obtained with the algorithm and the diagnosis based on the ICD-9 codes.

When we compared the subjects receiving 1 dispensation of A10 medicines in one year with those with 2 and 3 or more, we found little differences. Only 88 patients (13.3%) had a single dispensation in one year, but these would have been left out if we had taken into account only those with 2 or more dispensations per year. Moreover, considering only the patients with 3 or more dispensations per year would result in the loss of 82 additional subjects (12.4%), thus determining the overall loss of 170 diabetic patients (Fig 2).

Finally, we looked at the percentage of patients with T2D in the various ethnic groups. Obviously, these percentages give no insight into the prevalence and incidence of diabetes in these populations, since we have no measure of their dimensions. This information only gives an estimate of the relative importance of diabetes among the different diseases for which undocumented migrants seek medical assistance. These data are shown in table 2. As can be seen no differences in mean age among the various groups were noticed (P=0.10). On the contrary, for sex distribution a slight prevalence of females was present in Asians and Northern Africans (P=0.05).

DISCUSSION

Worldwide, the total number of migrants exceeds 200 million and is steadily increasing. In Italy, immigrants already account for almost 10% of the population. By 2050, in the United States 1 in 5 residents is expected to be an immigrant [12] and in the United Kingdom the majority of residents

TABLE 2. Patients with T2D in the different ethnic groups

Ethnic group	Number of patients		Percentage of patients		Mean age±SD (years)	
	Males	Females	Males	Females	Males	Females
Italy and EU	9	26	5.5%	6.9%	46±6	63±6
Eastern Europe	63	62	3.8%	5.5%	49±12	49±12
Northern Africa	13	116	2.6%	5.1%	48±10	44±10
Sub-Saharan Africa	22	48	7.9%	4.7%	62±8	42±9
Asia and Middle East	26	117	5.9%	7.1%	50±10	42±11
Latin America	102	56	5.8%	5.1%	51±15	45±9

aged 50 and over will be non-native [13]. Beside this growing (and ageing) population of documented migrants, there is a population of undocumented migrants, the number of which, though unknown, is estimated to account for up to 4% of residents [14].

While medical databases are available for legal migrants in the majority of countries including Italy [15], data on the health status and health needs of undocumented migrants are scanty to date. This poses a serious problem, since reliable information is fundamental not only to meet the present health needs of these populations of migrants, but also to implement adequate public health strategies and resource allocation for the coming years.

The problem is urgent, since what was optimistically called "the healthy migrant effect" [16] starts now to be reconsidered, not only for the inevitable exposure of settled migrants to ageing and environmental noxious factors such as smoking [17] in host countries, but also because many migrants, especially if they are refugees, have already medical problems upon arrival [18,19] and migration itself is an unhealthy and dangerous event [18].

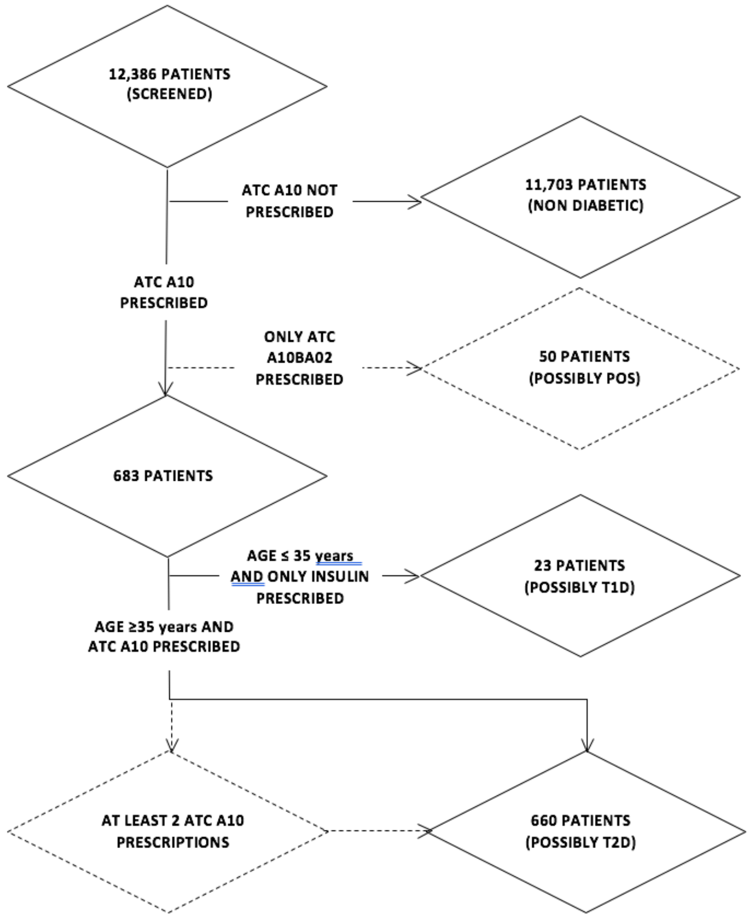
For undocumented migrants, we have recently demonstrated that infants and children are affected by the same diseases of their native counterparts [20], while adults seem to show a significant burden of chronic non-communicable diseases [11].

