World Development 140 (2021) 105368

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Using Google data to measure the role of Big Food and fast food in South Africa's obesity epidemic

Steffen Otterbach[a,b,c], Hamid Reza Oskorouchi[a,d,*], Michael Rogan[e], Matin Qaim[d]

[a] Institute for Health Care & Public Management, Universität Hohenheim, Germany
[b] Institute for the Study of Labor (IZA), Bonn, Germany
[c] Institute of Social and Economic Research, Rhodes University, South Africa
[d] International Food Economics and Rural Development, University of Goettingen, Germany
[e] Department of Economics and Economic History, Rhodes University, South Africa

A R T I C L E   I N F O

Article history:
Accepted 17 December 2020
Available online 9 January 2021

Keywords:
Google data
Supermarkets
Obesity
National Income Dynamics Study (NIDS)
South Africa

A B S T R A C T

Many developing countries face a rapid increase in overweight and obesity, inasmuch as the prevalence has now nearly converged to levels observed in high-income countries. Among other factors, the rise in obesity is caused by a nutrition transition involving higher affordability and consumption of heavily processed or otherwise unhealthy foods containing high amounts of added sugar, fat, and salt. This development is accompanied by the growing expansion of, and increased access to, large modern food retailers (Big Food) and fast food restaurants. Using a novel methodology, we are able to link proxies of exposure to modern food environments based on Google data with nationally representative micro-level nutrition and health data to examine the influence of Big Food and fast food on overweight and obesity. The micro-level data come from the National Income Dynamics Study (NIDS) in South Africa, a middle-income country with alarming and further rising levels of obesity. We find that proximity to Big Food retailers and fast food restaurants increases overweight and obesity significantly, even after controlling for income and other confounding factors. The results suggest that the shape of food environments needs higher policy attention to promote more healthy food choices, which is true in South Africa and beyond.

1. Introduction

Populations in developing countries do not only face a broad spectrum of health risks from undernutrition and infectious diseases, but they are increasingly also affected by many non-communicable diseases (NCDs) more commonly associated with high-income countries (Prentice, 2006). In fact, over the last 20 years the burden of NCDs has increased over-proportionally in developing countries (Popkin, Adair, & Ng, 2012), while levels of obesity have almost converged to those observed in high-income countries (Popkin, Corvalan, & Grummer-Strawn, 2020; Popkin & Slining, 2013; Swinburn et al., 2019).

Major factors associated with this rising prevalence of obesity in developing countries include urbanization, an ongoing nutrition transition,[1] the greater affordability and accessibility of highly processed or otherwise unhealthy foods, and, relatedly, the growing role and concentration of modern food retailing networks commonly referred to as "Big Food" (Hawkes, 2008; Timmer, 2009). Big Food retailers, as defined here, include modern supermarkets and hypermarkets belonging to national or international chains. Since 2006, nearly all sales growth in global Big Food has been concentrated in developing countries (Stuckler & Nestle, 2012). Similarly, the consumption of soft drinks, snacks, and other highly processed foods has grown over-proportionally in low- and middle-income countries (Stuckler, McKee, Ebrahim, & Basu, 2012). Part of this rise in the consumption of unhealthy foods is ascribed to changing food environments in general, and to the supermarket revolution observed in many parts of Africa, Asia, and Latin America in particular (Campbell, 2016; Popkin & Reardon, 2018; Qaim, 2017; Weatherspoon & Reardon, 2003). This expansion of supermarkets in developing countries has been accompanied by a proliferation of global fast food companies, as well as domestic fast food and soft

* Corresponding author at: Platz der Goettinger Sieben 5, 37073 Goettingen, Germany.

E-mail addresses: steffen.otterbach@uni-hohenheim.de (S. Otterbach), hamidreza.oskorouchi@uni-goettingen.de (H.R. Oskorouchi), m.rogan@ru.ac.za (M. Rogan), mqaim@uni-goettingen.de (M. Qaim).

[1] The term "nutrition transition" generally refers to the transition from a traditional diet rich in fiber and unprocessed or lightly processed components to a diet that contains more highly processed foods rich in sugar, fat, and salt (Popkin et al., 2012).

https://doi.org/10.1016/j.worlddev.2020.105368
0305-750X/

drink producers that imitate global brands at lower prices (Traill, 2017).

