Defining and Developing Measures of Checklist for Measuring Food Store Environment: A Systematic Review

Nahid ZERAFATI-SHOAE [1,2], **Mohammad Hossein TAGHDISI** [3], **Leila AZADBAKHT** [4], **Hamid SHARIF NIA** [5], ***Naheed ARYAEIAN** [1,6]

1. Department of Nutrition, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
2. Department of Community Nutrition, School of Nutrition Sciences and Food Technology, National Nutrition and Food Technology Research Institute, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Department of Health Education and Promotion, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
4. Department of Community Nutrition, School of Nutritional Sciences and Dietetics, Tehran University of Medical Sciences, Tehran, Iran
5. Department of Medical Surgical Nursing, School of Nursing and Midwifery, Mazandaran University of Medical Sciences, Sari, Iran
6. Research Center for Environmental Health Technology, Iran University of Medical Sciences, Tehran, Iran

*Corresponding Author: Email: aryaeian.n@iums.ac.ir

Abstract

Background: Food store measurement is important for planners and policy makers to improve unhealthy stores towards healthy stores. This review aimed to outline the concepts and measures development of checklists that assess food store environment in urban communities.

Methods: The search was carried out in PubMed, Embase, Web of Science and Scopus as well as reference lists of included studies for obtaining published articles between 1990 up to the date of search (30 June 2017). Eligibility criteria attempted to capture peer-reviewed articles aimed at development and validation of checklists for assessing food stores.

Results: From 3,862 records, 24 studies were included in this review. Findings showed constructs included in the instruments were availability (n=22); price (n=22); quality of fresh foods (n=13); promotion (n=6); product placement (n=6); advertisement (n=5); shelf space (n=3); display (n=3); store features or characteristics (n=2); marketing (n=2); accessibility; nutrition information; visibility; food variety; signage. There are differences on the conceptual definition of each constructs across the checklists. Only half of studies pursued fully systematic steps for the measures development.

Conclusion: Consensus for definition of constructs of food store measurement is necessary. Besides, the development of the measures of checklists needs to be done by high-quality methods.

Keywords: Measurement; Food store environment; Instrument; Development

Introduction

Food store as a component of food environment is an important influential factor on people's food-related decisions, eating patterns, behavior and health outcomes (1). Measurement of the food stores is important for planners and policy makers to improve unhealthy stores towards healthy stores (2). More healthful food stores sell good quality healthy foods and help customers

for better choices (3). Without access to healthy foods, a nutritious diet and healthiness are unreachable (4). What consumers face within and around a food store can be variables that affect food choice and purchase. These include price, promotions, placement, freshness, nutritional information and product variety intended to be measured (5).

Checklist is the most common instrument to measure food environment (6-8). There is a list of pre-defined foods and/or factors that affect food choice and purchase (7). The development of checklists is seen with looking in publications with the sharp increase recently (1). Several reviews and systematic reviews outlined some tools developed to measure type of food environment (6-9). Some studies have provided tools published from1990 and 2007, and the others have provided from 2007 to 2015 or 1990 to 2015. These reviews showed Instruments were different in addressing type of food environment. In these reviews, checklists used to measure the food stores were presented in limited numbers. Moreover, a great proportion of tools had not been tested for psychometric properties. Another finding of these reviews showed availability was the most construct measured. In another review, Kelly and et al. presented validity or reliability some tools, published from 2000 to 2010, to measure different constructs of both consumer food environment and community food environment, but have not criticized all possible instruments (10). A systematic review overviewed the measures to assess availability and/or accessibility in the home environment, the school, stores, and restaurant. This review found conceptualizations of availability and accessibility were different across the studies (11). When constructs with different conceptual definitions are used, it is inability to compare finding across different studies that measure in-store factors and assess relationships between health outcome and in-store environments.

The finding of a systematic review study showed a wide range of food store audit tools (market basket, inventory, checklist, audit tool) which varied in food items included in (1). The findings revealed, not only instruments were heterogenic

and validity and reliability of reported tools was not clear but also it was not clear to what extent are food items of the tools relevant to health of store. Moreover, other reviews that related to topic of association between food environment elements and dietary behaviors or health outcome, did not get a definite relationship (12-14). This is the result of using many different tools and do not attend to accuracy of tools. As well as, the construct validity of tools were unknown. Construct validity is important when study is built on how the food environment elements is related to healthy behavior (8)

Thus based on gaps in the literature, we do not know 1) how much the development of checklist measures are based on a high-quality method. Because researchers who need to develop a new measurement instrument should use adequate methods to develop a measurement instrument and then evaluate it to ensure the quality of the instrument is high (15). Quality of development of measures affects quality level of psychometric properties. 2) The reliability and construct validity of the checklists have been developed 3) Do the same constructs across the checklists have the same definitions.

