<div align="center">Original Article</div>

Suggestions for Better Data Presentation in Papers: An Experience from a Comprehensive Study on National and Sub-national Trends of Overweight and Obesity

Shirin Djalalinia MSc PhD Candidate[1,2,3], Roya Kelishadi MD[4], Mostafa Qorbani PhD[5,1], Niloofar Peykari MSc PhD Candidate[1,2,3], Amir Kasaeian PhD Candidate[6,1], Sahar Saeedi Moghaddam BS[7,1], Kimiya Gohari BS[7,1], Bagher Larijani MD[2], Farshad Farzadfar MD MPH DSc•[1,2]

Abstract

Background: The importance of data quality whether in collection, analysis or presenting stage is a tangible and undeniable scientific fact and the main objects of researches implementation.

Objective: This paper aims at explaining the main problems of the Iranian scientific papers for providing better data in the field of national and sub-national prevalence, incidence estimates and trends of obesity and overweight.

Methods: To assess and evaluate papers, we systematically followed an approved standard protocol. Retrieval of studies was performed through Thomson Reuters Web of Science, PubMed, and Scopus, as well as Iranian databases including Irandoc, Scientific Information Database (SID), and IranMedex.

Using GBD (Global Burden of Diseases) validated quality assessment forms to assess the quality and availability of data in papers, we considered the following four main domains: a) Quality of studies, b) Quality report of the results, c) Responsiveness of corresponding authors, and d) Diversity in study settings.

Results: We retrieved 3,253 records; of these 1,875 were from international and 1378 from national databases. After refining steps, 129 (3.97%) papers remained related to our study domain. More than 51% of relevant papers were excluded because of poor quality of studies. The number of reported total population and points of data were 22,972 and 29 for boys, and 38,985 and 47 for girls, respectively. For all measures, missing values and diversities in studies' setting limited our ability to compare and analyze the results. Moreover, we had some serious problems in contacting the corresponding authors for complementary information necessary (Receptiveness: 17.9%).

Conclusion: As the present paper focused on the main problems of Iranian scientific papers and proposed suggestions, the results will have implications for better policy making.

Keywords: Data presentation, data quality, obesity, overweight, pediatric

Introduction

Assessment of health related values and indicators as well as the estimates of their levels and trends are the most essential requisites for evidence-based health policies.[1–2] Different studies have emphasized that overweight and obesity are among the most important health priorities with increasing trends and need special attention and response.[3]

Undoubtedly, decisions about the design, feasibility, funding, implementation, and management of such multi dimensional programs require the development and use of accurate information and scientific evidence provided through related papers and reports.[4–5]

The validity and sensitivity of data analyses depend on the type and quality of the inputs that obtain from papers, reports, and other available data sources.[6]

The importance of data quality whether in collection, analysis or presenting stage is a tangible and undeniable scientific fact and the main objects of researches implementation. Moreover, the transparency and completeness of presented data would lead to more accurate and more effective policies.[7]

Two important aspects of data quality are the quality of presented data and data availability. The former refers to the accuracy of data, and the latter includes the extent to which materials, data, and associated protocols are promptly available to other researchers.[8–9]

Based on our experiences in estimating the National and Sub-National Prevalence, Trend, and Burden of Cardiometabolic Risk Factors in Iranian Children and Adolescents,[10] as a sub-component of National and Sub-nation Burden of Diseases, Injuries, and

Authors' affiliations: [1]Non-communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran. [2]Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Research Institute, Tehran University of Medical Sciences, Tehran, Iran. [3]Development of Research and Technology Center, Deputy of Research and Technology, Ministry of Health and Medical Education, Tehran, Iran. [4]Child Department of Pediatrics, Child Growth and Development Research Center, Research Institute for Primary Prevention of Non-communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran. [5]School of Medicine, Community Medicine Department Alborz University of Medical Sciences, Karaj, Iran. [6]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. [7]Department of Biostatistics, Faculty of Paramedical sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

•**Corresponding author and reprint:** Farshad Farzadfar MD MPH DSc, Non-communicable Diseases Research Center, Endocrinology and Metabolism Research Institute, Tehran University of Medical Sciences, Tehran, Iran. Address: No. 4, 4th Floor, , Ostad Nejatollahi St., Enqelab Ave., Tehran 1599666615, Iran. Telefax: 98-21-88913543; E-mail: f-farzadfar@tums.ac.ir.
Accepted for publication: 12 November 2014

Risk Factors from 1990 to 2013 (NASBOD study) in Iran,[11] we faced numerous challenges in both quality and availability of papers' data.