On the basis of these considerations we thought that a tool to clearly identify populations of undocumented migrants with chronic illnesses could be useful in a number of ways. Usually these tools are developed with the help of detailed databases containing diagnostic codes (e.g. ICD-9) or medical records, and this is the case also for diabetes [5,21]. Recently, two excellent algorithms have been developed: one allows to distinguish and classify patient with T1D and T2D on treatment in primary care, starting from the Read diagnostic codes, age, and type of treatment [7]. The second (DIABECOLUX algorithm) is able to identify patients with treated T2D, using medico-administrative data containing health insurance claims, with demographic patients' data, and type and number of hypoglycaemic drugs reimbursed in a given period of time [6]. In our case, with a population of undocumented migrants, we could not rely either on the occasional ICD-9 codes or on complete medical records, let alone insurance reimbursements. So, in our attempt, we were forcibly taken to consider only age and dispensed antidiabetic agents. Our algorithm was able to identify all treated patients with T2D, but we had a minor problem when we excluded women of childbearing age taking only metformin. This was done to exclude those on metformin treatment for POS [7], but when we checked the records manually we found that all of them had a diagnosis of T2D. This can be due to a lower prevalence of POS in these populations than in Caucasians, to a lesser awareness of this problem by doctors operating in the charities, to a tendency to treat this disease differently in these populations, or to a combination of these factors. We therefore decided to skip this step in the final version of our algorithm. In this we are also supported by other authors, who, as well, do not use this correction in their algorithm, though they developed it in a population assisted by an EU National Health Service [6].

Another possibly problematic issue is the choice of an age threshold of 35 years to distinguish patients with T1D from those with T2D, as suggested by other authors [7] in agreement with what is proposed by the Royal College of General Practitioners [22]. A higher age threshold has also been used [6], but in our patients we have chosen a lower limit, not only in consideration of their relative low mean age, but also for the earlier onset of diabetes in some ethnical groups [23]. Considering as potential patients with T1D all the subject on insulin aged less than 35 years does not exclude the pediatric patients who develop T2D, a reality which must not be underestimated also in consideration of the early onset of severe complications [24].

A few considerations have to be made on the fact that we have chosen one single dispensation of any A10 drug in one year as sufficient evidence for diagnosing a patient as diabetic. Different approaches can be found in literature. Some authors consider necessary two or three prescriptions per year to avoid diagnostic errors [6], while others suggest that when ICD-9 codes are available, a single physician's service claim in two years is enough [5]. Increasing the number of observations/events to increase specificity did not seem acceptable in our unstable population of undocumented migrants because of the unacceptable reduction in sensitivity. Indeed, our algorithm

FIGURE 1. Flowchart of the algorithm. The dotted lines indicate steps skipped in the final version.

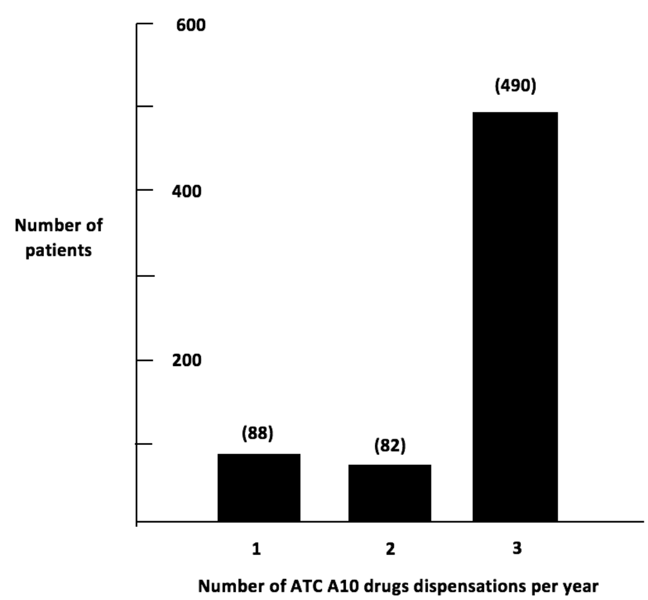


T1D=Type 1 diabetes; T2D=Type 2 diabetes; POS=Polycystic Ovary Syndrome. See text for further details