This paper addressed these gaps in the literature by presenting results from a systematic review on development and evaluation of checklist, the most common instrument used, for measuring consumer food store environment in urban settings. Our objective was to outline the conceptualization of constructs and the methodological quality of item development. The findings help us to have a view of future research needs.

Methods

This systematic review was planned, conducted, and reported according to the PRISMA guidelines (16).

Search strategy
A systematic review of the literature was conducted using the following electronic databases: PubMed, Embase, Web of Science and Scopus and reviewing the references of included articles. Free

text method is used to find appropriate key terms. Peer-reviewed literature were searched published from 1990 up to the date of search (30 June 2017) in English language. Search strategy terms included supermarket, food shop, convenience store, food store, food outlet, food market, food desert, grocery store, corner store, food retail, retail food, food environment, nutrition environment, consumer retail food environment, consumer food environment, retail food environment, local food environment, specialty food store, consumer food store environment, food store environment, consumer nutrition environment, food access, food accessibility, healthy food, unhealthy food, food availability, food available, food affordability, AND assess, assessment, measure, measurement, monitor, monitoring, evaluate, evaluation, indicator, instrument, questionnaire, checklist, tool, audit AND develop, development, valid, validity, validation, reliable, reliability. In addition, previous literature reviews and reference lists of included studies were manually searched.

Eligibility criteria

Eligibility criteria attempted to capture peer-reviewed articles aimed at development and validation of checklists for assessing food stores. The studies included in this review met the following criteria: 1) developing and validation of checklist 2) type of the food store that intended to be assessed such as grocery stores, supermarkets, convenience stores 3) studies in urban settings.

The authors applied the following exclusion criteria: 1) Development of interview/questionnaire, inventory, market basket 2) Development of instrument for measuring the other food environment categories such as home, schools, restaurants, workplaces, public facilities.

Quality assessment

Quality assessment was done for the following two aspects:

- Quality assessment for conceptualization of instrument was done based on method with 4-point rating scale (17). Scores were categorized according to clear conceptualization. We gave the highest rates if the concepts were named and clearly defined. The lowest rates were given when the concepts were not named or defined.

- Quality assessment for item development methods was done based on manner with 4-point rating scale (11). Rates were modified from the rating developed by previous study (17). The rate categorizations were based on how systematic the process of item development or refinement was done. The highest rate was given wherever item development was done with fully systematic process and item refinement was done with at least one method. The lowest rate was used if no systematic process was reported for the development or refinement of items. See the two sources for details. Notably, rating on a 4-point rating scale depended on the information reported by the authors.

Data extraction

Two researchers independently screened all citations by title and abstract. Full articles were examined, and data were extracted and entered into an abstraction form. Disagreement between researchers was agreed upon through discussions or by a third researcher. We designed and used structured forms to extract information on the name of instrument (if was available), year of publication, country, target population, constructs measured, construct definition (if explicitly reported in paper), and methods of item development and refinement.