The present paper is intended to reveal the main problems of Iranian scientific papers and some data provision based comments that can be helpful for providing better data in the field of national and sub national prevalence, incidence estimates and trends of childhood obesity and overweight.

Materials and Methods

To assess and evaluate papers, we systematically followed an approved standard protocol. The details of aim and methods are described previously and here we refer to some points in brief.[10]

We conducted a systematic review to identify the trend of overweight and obesity in a 24-year interval from 1990 to 2013 in Iranian children and adolescents. Retrieval of studies was performed through Thomson Reuters Web of Science, PubMed, and Scopus, as well as Iranian databases including Irandoc, Scientific Information Database (SID), IranMedex based on our search strategy. For more accuracy, in addition to the searched articles and national and sub-national studies, we detected all papers that were cited in the references of retrieved papers and reports.

To assess the quality and availability of data in papers, we considered the following four main domains: a) Quality of studies, b) Quality report of the results, c) Responsiveness of corresponding authors, and d) Diversity in study settings.

We used the GBD (Global Burden of Diseases) validated quality assessment form which has three parts: general information, sampling quality, and measurement quality. Based on the total score, the quality of article might be ranked as excellent (13–18), good (6–12) or poor (≤5). Papers that had poor ratings were excluded and data were extracted from moderate and high quality studies (12). The quality assessment was followed independently by two research experts and probable discrepancy between them was resolved based on third expert opinion. Agreement was assessed using Cohen's kappa statistic. The kappa statistic for agreement on quality assessment was 0.92.

Quality of reporting was assessed based on standardized data extraction forms of GBD study.

In order to contact the corresponding authors for complementary information required, information request forms were sent along with the letter, including the goals, methods, and other required details of study. In this letter, which was signed by the principle investigator and main researcher, all of the intellectual property rights of participants were clearly specified. E-mails were sent three times with an interval of two weeks to the active e-mail of corresponding authors. Not receiving a reply was considered as non-responsive.

To assess the quality and availability of papers' data, all the information was analyzed and the results were organized as main problems and corresponding suggestions, in the four mentioned domains.

Results

We retrieved 3,253 records; of these 1,875 were from international and 1,378 from national databases. After refining steps, 129 (3.97%) papers remained related to our study domain.

Through these stages, the main reasons for exclusion of papers were selection of vague irrelevant topics (about 56% of excluded papers), and wrong selection of key words (more than 40% of excluded papers), both of which distort the process of searching and retrieval of relevant articles. Figure 1shows the flowchart for data collection and selection process.

Quality of studies

Considering the first mentioned subjects or quality of studies, only 62 (48.06%) articles that met our eligibility criteria were selected and from them, the results of 53(85.48%) papers were extracted as remaining reliable data. In other words, about 51% of relevant papers were excluded because of poor quality of studies. From 67 excluded papers, 8 papers did not mention the sample size, 23 papers were based on non random sampling methods, 5 papers did not report measures by sex, and the others did not get the minimum quality rating. The results of quality scoring of included papers are presented in Table 1.

Quality report of the results

The number of reported total population and points of data were 22,972 and 29 for boys, and 38,985 and 47 for girls, respectively. We found five studies that did not report BMI separately for boys and girls. Regarding the geographical distribution, we found 9 national, 14 provincial, and 58 district level data points. Figures 2 and 3 show the data points for BMI separately for boys and girls.

Using linear regression model, for 25 studies the mean of BMI was crossed based on prevalence of obesity and for 10 studies, the prevalence of obesity was crossed based on reported mean of BMI. Also, for 15 studies, the prevalence of overweight was crossed based on reported mean of BMI.