However, although the rise of large supermarket chains has been on the development agenda for over a decade (Asfaw, 2008; Hawkes, 2008; Timmer, 2009), it is only very recently that the literature has begun investigating its impact on health and nutrition outcomes in developing countries (Demmler, Klasen, Nzuma, & Qaim, 2017). A substantial share of this small but evolving literature documents that food purchased in a supermarket is related to higher processing levels, a higher body mass index (BMI), and a higher likelihood of being obese and suffering from NCDs (Asfaw, 2008; Demmler, Klasen, Nzuma, & Qaim, 2017; Khonje, Ecker, & Qaim, 2020; Kimenju, Rischke, Klasen, & Qaim, 2015; Umberger, He, Minot, & Toiba, 2015). Admittedly, a few studies also suggest that supermarket distribution systems can improve dietary diversity by facilitating people's access to certain nutritious foods, such as fruits, vegetables, and animal products (Debela, Demmler, Klasen, & Qaim, 2020; Rischke, Kimenju, Klasen, & Qaim, 2015; Tessier et al., 2008). One general drawback of almost all existing studies on the link between supermarkets and nutrition in developing countries is that they build on relatively small surveys in purposively selected urban settings.[2] While nationally representative data sets often include details on household-level food consumption, they rarely provide information on the types of retailers used for food purchases. Given the limited and somewhat contrasting empirical evidence, it is still unclear how the modernization of food environments in developing countries will influence nutrition and health. This is a major research gap, especially against the background of the rapid growth of Big Food and fast food in many low- and middle-income countries.

In this study, we propose a new approach to overcome some of the data constraints in the existing literature. In particular, we merge nationally representative, georeferenced survey data with Google-based spatial data on the location of supermarkets and fast food restaurants. We then use this combination of data to examine the link between supermarket and fast food proximity and obesity. The empirical analysis focuses on South Africa, a middle-income country that faces a profound and ongoing nutrition transition and rapidly changing food environments. Our analysis shows that – also after controlling for household sociodemographic characteristics such as income and lifestyle – proximity to modern food retailers is associated with higher BMI and various binary measures of overweight and obesity. The magnitude of this effect is relatively small, but it stays constant and statistically significant in a variety of robustness checks.

2. Modern retailers, obesity, and the rise of NCDs

2.1. Supermarkets and obesity

Much of the existing literature on supermarkets and nutrition in developing countries finds a positive association between supermarket access and unhealthy diets and obesity (Qaim, 2017). Several studies focusing on urban households suggest that purchasing food in supermarkets may increase the share of processed foods consumed at the expense of non-processed foods (Asfaw, 2008; Khonje, Ecker, & Qaim, 2020; Kimenju, Rischke, Klasen, & Qaim, 2015; Neven, Reardon, Chege, & Wang, 2006; Pingali, 2006; Rischke, Kimenju, Klasen, & Qaim, 2015; Tessier et al., 2008). One of the first studies to make this link presented evidence from Guatemala showing that access to supermarkets is associated with lower consumption of fresh fruits and vegetables and higher consumption of processed foods with added fats and sugars (Asfaw, 2008). Rischke et al. (2015) used data from three towns in Kenya and also found evidence that supermarket purchases are significantly associated with the consumption of processed foods and calories. A more recent study, also focusing on Kenya, used panel data to suggest that the supermarket expansion actually contributes to the nutrition transition, as opposed to simply reflecting existing food preferences (Demmler, Ecker, & Qaim, 2018). In particular, the study found that better supermarket access results in dietary shifts away from fresh fruits and vegetables in favor of processed foods, snacks, and animal products (Demmler et al., 2018).

In terms of the direct association between supermarkets and overweight and obesity, the study from Guatemala found that a higher level of income spent in supermarkets was associated with a higher BMI and a greater risk of being overweight or obese (Asfaw, 2008). This was also found in a recent study with data from urban Zambia (Khonje, Ecker, & Qaim, 2020). Similarly, several analyses based on the data from urban Kenya suggest that supermarket purchases are associated with an increased BMI, a higher likelihood of being overweight or obese, and an increase in the probability of being pre-diabetic or suffering from metabolic syndrome (Demmler et al., 2017; Kimenju et al., 2015). Demmler et al. (2018) used panel data models with household fixed effects to show a positive association between supermarket purchases and BMI but did not identify a significant effect on being overweight or obese.