Results

Identification of the checklist

Search results are summarized in the PRISMA Fig. 1. The initial search identified 3,862 publications. After de-duplication, the number of 2,624 potentially relevant articles remained. Title and abstract screening of the 2,624 citations resulted in the inclusion of 58 citations for further review. After examination of full-text articles, 20 articles were identified as being eligible and manual reference searching identified 4 additional papers (18-21). The search resulting in 24 articles finally included.
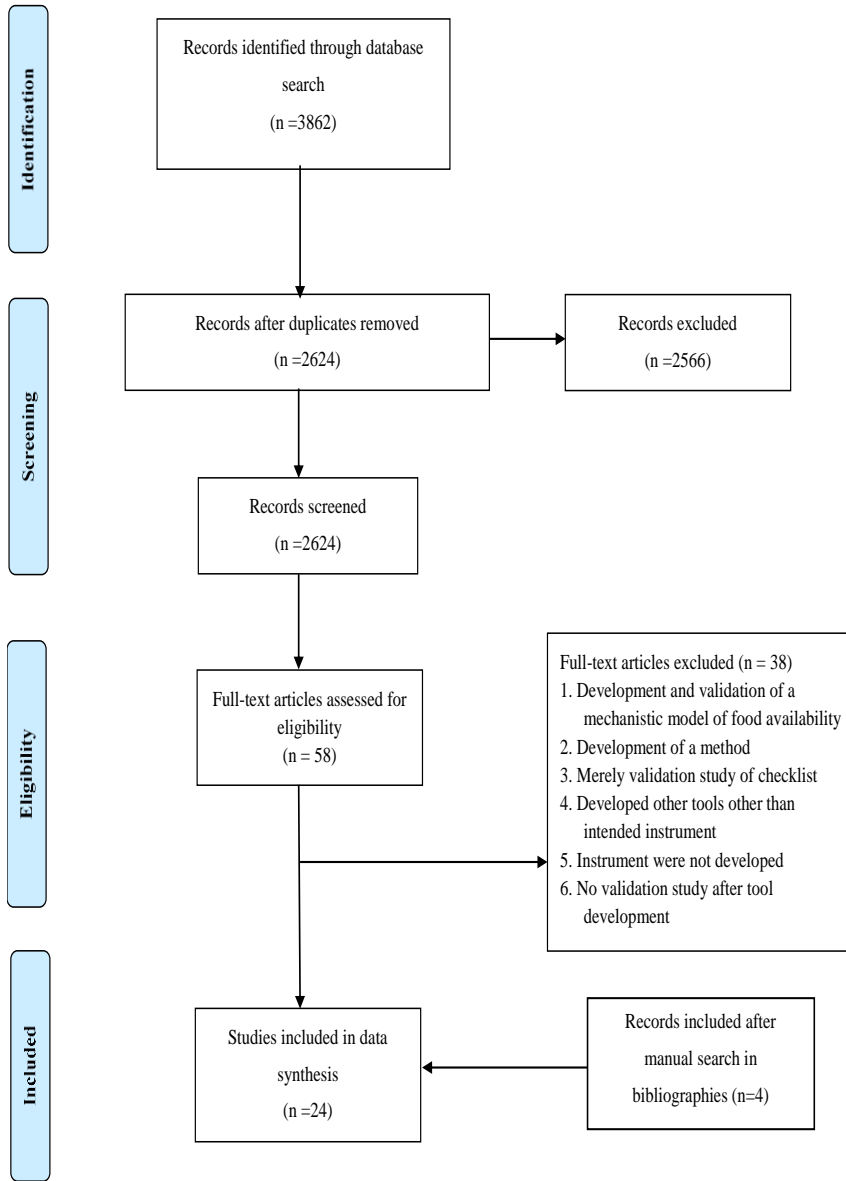
Fig. 1: PRISMA flow diagram

Characteristic and development methods of the tools

Table 1 provides a brief overview of the description, development, and refinement processes employed for each instrument to measure food store environment. Instruments have been sorted based on publication year. The development of the checklist has also been attended since 1990 year, but increased since 2007. In 1990 and also fourteen years later, 2004, only two instruments have been developed, evaluated the properties and published, but in each year of 2007, 2010, 2011, 2012, 2013, 2014, 2015, and 2016 more than one instrument have been published. Instruments were developed in 7 countries, the USA was the one with the highest number of instruments (n=17); the rest were conducted in Brazil (n=2), Canada (n=1), China (n=1), UK (n=1), Europe (n=1), Australia (n=1).

Most of the instruments were developed in specific settings or neighborhoods including predominantly low-income, racial/ethnic minority, or low socioeconomic status residents (19 of 24), but have been intended without specific disease. Two instruments designed for schools surrounded by food stores and one study included no specific condition areas. There were not feature demographic characteristics or prevalence of nutrition-related disease.

The most common respectively, constructs contained within the tools include following: foods availability(n=22); prices(n=22); quality of fresh foods (n=13); promotion (n=6); products placement(n=6); advertisements (n=5); shelf spaces (n=3); display (n=3); store features or characteristics (n=2); marketing (n=2); accessibility; nutrition information; foods variety; visibility; signage. Except for two instruments, almost in all (22 of 24), food availability and/or prices are aspects of food store measurement. Of these 22 instruments, two instruments exclusively focused on measuring food availability (22, 23), two instruments were added pricing to availability construct (namely, the tool contained two constructs: availability and pricing) (20, 21). Seven instruments are modified NEMS-S, so the constructs are the same (18, 23-28). Only NEMS-S originated from a conceptual model, community nutrition environment.

Conceptual definitions of constructs were provided in one-third of the studies (n=8), two studies generally described and seven studies only named the concepts. In a large number of instruments, availability was defined as presence or absence of food options by choosing between "Yes" or "No" scale, but the TXNEMS-WIC described as amount of shelf space, number of varieties, stocking of products, quality of fresh produce. Other instruments contain the shelf space or quality of products separate variable from availability construct (19, 29). There are different concepts of food availability. In the TXNEMS-WIC tool, the concept of accessibility was defined as visibility or display of each product, and presence of WIC labels. Display has been used as a dimension of accessibility (30) or promotion (31); it has also been defined as a construct by itself, separate from accessibility or promotion (32). Advertisement is an aspect of promotion in the ESAO-S instrument (31), but they are two separate constructs in the CX3 Food Availability and Marketing Survey instrument (33).