Most studies reported scattered point estimations of body mass index (BMI) in different sub age groups, and there was no study on trend of prevalence of obesity in Iranian pediatric population.

Most studies focused only on mean of BMI, neglecting the important values of WHR, WC, and WHtR [only 9 (6.97%) eligible papers took these auditing values into consideration]. Figure 4 shows the other measures points of data by sex, measures, and provinces during the study period.

Another important consideration is that for all measures, the reports had significant missing in reporting the confidence intervals for both anthropometric values and rates which further limited our ability to compare and analyze the results. Only 9 papers reported the Lower and Upper Level of 95% CI for mean of BMI or obesity/overweight prevalence. The missing report for categorized age groups, age range, mean of age, and SD for mean of age were respectively: 75.8%, 22.6%, 61.3%, and 64.5%.

It is noteworthy that in 16 25.8% of papers, years of the study were not reported. Table 2 shows the quality of data reporting in included papers

Responsiveness of corresponding authors

We also had serious problem in contacting the corresponding authors for complementary information required. For 28 studies, we requested the required information from their corresponding authors. From them, despite our considerations on intellectual property rights (see methods section), we received only 5 responses (Receptiveness: 17.9%).

Diversity in study settings

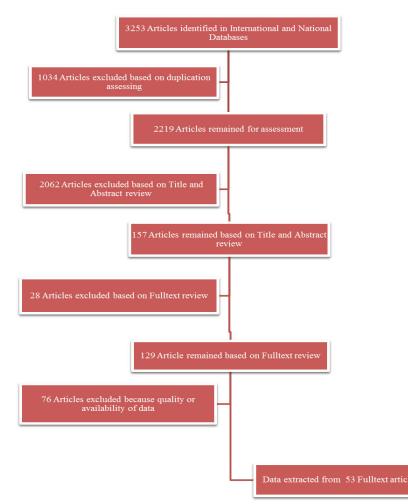The reported data had large diversity in different fields. As an

Figure 1. Flowchart for data collection and selection process.

important consideration, the reported values on prevalence of overweight and obesity and related anthropometric measures, including Body Mass Index (BMI), Waist Circumference (WC), Waist-Hip Ratio (WHR), and Waist-to-Height Ratio (WHtR), had significant variations from one study to another. Studies reported data based on different measures from various target groups of different study scopes (urban or rural residency area) with quite different age groups and sex. For instance, in age range of 6 to 11 years, we found 4 different age groupings. Considering the scope of studies, 55 (88.75%) of studies were set for urban areas, 4 (6.5%) papers did not mention their scope of study, and the remaining 3 (4.8%) papers were from rural settings. Non-homogeneous data prohibited us from conducting meta-analysis. These diversities in scattered data necessitate the use of sophisticated statistical methods that would be referred to in the following sections.

Discussion

Most of the time, using primary or secondary data is not optimal. In order to adopt a more useful approach, we should focus on selecting data that are appropriate to the research question and the available resources to the researcher. Other important determinants are time, money, and personal expertise.[9,13–14] It is also noteworthy that working with secondary data is efficient economically and provides more extensive data. At the same time, in some situations, ambiguity in details of data collection processes and aims of studied misleads the researchers.[13]

Shortcomings in accuracy of parameter estimations or even gaps in data presenting methods and skills limited our access to targeted accurate reliable data. For instance, what is reported as missing data in a paper/report, whether it was an exact report or if it was extracted from researcher error in data estimation, will be decisive in the strength of the association and analysis that is calculated in future chains of studies based on them.[15–16] The quality of presented data also has an important role in the accuracy of estimating the causes of health problems and related programs.[6,15]

Considering the above, the present paper focuses on evaluation of quality and reporting of data of papers in the field of obesity and overweight. To summarize the findings, the main problems are classified under four main domains: a) Quality of studies, b) Quality report of the results, c) Responsiveness of corresponding authors, and d) Diversity in study settings in Table 3. The table also shows the corresponding suggestions on papers' data quality and data availability.