In contrast to these findings from Guatemala, Zambia, and Kenya, two studies in different developing countries found positive or non-significant nutrition and health effects of supermarket access. A study from Tunisia, where supermarket penetration is still relatively weak and largely observed in higher-income areas, found that the use of supermarkets was associated with higher dietary quality (Tessier et al., 2008). In a more robust study from urban Indonesia, Umberger et al. (2015) found no evidence of supermarket purchases being associated with either higher adult BMI or increased rates of overweight and obesity. The authors did find some indication that supermarkets are linked to childhood obesity but this was only observed in the highest-income group.

More generally, research on the drivers of overweight and obesity in developing countries is still relatively scant. One area of interest is to what extent the epidemiology of obesity in developing countries shares the same spatial and socioeconomic characteristics as obesity in high-income countries (Popkin et al., 2012). For example, the prevalence of obesity in high-income countries is higher in rural areas and among the poor, whereas the opposite is true in developing countries (i.e., where obesity is a larger concern in urban areas and among wealthier groups). Crucially, however, poorer population segments and rural areas in developing countries are rapidly catching up in terms of rising obesity levels (NCD Risk Factor Collaboration, 2019; Popkin et al., 2012). Asfaw (2008) analysis in Guatemala provided some support for this, as it showed that the association between supermarket access and the purchase of highly processed foods was larger for households below the national poverty line.

More broadly, since the effect of supermarkets on obesity and nutrition-related NCDs depends largely on the context it is important to investigate this association in countries with different socioeconomic conditions and at different stages of supermarket penetration (Hawkes, 2008; Qaim, 2017). One reason for context specificity is that the effects depend on the initial nutrition situation of the population. Obesity-increasing effects are more likely when the average BMI is already high than in situations where undernutrition is still more widespread (Debela et al., 2020; Kimenju et al., 2015). Another reason is that supermarkets often primarily sell processed foods in the early stages of expansion in

[2] The only exception we are aware of is the study by Asfaw (2008), which uses nationally representative data from Guatemala.

a country, but then also broaden their portfolio to fresh foods at later stages of development (Kimenju et al., 2015; Mergenthaler, Weinberger, & Qaim, 2009; Neven et al., 2006; Reardon, Timmer, Barrett, & Berdegué, 2003).

To the best of our knowledge, there is no published research on the link between supermarket purchases and obesity in developing countries beyond the examples discussed above. All of the existing studies focus on countries where the supermarket revolution is still at a relatively early stage. Moreover, almost all use relatively small surveys from purposively selected urban settings, with the only exception of Asfaw (2008), who used a nationally representative survey in Guatemala. As mentioned above, most existing nationally representative surveys do not include information on where households purchased their food or on the food environment more generally. Our study in South Africa is the first that combines nationally representative survey data with spatially explicit data on food environments. Another novel contribution is that South Africa is a country where the supermarket revolution is already relatively advanced both in urban and rural areas.

2.2. Fast food and obesity

Beyond supermarkets, a parallel body of literature has focused on the role of fast food restaurants for overweight, obesity, and NCDs. Two systematic reviews provide a useful background. The first, conducted in 2011, reports that about half of the studies reviewed found a significant association between living close to a fast food outlet and obesity (Fleischhacker, Evenson, Rodriguez, & Ammerman, 2011).[3] One of the most common findings across the reviewed studies is that fast food restaurants are predominantly concentrated in low-income neighborhoods (Fleischhacker et al., 2011). Over half of the studies reviewed were conducted in 2007 and 2008, underlining that this is a relatively young but fast growing area of research. Notably, all of the studies were conducted in high-income countries, especially in the United States.

Reflecting the rapidly expanding interest in fast food access and obesity, the second systematic review was conducted in 2018. It found 'inconsistent associations between the fast food environment and rates of obesity/overweight' across 46 peer-reviewed original studies (Chennakesavalu & Gangemi, 2018: 381). While a direct link between the fast food environment and obesity was difficult to identify, the review again found that fast food was consistently associated with higher rates of obesity in areas with lower income (Chennakesavalu & Gangemi, 2018).[4] Strikingly, also in this recent systematic review, most of the original studies focused on high-income countries. The dearth of empirical studies in developing countries was noted as a particular limitation of the fast food literature (see also Fraser, Edwards, Cade, & Clarke, 2010).[5] In addition, we are not aware of any previous study that has investigated access to both supermarkets and fast food outlets as correlates of overweight and obesity, as we do here for the case of South Africa.