In table (column 4) provides details of the item development and refinement for each included study. Many of the included studies provided a clear description of how the items were chosen or developed and followed a systematic step in item development and refinement. Almost half of the studies, 14(58%), received the maximum received a score of 2, and one study receives a score of 1, indicating no description regarding how items were developed or refined.

Methods have been used in the development of items included using or building on available instruments (n=13) (19, 22-28, 30, 31, 34-36), literature review (n=14) (22, 24, 25, 28, 29, 31, 34-41), expert opinion (n=9) (21, 23, 27-29, 31, 34, 38, 41) and exploring from GPS (40). In 18 (75%) of instruments items refinement was done; however, in six other instruments were not done. Only one instrument originated from using theory to guide development of instrument. Several studies combined two or more of these methods.

Table 1: Description and psychometric properties [a] of checklists for measuring food store environment

Instrument place, Year	Target population	Constructs assessed (conceptual definitions)/ Concept score[b]	Development methods/ Method score[c]		Reliability
			Item development	*Item refinement*	
Unnamed, USA, 1990 (32)	Not mentioned	Promotion (health education activities usually printed material for example shelf labeling, posters), display (proportion of shelf	No reported score= 1	Not reported	-[d]

		space)/ score=4			
Unnamed, USA, 2004 (21)	Racial/ethnic minority neighborhood	Availability, price/ score=2	A nutrition committee selected appropriate food to be included in the tool based on recommendation, culturally acceptable and field work/ score= 4	Field work	Inter-rater reliability: κ= 0.94-1.00[f]
NEMS-S, USA, 2007(29)	People who live in neighborhoods with High/ low in SES	Availability of food items (present/absent of some food items) , Quality of fresh fruits and vegetables (acceptable/unacceptable freshness rating), cost (price per pound or item)/ score=4	Review of literature, Review of existing tools, expert consultation, Use of theory/ score= 4	pretesting proposed tool	Test-retest reliability for all food items: 0.73-1.00 Inter-rater reliability for all food items: κ=0.84-1.00
Unnamed, Australia, 2007 (20)	different in socioeconomic neighborhood	Availability, price/ score=2	Food were chosen based on guide to healthy eating, purchasing behavior and minimizing risk for diet-related diseases/ score= 2	Not reported	Inter-rater reliability: κ=0.74±0.03
Unnamed, USA, 2007, (19)	Not mentioned	Availability of alcohol and selected foods (presence/absence), advertising of alcohol (extent of ads inside the store/ number of ads on exterior of store), price of alcohol, placement of alcohol, length shelf space for alcohol and F&V (feet)/ score=4	Use of existing instrument/ score= 3	pilot- testing	Inter-rater reliability: κ=0.87 ICC=0.83
TxNEA-S, USA, 2010, (27)	Low-income and high-income neighborhoods	The same as NEMS-S/ score= -	The list of previously validated instrument (NEMS-S) was modified and expanded by team of experts opinion and use of recommendations in dietary guidelines/ score=4	Expert opinion	Test-retest: % Agreement=%92±6 Inter-rater reliability: % Agreement=%95±6
Unnamed, USA, 2010 (36)	Racially/ethnically and socioeconomically diverse community areas	Availability food options(presence/absence), prices some food (per pound/per item), quality of fresh fruit and vegetables (external appearance e.g. color, texture, form, damage)/ score= 4	Survey items were based on existing instrument, dietary recommendation, food items commonly consumed in the USA, food preferences of racial/ethnic populations, field testing was conducted/ score=3	Pretesting	Test-retest: Availability was consistent Inter-rater reliability: Nearly all food items had % Agreement=%87.5
FEAD-N, USA, 2011 (35)	Low-income and racial/ethnic minority neighborhoods	NEMS-S constructs Added physical and social store features/ score= 3	Use of several existing instruments, food items added based on culturally specific foods, store physical and social features were identified based on prior studies, previous tools, informal observation at stores, interviews with community residents, focus group/ score=3	Not mentioned	Inter-rater reliability: Almost more than 75% of items had κ=0.80-1.00
FROST, USA, 2011 (41)	a racially diverse areas with predominantly low socioeconomic status and moderate-income	Availability of food items, placement of fresh produce, prices and sizes of selected food, stores' physical characteristics/ score= 3	Based on researcher experience of previous studies, review of existing tools, use of nutritionally important foods recommendations, identifying food items according to local food culture/preference, review of available tools, and advisory board members / score=4	the first draft of tool was pilot tested and revised to increase operational efficiencies	Inter-rater reliability: The most of items had κ=0.80-1.00