As a feasible solution, both the authors and scientific journals must be simultaneously more responsive for the paper quality. Journals should adopt more policies for data quality and data

Table 1. Quality characteristics of included papers.

Index	Response rate for taking sample	Sample size by gender	Sampling design	Weight measured Calibration	Height measured Calibration	*Waist measured Calibration	Quality Score
First category	Over 90% — n:56 — 90.4%	over 1000 — n:36 — 58.1%	Simple random sampling/ stratified — n:8 — 12.9%	One center - using one kind of equipments — n:4 — 6.5%	One center - using one kind of equipments — n:4 — 6.5%	One center - using one kind of equipments — n:3 — 25%	19 or more — n:49 — 79%
Second category	76–90% — n:2 — 3.2%	500–1000 — n:11 — 17.7%	One-stage systematic, clustered, or stratified sampling — n:15 — 24.2%	Multi center -using one kind of equipments — n:52 — 83.9%	Multi center -using one kind of equipments — n:52 — 83.9%	Multi center -using one kind of equipments — n:1 — 8.3%	Excellent (13–18) — n:6 — 9.7%
Third category	60–75% — n:2 — 3.2%	250–500 — n:9 — 14.5%	Multi-stage clustered stratified sampling — n:39 — 62.9%	Multi center -using different kinds of equipments — n:4 — 6.5%	Multi center -using different kinds of equipments — n:4 — 6.5%	Multi center -using different kinds of equipments — n:2 — 16.7%	Good (12–6) — n:3 — 4.8%
Fourth category	Less than 60% — n:0 — 0%	Less than 250 — n:6 — 9.7%	not reported — n:0 — 0%	not reported — n:0 — 0%	not reported — n:2 — 3.1%	not reported — n:3 — 25%	Poor (≤5) — n:4 — 6.5%
Fifth category not reported	not reported — n:2 — 3.2%	not reported — n:0 — 0%					

*This column is calculated for 12 papers with waist measurment reports.

Table 2. Quality of data reporting in included papers.

Idex	Sex	Age groups	Age range	Median Age Range	Mean Age	SD Mean Age	Time of study	Scope of Study	Sample Size	Lower/ Upper Level of 95% CI	BMI Mean/ prevalence of obesity or overweight	SE Mean	*WHR/ WC/ WHtR mean	*SE of WHR/ WC/ WHtR mean	*Definition of measur	*Lower/ Upper Level of 95% CI
Reported	57 (91.9%)	47 (24.2%)	48 (77.4%)	0 (0%)	24 (38.7%)	22 (35.5%)	46 (74.2%)	58 (93.5%)	51 (82.3%)	11 (17.7%)	62 (100%)	0 (0%)	9 (100%)	9 (100%)	7 (77.8%)	9 (100%)
Not reported	5 (0.08%)	15 (75.8%)	14 (22.6%)	62 (100%)	38 (61.3%)	40 (64.5%)	16 (25.8%)	4 (6.5%)	11 (17.7%)	51 (82.3%)	0 (0%)	62 (100%)	0 (0%)	0 (0%)	2 (22.2%)	0 (0%)
Total	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**62 (100%)**	**9 (100%)**	**9 (100)**	**9 (%100)**	**9 (100%)**

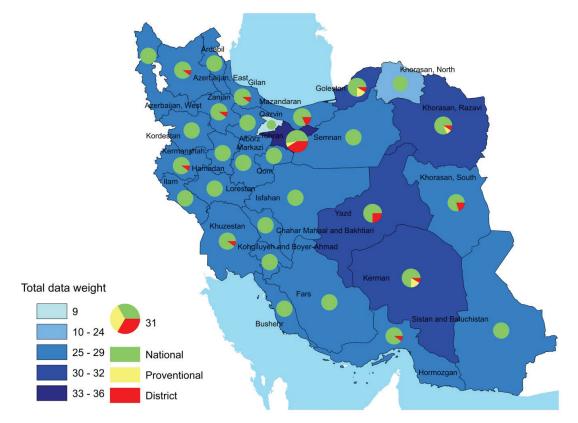*This column is calculated for 12 papers with WHR/ WC/ WHtR reports.