[3] About 60% of the studies reviewed by Fleischhacker et al. (2011) included a spatial analysis of fast food access using GIS software.

[4] This finding resonates with insights from the literature on 'food deserts', which suggests that low-income areas are more likely to have limited access to food options apart from fast food outlets and convenience stores (Kwate, 2008; Lamichhane et al., 2013).

[5] In one of the few emerging-country studies and, against the backdrop of a dramatic increase in the presence of 'western-style' fast food outlets in many East Asian contexts, Odegaard, Koh, Yuan, Gross, and Pereira (2012) found that more consumption of fast food is correlated with type 2 diabetes and coronary heart disease in Singapore.

3. The South African context

Over the last 25 years, South Africa has experienced a particularly rapid rise in the number of modern supermarkets. South Africa is not only the country with the largest share of supermarkets in food retailing on the African continent, but several South African supermarket chains have also promoted the supermarket revolution in a number of other African countries (Khonje & Qaim, 2019; Weatherspoon & Reardon, 2003). South African supermarket chains – such as Shoprite and Pick n Pay – have been described as being very similar in function and form to the global chains which dominate most of the recent supermarket literature (Campbell, 2016; Peyton, Moseley, & Battersby, 2015; Popkin et al., 2012).

In the early 2000s, supermarkets already accounted for over half of all South African food retail sales (Weatherspoon & Reardon, 2003). Growth has continued with the supermarket share increasing to 68% by 2010 (Battersby & Peyton, 2014). The supermarket sector in South Africa is highly concentrated, with four major chains accounting for over 95% of formal retail market sales (Battersby & Peyton, 2014; Weatherspoon & Reardon, 2003). While supermarkets were initially only found in South Africa's large cities, more recently they also expanded to urban townships and rural areas. This means that supermarkets in South Africa are now also accessible to many low-income consumers (D'Haese & van Huylenbroeck, 2005; Okop et al., 2019).

Concerning nutritional trends, South Africa is experiencing an unprecedented increase in the prevalence of obesity (Sartorius et al., 2017). The most recent Demographic and Health Survey suggests that 68% of adult women and 31% of men are either overweight or obese (Statistics South Africa, 2017). For women, the prevalence of obesity is about three times higher than the global average. The rate of adolescent obesity is also rapidly increasing and much higher than what is observed in most other middle-income countries (Sartorius et al., 2017). While rising obesity rates are observed in urban and rural areas of South Africa, the prevalence is higher in urban areas and the increases are particularly rapid in poor and informal urban settlements (Sartorius et al., 2017). The prevalence of NCDs has also been shown to be highest in poor urban areas (Mayosi et al., 2009). However, especially in poor communities, undernutrition persists, resulting in a dual burden of malnutrition (Kimani-Murage et al., 2010).

We are not aware of any research that directly links supermarket growth to nutrition and health outcomes in South Africa. Yet there are a few studies that provide interesting insights into the cost of healthy diets and the types of foods that supermarkets offer in different settings. One estimate suggests that a healthy diet in South Africa costs 69% more than observed average diets and that such a healthy diet is not affordable for the majority of the population (Temple & Steyn, 2011; Temple, Steyn, Fourie, & De Villiers, 2011). There is also some evidence that supermarkets located closer to low-income households tend to stock less healthy foods (Battersby & Peyton, 2014), and that healthier food options are particularly expensive in low-income areas.[6] In comparison to traditional retailers, South African supermarkets also seem to expose consumers more to highly processed, cheap and energy-dense foods, including ready-made meals, snacks, and sugar-sweetened or alcoholic beverages (Hawkes, 2008).

[6] Using price data for Cape Town from Temple and Steyn (2009), it can be shown that percentage cost differences between healthy diets and less-healthy diets are larger in poorer than in richer neighborhoods. That healthy food options – such as fresh fruits and vegetables – are often very expensive in poor regions and subject to strong seasonal price fluctuation was recently also demonstrated with data from other African countries (Bai, Naumova, & Masters, 2020).