CX³ Food Availability and Marketing Survey, USA, 2011 (33)	Low-income areas	Price, availability of foods, quality of fruit & vegetable, advertising, marketing, promotions, product placement/ score= 2	Was designed by expert opinion and field working/ score= 3	Pilot testing	Inter-rater reliability: κ=0.681-0.800
EURO-PREVOB Community Questionnaire, Europe, 2012 (38)	Areas of varying levels of affluence	Food environment: (number and types of food store, cost and availability of indicative food items, marketing in and outside the grocery stores, cost and marketing related to selected fast food items) build environment/ score= 2	The literature Review conducted to identify questionnaires aspects of obesogenic environment, the first draft was refined at an expert meeting/ score= 4	The instrument was pilot tested	Inter-rater reliability: ICC=0.95-0.98
GroPromo Audit Tool, USA, 2012 (37)	High/low income, ethnic group	Placement and promotional prominence of healthy and less-healthy food items/ score= 2	Reviewing literature and selecting Food items based on nutritional value and associations with childhood obesity/ score= 2	Not reported	-
BTG-FSOF, USA, 2013 (34)	racially mixed area (withe and non-white residents)	Food availability, price, advertisements, store characteristics, product placement/promotion/ score= 2	Use of several existing instruments, review of previous surveys on dietary habits, consulting with experts/ score= 4	Pre-testing the initial draft for modification	Inter-rater reliability: κ=0.84 ICC=0.90
Unnamed, Brazil, 2013 (28)	Three socioeconomic levels areas	The same as NEMS-S/ score= -	The original NEMS-S was adapted across a series of meetings of researchers, identifying food items according to commonly eaten in Brazil, recommendations and degree of industrial processing of food/ score= 4	pretesting the tool	Inter-rater reliability: κ=0.77e ICC=0.98e Test-retest reliability: ICC=0.98e
NEMS-CS, USA, 2013 (18)	Low-income areas	The same as NEMS-S/ score= -	Items were expanded from previously validated tool (NEMS-S) / score= 3	Not reported	Inter-rater reliability: κ=0.79-1.00 Test-retest reliability: κ=0.37-1.00
Unnamed, China, 2014 (24)	Ethnic minority groups	The same as NEMS-S/ score= -	Use of validated instrument (NEMS-S) to provide ideas for conceptualizing the tool, reviewing the literature to determine appropriate survey items/ score= 4	pilot testing	Inter-rater reliability: κ=0.5e
SNACZ Food Store Checklist, USA, 2014 (22)	Students in elementary and middle schools located around the food stores	Availability of healthier alternatives to the energy-dense snacks and beverages/ score= 2	items were identified by reviewing literature and tools / score= 3	pretesting in stores	Inter-rater reliability: 73% of items had κ=0.61-1.00
Unnamed, UK, 2014 (39)	different in socioeconomic neighborhood	Variety (number of different choice), price (pound per portion), quality(based on quality indicator), promotions, shelf placement, store placement, nutrition information, single fruit sale (single sale was possible) / score= 4	Food items were selected based on frequently consumed in England and represent the recommendations and contribute to nutrition-related chronic diseases/ score= 2	Not reported	Inter-rater reliability: κ=≥ 0.85
ESAO-S, Brazil, 2015 (31)	different in socioeconomic levels areas	Availability of selected foods, variety (number of different brands available for purchase), quality (unacceptable if 75% of the products was bruised, old looking, overripe, or spotted), pricing (price per kilogram/per	Building on existing tools and literatures and inputs from a panel of experts in food environment from other country / score= 4	The initial draft was modified based on the pretesting	Inter-rater reliability: most items had κ=≥ 0.70 Test-retest reliability: most items had