Figure 2. Boys' body mass index points of data by levels and provinces during the study period.
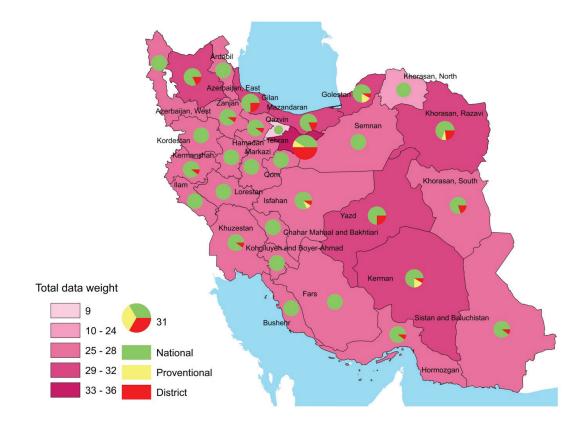

Figure 3. Girls' body mass index points of data by levels and provinces during the study period.

sharing. They must consider more obligatory standard protocols for data qualities, more exact peer-reviews, and providing instructions to authors regarding the public availability and sharing of data from submitted manuscripts. Using journal authority, non-responsive corresponding authors have to be obliged to share their data when researchers ask them. Furthermore, it is essential to develop regulatory mechanisms for journals to ensure that such policies would be consistently followed by authors and researchers.[9,14,17]

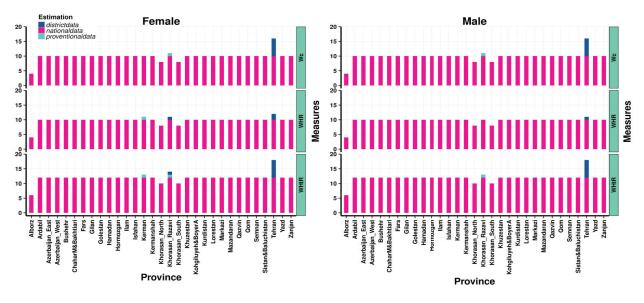Another simultaneous approach is application of advanced sta-

Figure 4. Other measures points of data by sex, measures, levels and provinces during the time period of study.

Table 3. Problems and corresponding suggestions on papers' data quality and data availability.

Problems	Suggestions
a) Quality of studies	
Validity of data	Valid methodologies in designing, conducting and documentation of studies (quality of sampling, quality of measurements, quality of analyzing of data, quality of publications, following the standard documentation principles)
Non standard age groups	Application of standard defined age groups (mean ± SD, Median, range)
b) Quality report of the results	
Ambiguity in details of the design and implementation of study	Referring to methodological approach (Clear expression of aim and scope of study, sampling methods, sample size, details of data collection, sampling time period, validity of instruments, time and duration of study, direct sampling/ Tel surveys', urban/rural area scopes, place and level of the study)
Ambiguity in the use of specialized terms	Appropriate referenced based terminology (define the terms, measures, and other technical terms)
Ambiguity in data application scopes	Referring to precision in measurements (measure, unit, measurement/self report)
	Referring to statistical analyses tools and techniques
	Referring to limitation of the study
Ambiguity in study population and representatively of samples	Description of study population (referring to inclusion and exclusion criteria, exact definition of target groups)
Retrieval of papers by other researchers	Selection the accurate illustrative title selection
	Setting the right and standard key words (based on the Mesh of Pubmed , Emtree of Scopus, and review the other related papers key words)
c) Responsiveness of corresponding authors	
More requirement information or clarification regarding the some contents of the papers	Explanation of exact Authors affiliation and available addresses (especially corresponding emails')
	Responsiveness and accountability to the papers
d) Diversity in study settings	
Diversity in presented data	Better documentation for values (more elaboration on designing, conducting and documentation of national and sub-national studies, standard analyzing of data, publishing the results, following the standard documentation principles)
Interaction with the other related data (national and sub national studies, non-published data)	

tistical strategies. Currently, available information is the only available data sources that should be used with the highest possible accuracy. To deal with mentioned shortcomings, including variations in groups studied, differences in living areas (urban/rural), problems of measures, and variations in methodological approaches during the time, and also other limitation of data presentation, we provide modern practical statistical analysis. These approaches and methods have been discussed previously.[11,18–19]