yielded no false positive, while dropping the subjects with only one prescription we would have lost 13.3% of patients. However, it is interesting to notice that more than 70% of our patients had three or more dispensations per year. This could be an indicator of a certain adherence to treatment, which is quite unexpected in such a population, since it depends from many factors, including attitudes, intentions and perceived behavioural control [24] and religious beliefs and habits [26].

When we looked at the contribution given by the different ethnic groups, we found that diabetes is a reason for seeking free medical help for all of them, without significant differences, as demonstrated by the fact that the percentage of diabetics is comparable in all the groups. Also, no difference was seen between the EU group and any of the other ethnic groups. This is partly surprising since diabetes is known to have an increased prevalence in some populations of migrants [18,27,28], so it was reasonable to expect a greater contribution by the ethnic groups with a greater prevalence. Of course we have no indication on the prevalence of diabetes in the various ethnicities to which our groups of patients belong.

FIGURE 2. Number of diabetic patients with only 1, only 2 and 3 or more A10 drug dispensations per year



Therefore we can draw no general conclusions, also because other factors, besides ethnicity, are playing a role in the ethnic disparities in T2D, and socioeconomic factors can be more relevant than ethnicity itself [29]. We can only say that the algorithm described in the present study is able to reliably detect patients treated for T2D without recurring to administrative codes, but only on the basis of drug prescriptions/dispensations, though we are aware of its many limitations. If our preliminary observations are confirmed in better defined and sampled populations, they could represent the starting point for further studies aimed to investigate cardiovascular comorbidities in diabetic undocumented migrants, in consideration of the increasing prevalence of T2D, its diminishing age of onset and its association with early complications, especially in certain ethnic groups [23,24,27,30].

Acknowledgements

Since this is a retrospective study, no ethical approval was required. All data were completely and permanently anonymised.

We thank the "Osservatorio Donazione Farmaci" of the Banco Farmaceutico Foundation for its support. We are equally indebted with the "Opera San Francesco" and the "Fratelli di San Francesco" for allowing us to use their databases.

Funding acknowledgement and conflict of interest

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

The authors have no conflicts of interest to disclose, nor they have specific financial interests and relationships and affiliations relevant to the subject of this manuscript.