					κ=≥ 0.70
		unit), promotion (number of different advertisements and signs: nutrition information, displays) / score= 4			
Unnamed, USA, 2015 (26)	Not mentioned	the same as NEMS-S (Short form of NEMS-S) / score= -	Validated tool (NEMS-S) / score= 4	Items reduced by data mining techniques	-
The Outdoor MEDIA DOT, USA, 2015 (40)	Students in middle and high schools located around the food stores	Food and beverage advertising (any sign promoting food or beverages) / score= 4	Categories, types of advertisement were determined based on a review of the literature or were created by study/ score= 4	Field study	-
SCAT, USA, 2016 (23)	Low-income, high-minority communities	availability (the same as NEMS-CS, Short form of NEMS-S) / score= -	NEMS-CS survey items that related to availability of food items were selected but items related to price and quality were excluded, additional items were generated based on input from community partners and expert panel / score= 4	items redact after analytic approaches	-
ToNEMS-S, Canada, 2016 (25)	The neighborhood was composed of low-income residents (ethno-cultural diverse groups and immigrants)	the same as NEMS-S/ score= -	Items were added to validated tool (NEMS-S) based on healthy eating recommendations, commonly food consumption by population under study, field observation and key informant interviews/ score= 4	Field-testing	Inter-rater reliability for availability: κ=0.91 Inter-rater reliability for variety: ICC=0.806-0.995
TXNEMS-WIC, USA, 2016 (30)	Across the state of Texas	Availability (amount of shelf space, number of varieties of F&V, stocking of products, quality of fresh produce), accessibility (visibility or display of each product, presence of WIC labels), affordability (cost of the LEB item) / score= 4	Additional foods (WIC food package) that are culturally specific to minority populations added to previous validated instrument (TxNEA-S) / score= 4	field testing improved the tool	Inter-rater reliability: %Agreement ≥ 0.95

SES socio-economic status; *F&V* fruits and vegetables; *LEB* least expensive brand item

a: psychometric properties were reported for constructs of availability or variety of food items (ICC for counts of selected food items options, ☐ for absence/presence of food items. In some studies %Agreement were reported). b: based on method of Vaughn et al. c: based on manner of Gebremariam et al. d: availability was not construct of the tool or the tool is the short form of previous validated tool that psychometric testing were done before. e: construct validity means the degree to which the measure is correlated with other constructs in the path that one would expected. f: values reported mean or range . g Result is from other study: Minaker et.al. American journal of epidemiology. 2013 Nov 20;179(4):519-28.

Discussion

This study was able to identify 24 checklists developed for measuring different aspects of food store content in urban settings and presented the critique on checklists to give ideas for the creation of high-quality instruments.

Most of these instruments were developed and performed in low-income and minority neighborhoods; however, researchers in this review believe that to develop evaluative instruments for measuring in-store to improve rate of diet-related diseases prevalence, demographic characteristics of target population surrounded by stores better would be noted.

The availability and price were the most constructs that have been used. Although some instruments included other constructs (e.g. marketing, placement and promotion, which are aspects of food, store environment that affect food-purchasing decisions), none of them comprehensively included all constructs.

Few studies clearly defined the concepts. Different conceptual definitions of availability, accessibility, display and some other constructs were presented in studies. Most studies refer the avail-

ability to the physical presence of food, but there is a study that availability is multidimensionality and operationalized using four parts: the variety of items to assess the diversity, amount of shelf space dedicated to product, stocking of products, quality of produce. Other studies isolate availability from food quality or shelf space. We found that availability is not conceptualized in the same manner. There is also variation in the concept definition of accessibility. In the literature, accessibility is defined as the location of food supply and the ease of getting to that location (42). However, in TXNEMS-WIC instrument, it covers different aspects including visibility or display of each product and presence of WIC labels. There is a need to uniform its definition to compare the studies. Visibility position as an independent or subdomain of construct is unknown. Our findings confirmed results of previous systematic review that aimed to assess conceptualization of availability and accessibility of food among youth (11). As this study pointed out, the heterogeneity in the definition of the same constructs leads to inconsistent and conflicting results in studies on relationship between food availability and accessibility with health behavior. Therefore, global concepts of measures allow for comparison between countries and studies. In our study, all of the studies, except NEMS-S, decision on what constructs need to be included in the measurement instruments were based on review of the literature, however, for this purpose, applying a conceptual model is a better manner. Despite significant conceptual limitations in researches about the environment and health, consensus experts' opinions can provide further clarity to the conceptualization of constructs.

Ideally, instrument development and refinement of the items bank are multi-staged processes and iterative, in a continuous process of evaluation and adaptation to enhance the quality of checklist and reach to high valid instrument (17). Reviewing the literature is a basic methodological principles (15), while almost half of instruments reported review literature as part of their process. Moreover, half of instruments studies reported pulling items from existing instruments, but it is unknown how systematically existing instruments were reviewed before selecting which instruments and items to use for the new measurement. Examining similar instruments in the literature might help to provide relevant items pool. Consulting with experts help to make a selection the appropriate items, but no clearly were reported how experts judge the relevance and the comprehensiveness of the items. Once an initial item pool is created, it is also important to evaluate and refine that item pool. In the last steps, pilot-testing, final selection and evaluation of items take place. Half of studies provide a useful example of a thorough and iterative process combining multiple strategies to generate and refine an item pool. For example researchers for developing the Food Retail Outlet Survey Tool (FROST) (41), to create an initial item pool drew items from exciting instruments, previous studies, and input from advisory board. The original items pool was piloted within target stores. Studies vary widely in terms of items development rigor, scientific ways require tolerance, maybe takes years.