In conclusion, the present findings could provide practical information on better data presentation in papers, and more detailed design studies in this field. Because of the importance of data quality and data availability, more attention should be given to more efficient data presentation in papers, which could be also used for designing and conducting the future related studies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All co-authors had contribution in the general designing of paper, designing of systematic review, primary draft preparation, and revision. All authors have given approval to the final version of the manuscript.

Acknowledgments

References

1. Day C, Gray A, Budgell E. Health and related indicators. *SAHR*. 2011; **(1):** 119 – 248.
2. Chan M. From new estimates to better data. *Lancet*. 2012; **380(9859):** 2054.
3. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet*. 2011; **377(9765):** 557 – 567.
4. Heinrich CJ. Evidence-based policy and performance management challenges and prospects in two parallel movements. *ARPA*. 2007; **37(3):** 255 – 277.
5. Van der Sluijs JP, Petersen AC, Janssen PH, Risbey JS, Ravetz JR. Exploring the quality of evidence for complex and contested policy decisions. *ERL*. 2008; **3(2):** 024008.
6. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 2006; **367(9524):** 1747 – 1757.
7. Murray CJ, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C, et al. GBD 2010: design, definitions, and metrics. *Lancet*. 2012; **380:** 2063 – 2066.
8. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *BJM*. 2003; **14(3):** 207 – 222.
9. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP. Public availability of published research data in high-impact journals. *PloS one*. 2011; **6(9):** e24357.
10. Kelishadi R, Hovsepian S, Qorbani M, Jamshidi F, Fallah Z, Djalalinia S, et al. National and sub-national prevalence, trend, and burden of cardiometabolic risk factors in Iranian children and adolescents, 1990–2013. *Arch Iran Med*. 2014; **17(1):** 71 – 80.
11. Farzadfar F, Delavari A, Malekzadeh R, Mesdaghinia A, Jamshidi HR, Sayyari A, et al. NASBOD 2013: design, definitions, and metrics. *Arch Iran Med*. 2014; **17(1):** 7 – 15.
12. Peykari N, Sepanlou SG, Djalalinia Sh, Kasaeian A, Parsaeian M, Ahmadvand A, et al. National and sub-national prevalence, trend, and burden of metabolic risk factors (MRFs) in Iran: 1990–2013, study protocol. *Arch Iran Med*. 2014; **17(1):** 54 – 61.
13. Boslaugh S. Secondary data sources for public health: a practical guide. Cambridge University Press; 2007. Available from: URL: www.cambridge.rg/9780521870016 (Accessed Date:9/9/2014).
14. Baggerly K. Disclose all data in publications. *Nature*. 2010; **467(7314):** 401.
15. Chapman AD. Principles of data quality: GBIF; 2005. Available from: http://books.google.com (Accessed Date:9/9/2014)
16. Doménech R. Human capital in growth regressions: how much difference does data quality make? *JEEA*. 2006; **4(1):** 1-36.
17. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PloS One*. 2009; **4(9):** e7078.
18. Kasaeian A, Eshraghian MR, Rahimi Foroushani A, Niakan Kalhori SR, Mohammad K, Farzadfar F. Bayesian autoregressive multilevel modeling of burden of diseases, injuries and risk factors in Iran 1990–2013. *Arch Iran Med*. 2014; **17(1):** 22 – 27.
19. Parsaeian M, Farzadfar F, Zeraati H, Mahmoudi M, Rahimighazikalayeh G, Navidi I, et al. Application of spatio-temporal model to estimate burden of diseases, injuries and risk factors in Iran 1990–2013. *Arch Iran Med*. 2014; **17(1):** 28 – 33.