References

1. WHO, Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, vol. 1, World Health Organization, Geneva, Switzerland, 1977.
2. Foreman KJ, Naghavi M, Ezzati M. Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death. Population Health Metrics 2016; 14:14-23
3. Kottle TE, Baechler CJ. An algorithm that identifies coronary and heart failure events in the electronic health record. Prev Chronic Dis 2013; 10:120097. DOI:http://dx.doi.org/10.5888/pcd10.120097
4. Coloma PM, Valkhoff VE, Mazzaglia G, et al. on behalf of the EU-ADR Consortium. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. BMJ Open 2013; 3:e002862. doi:10.1136/bmjopen-2013-002862
5. Hux JE, Ivis F, Flintoft V, et al. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002; 25:512-16
6. Renard LM, Bocquet V, Vidal-Trecan G, et al. An algorithm to identify patients with treated T2D using medico-administrative data. BMC Medical Informatics and Decision making 2011; 11:23-29
7. Sharma M, Petersen I, Nazareth I, et al. An algorithm for identification and classification of individuals with type 1 and T2D mellitus in a large primary care database. Clinical Epidemiology 2016; 8:373-80
8. Ali MK, Jaacks LM, Kowalski AJ, et al. Non-communicable Diseases: Three Decades Of Global Data Show A Mixture Of Increases And Decreases In Mortality Rates. Health affairs 2015; 34:1444-55
9. Tsiodras S. Irregular migrants: a critical care or a public health emergency. Intensive Care Med 2016; 42:252-55
10. Norwegian Institute of Public Health. ATC: structure and principles. Oslo, http://www.whocc.no/atc/structure_and_principles/; 2011.
11. Fiorini G, Cerri C, Bini S, et al. The burden of chronic non-communicable diseases in undocumented migrants: a 1-year survey of drugs dispensation by a non-governmental organization in Italy. Public Health 2016; 141:26-31
12. Passel JS, Cohm D. US population projections: 2005-2050. Washington, DC: Pew Research Center, 2008. Available from: URL: http://www.pewsocialtrends.org/files/2010/10/85.pdf
13. Lievesly N. The future ageing of the ethnic minority population in England and Wales, older BME people and financial inclusion report. Runnymede and the Centre for Policy on Ageing, July 2010
14. Karl-Trummer U, Metzler B, Novak-Zezula S. Health care for undocumented migrants in the EU: concepts and cases. Brussels, IOM, 2009. URL: http://www.migrant-health-europe.org/files/Health Care for Undocumented Migrants_Background Paper(6).pdf
15. Firenze A, Restivo V, Bonanno V, et al. Health status of immigrants arrived to Italian coast. Epidemiol Prev 2014; 39 (suppl. 2):78-82
16. Razum G, Zeeb H, Akgun HS, et al. Low overall mortality of Turkish residents in Germany persists and extends into a second generation: merely a healthy migrant effect? Trop Med Int Health 1998; 3:297-303.
17. Aspinall PJ, Mitton L. Smoking prevalence and the changing risk profiles in the UK ethnic and minority populations: implications for stop smoking services. Public Health 2014; 128:297-306
18. Trovato A, Reid A, Takarinda KC, et al. Dangerous crossing: demographic and clinical features of rescued sea migrants seen in 2014 at an outpatient clinic at Augusta Harbor, Italy. Conflict and Health 2016; 10:14-25
19. Diaz E, Kumar BN, Gimenu-Feliu LA, et al. Multimorbidity among registered immigrants in Norway: the role of reason for migration and length of stay. Tropical Medicine and International Health. 2015; DOI: 10.1111/tmi.12615
20. Bini S, Clavenna A, Rigamonti AE, et al. Drugs delivery by charities. A possible epidemiologic indicator in children of undocumented migrants. J Immigrant Minority Health 2016; DOI 10.1007/s10903-016-0471-6
21. Ng E, Dasgupta K, Johnson JA. An algorithm to differentiate diabetic respondents in the Canadian community health survey. Health Rep 2008; 19:71-79
22. Royal college of General Practitioners. Coding, classification and diagnosis of diabetes. Departmen of Health Gateway clearance

number 14664. Pp 52; London, march 2011.
23. Wilkinson E, Waqar M, Sinclair A, et al. Meeting the challenge of diabetes in ageing and diverse populations: a review of literature from the UK. Journal of Diabetes Research. 2016; DOI 10.1155/2016/8030627
24. Pinhas-Hamiel O, Zeitler P. Acute and chronic complications of T2D mellitus in children and adolescents. Lancet 2007; 369:1823-31.
25. Fai EK, Anderson C, Ferreros V. Role of attitudes and intentions in predicting adherence to oral diabetes medications. Endocrine Connections 2017; 6:63-70
26. Michael EM. Effectiveness and safety of newer antidiabetic medications for Ramadan fasting diabetic patients. Journal of diabetes research. 2016 DOI: 10.1155/2016/6962574
27. Montesi L, Caletti MT, Marchesini G. Diabetes in migrants and ethnic minorities in a changing world. World J Diabetes 2016; 7:33-44
28. Agyemang C, Meeks K, Beune E, et al. Obesity and T2D in sub-Saharan Africans – Is the burden in today's Africa similar to African migrants in Europe? The Rodam study. BMC Medicine 2016. DOI 10.1186/s12916-016-0709-0
29. Piccolo RS, Subramanian SV, Pearce N, et al. Relative contributions of socioeconomic, local environmental, psychosocial, lifestyle/behavioral, biophysiological, and ancestral factors to racial/ethnic disparities in T2D. Diabetes Care 2016; 39:1208-17
30. Amutha A, Anjana RM, Venkatesan U, et al. Incidence of complications in young-onset diabetes: comparing type 2 with type 1 (the young diab study). Diabetes research and clinical practice 2016. DOI 10.1016/j.diabres.2016.11.006

*