This study had a few limitations. We cast some databases to gather instruments for this review. However, "gray" literature or professional networks that develop such instruments have been overlooked. So few instruments may have been missed. Similarly, this review is limited to articles written in English. While maybe, some instruments were developed in other languages.

Conclusion

This review offers critiques of the checklists used in the Measuring Food Store Environment to provide ideas for creating high-quality tools. The first is that a general and global consensus is needed to create a clear conceptual model of what structures should be included and how those structures should be defined. In addition, the existing tools haven't sufficient quality, so it is recommended that new tools be designed for their improvement. Therefore, the present study offers solutions for designing the appropriate tool and process for its production and increasing the quality of the required checklist. In the end,

developed goal-based tools to prevent NCDs are proposed for future work.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

We would like to thank Farzaneh Asgari Taee for co-working in searching, data extraction and quality appraisal.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Gustafson A, Hankins S, Jilcott S (2012). Measures of the consumer food store environment: a systematic review of the evidence 2000-2011. *J Community Health*, 37(4):897-911.
2. Treuhaft S, Karpyn A (2010). The grocery gap: Who has access to healthy food and why it matters. Oakland, CA and Philadelphia, PA, PolicyLink and The Food Trust, pp. 5-24.
3. Hawkes C (2008). Dietary implications of supermarket development: a global perspective. *Dev Policy Rev*, 26(6):657-92.
4. Bell J, Mora G, Hagan E, et al (2013). Access to healthy food and why it matters: A review of the research. PolicyLink, pp. 4-20.
5. Glanz K, Sallis JF, Saelens BE, Frank LD (2005). Healthy nutrition environments: concepts and measures. Am J Health Promot, 19(5):330-3.
6. Ohri-Vachaspati P, Leviton LC (2010). Measuring food environments: a guide to available instruments. *Am J Health Promot*, 24(6):410-26.
7. McKinnon RA, Reedy J, Morrissette MA, et al (2009). Measures of the food environment: a compilation of the literature, 1990-2007. *Am J Prev Med*, 36(4 Suppl):S124-33.
8. Lytle LA, Sokol RL (2017). Measures of the food environment: A systematic review of the field, 2007-2015. *Health Place,* 44:18-34.
9. Glanz K, Johnson L, Yaroch AL, et al (2016). Measures of retail food store environments and sales: review and implications for healthy eating initiatives. *J Nutr Educ Behav*, 48(4):280-8.e1.
10. Kelly B, Flood VM, Yeatman H (2011). Measuring local food environments: an overview of available methods and measures. *Health Place*, 17(6):1284-93.
11. Gebremariam MK, Vaqué-Crusellas C, Andersen LF, et al (2017). Measurement of availability and accessibility of food among youth: a systematic review of methodological studies. *Int J Behav Nutr Phys Act,* 14(1):22.
12. Gamba RJ, Schuchter J, Rutt C, Seto EY (2015). Measuring the food environment and its effects on obesity in the United States: a systematic review of methods and results. *J Community Health*,40(3):464-75.
13. Caspi CE, Sorensen G, Subramanian S, Kawachi I (2012). The local food environment and diet: a systematic review. *Health Place*, 18(5):1172-87.
14. Engler-Stringer R, Le H, Gerrard A, et al (2014). The community and consumer food environment and children's diet: a systematic review. *BMC Public Health*, 14:522.
15. de Vet HCW, Terwee CB, Mokkink LB, Knol DL (2011). Measurement in Medicine: A Practical Guide. Cambridge, Cambridge University Press.
16. Moher D, Liberati A, Tetzlaff J, et al (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*, 8(5):336-41.
17. Vaughn AE, Tabak RG, Bryant MJ, Ward DS (2013). Measuring parent food practices: a systematic review of existing measures and examination of instruments. *Int J Behav Nutr Phys Act*, 10:61.
18. Cavanaugh E, Mallya G, Brensinger C, et al (2013). Nutrition environments in corner stores in Philadelphia. *Prev Med*, 56(2):149-51.

19. Cohen DA, Schoeff D, Farley TA, et al (2007). Reliability of a store observation tool in measuring availability of alcohol and selected foods. *J Urban Health*, 84(6):807-13.

20. Giskes K, Van Lenthe FJ, Brug J, et al (2007). Socioeconomic inequalities in food purchasing: the contribution of respondent-perceived and actual (objectively measured) price and availability of foods. *Prev Med*, 45(1):41-8.

21. Horowitz CR, Colson KA, Hebert PL, Lancaster K (2004). Barriers to buying healthy foods for people with diabetes: evidence of environmental disparities. *Am J Public Health*, 94(9):1549-54.

22. Izumi BT, Findholt NE, Pickus HA, et al (2014). Inter-rater reliability of a food store checklist to assess availability of healthier alternatives to the energy-dense snacks and beverages commonly consumed by children. *Child Obes*,10(3):266-71.

23. DeWeese RS, Todd M, Karpyn A, et al (2018). Short-Form Audit Instrument for Assessing Corner Store Healthfulness. *Am J Health Promot*, 32(1): 224-232.

24. Hua J, Seto E, Li Y, Wang MC (2014). Development and evaluation of a food environment survey in three urban environments of Kunming, China. *BMC Public Health*, 14:235.

25. Lo BK, Minaker LM, Mah CL, Cook B (2016). Development and Testing of the Toronto Nutrition Environment Measures Survey–Store (ToNEMS-S). *J Nutr Educ Behav*, 48(10):723-729.e1.

26. Partington SN, Menzies TJ, Colburn TA, et al (2015). Reduced-item food audits based on the nutrition environment measures surveys. *Am J Prev Med*, 49(4):e23-33.

27. Gloria CT, Steinhardt MA (2010). Texas nutrition environment assessment of retail food stores (TxNEA-S): development and evaluation. *Public Health Nutr*, 13(11):1764–72.

28. Martins PA, Cremm EC, Leite FH, et al (2013). Validation of an adapted version of the nutrition environment measurement tool for stores (NEMS-S) in an urban area of Brazil. *J Nutr Educ Behav*, 45(6):785-92.

29. Glanz K, Sallis JF, Saelens BE, et al (2007). Nutrition Environment Measures Survey in Stores (NEMS-S). Development and Evaluation. *Am J Prev Med*, 32(4):282-9.

30. Lu W, McKyer ELJ, Dowdy D, et al (2016). Evaluating the Influence of the Revised Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Food Allocation Package on Healthy Food Availability, Accessibility, and Affordability in Texas. *J Acad Nutr Diet*, 116(2):292-301.

31. Duran AC, Lock K, Latorre MRDO, Jaime PC (2015). Evaluating the use of in-store measures in retail food stores and restaurants in Brazil. *Rev Saude Publica*, 49:80.

32. Cheadle A, Psaty B, Wagner E, et al (1990). Evaluating community-based nutrition programs: Assessing the reliability of a survey of grocery store product displays. *Am J Public Health*, 80(6):709-11.

33. Ghirardelli A, Quinn V, Sugerman S (2011). Reliability of a retail food store survey and development of an accompanying retail scoring system to communicate survey findings and identify vendors for healthful food and marketing initiatives. *J Nutr Educ Behav*, 43(4 Suppl 2):S104-12.

34. Rimkus L, Powell LM, Zenk SN, et al (2013). Development and reliability testing of a food store observation form. *J Nutr Educ Behav*, 45(6):540-8.

35. Izumi BT, Zenk SN, Schulz AJ, et al (2012). Inter-rater reliability of the food environment audit for diverse neighborhoods (FEAD-N). *J Urban Health*, 89(3):486-99.

36. Zenk SN, Grigsby-Toussaint DS, Curry SJ, et al (2010). Short-term Temporal Stability in Observed Retail Food Characteristics. J Nutr Educ Behav, 42(1):26-32.

37. Kerr J, Sallis JF, Bromby E, Glanz K (2012). Assessing Reliability and Validity of the GroPromo Audit Tool for Evaluation of Grocery Store Marketing and Promotional Environments. *J Nutr Educ Behav*, 44(6):597-603.

38. Pomerleau J, Knai C, Foster C, et al (2013). Measuring the food and built environments in urban centres: Reliability and validity of the EURO-PREVOB Community Questionnaire. *Public Health*, 127(3):259-67.

39. Black C, Ntani G, Inskip H, et al (2014). Measuring the healthfulness of food retail stores: Variations by store type and

Available at: http://ijph.tums.ac.ir

neighbourhood deprivation. *Int J Behav Nutr Phys Act*, 11:69.

40. Poulos NS, Pasch KE (2015). The Outdoor MEDIA DOT: The development and inter-rater reliability of a tool designed to measure food and beverage outlets and outdoor advertising. *Health Place*, 34:135-42.

41. Hosler AS, Dharssi A (2011). Reliability of a survey tool for measuring consumer nu-trition environment in urban food stores. *J Public Health Manag Pract*, 17(5):E1-8.

42. Thornton LE, Pearce JR, Kavanagh AM (2011). Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. *Int J Behav Nutr Phys Act*, 8:71.