There is also hardly any literature that links fast food to obesity in a South African context. One study suggests that high-income households are the most frequent consumers of western fast food in South Africa, while low-income households are more likely to purchase local street food (Steyn, Labadarios, & Nel, 2011). However, both western fast food and local street food are fat- and energy-dense and associated with higher consumption of soft drinks, which are all factors that likely contribute to South Africa's obesity epidemic (Ronquest-Ross, Vink, & Sigge, 2015; Steyn et al., 2011). In terms of the country's fast food market structure, research suggests that both local franchises and multinational franchise corporations share a unique coexistence with supermarkets and informal traders (Maumbe, 2012). The popularity of shopping centers that include supermarkets and fast food chains in the same location are also a widely observed characteristic of the country's retail landscape (Wingrove & Urban, 2017).

4. Materials and methods

4.1. Survey data

We use data from the fifth wave (2017) of the South African National Income Dynamics Study (NIDS) (SALDRU, 2018), a nationally representative survey that conducts face-to-face interviews with private households in all of South Africa's nine provinces. A child questionnaire is administered to children up to the age of 15 (N = 14,993), and an adult questionnaire to individuals aged 15 years and older (N = 30,110) (Brophy et al., 2018). The questionnaires cover a wide range of household and individual socioeconomic and health characteristics, including a number of anthropometric measures. The secure version of the NIDS survey also provides full geocoding for each household, which is essential information for the construction of our main explanatory variable.[7]

We restrict our analysis to adults aged 20 years and older, excluding children and adolescents below 20 years of age. The reason is that the BMI, one of our main outcome variables of interest, is a measure of nutritional status in adults above 20 years of age and cannot be interpreted in the same way for adolescents and children (WHO, 2020). There are also systematic differences in the human body's metabolism between adults, adolescents, and children, so that pooled analyses of anthropometric indicators and their determinants are problematic (WHO, n.d.). The NIDS data include observations from 20,205 adults aged 20 years and older. After excluding individuals with missing relevant variables or implausible anthropometric values (BMI > 60 kg/m^2), our study sample comprises 19,091 observations.[8]

4.2. Google data

We use Google data to identify the location of supermarkets and fast food restaurants in the entire country of South Africa. Previous studies often differentiated modern supermarkets from traditional grocery stores based on criteria such as shop size, self-service options, or the number of cash registries (Demmler et al., 2018, Khonje & Qaim, 2019, McClelland, 1962). Unfortunately, such information is not provided by Google. Instead, our definition of supermarkets is based on BusinessTech South Africa (2018) and includes all Big Food retailers belonging to any of the well-known national or international retail chains operating in South Africa (e.g., Shoprite, Checkers, Pick n Pay, Spar, Woolworth).[9] Similarly, fast food restaurants are defined here as outlets belonging to any of the well-known fast food chains, as also listed by BusinessTech South Africa (2018).

This information on modern retail and restaurant chains is used to construct the distance (proximity) of each household to the closest supermarket and fast food outlet. Our data collection procedure is based on the Google Places Application-Programming-Interface (API), which enables the search for place information within a specified area (nearby search). In a first step, we created a geocoded grid with fine resolution of 1/64° longitude-latitude, covering the national territory of South Africa using a level-0 shapefile to account for international state borders. In both the shape file and the geocoded grid we use the World Geodetic System 1984 (WGS84) as the reference system for uniform position information, which appropriately accounts for the earth's curvature. This results in a two-column matrix, containing the latitudinal and longitudinal coordinates of each grid intersection point.

In a second step, we created an R program, using the Google Places API together with a verified Google developer account and processing each pair of coordinates for the grid intersections to generate a list of supermarkets and fast food restaurants for each intersection, including their names and coordinates within a specified radius (reverse geocoding). For each grid intersection point, we sent a google places API search request to gather place information within a radius of 1.2 km. This radius is slightly more than half the diagonal of the rectangles that make up the grid. The radius value was chosen as to minimize the number of duplicates from contiguous API requests while assuring full and precise country coverage.

The queries were made in two rounds. In a first round, we refined the type of location of our search request to the Google category "*grocery_or_supermarket*" and collected a list of names and geocodes of supermarkets belonging to a well-known chain within a radius of 1.2 km from each grid intersection point. In a second round, we restricted our search results to the types of places that matched the categories "*meal_delivery*", "*meal_take_away*", and "*restaurant*", and collected names and coordinates of fast food outlets belonging to a chain.

It should be noted that the user conditions of the free Google developer account restrict the maximum number of outputs per request (i.e., per grid intersection point) to 10 items. Creating a grid of high resolution is critical for obtaining complete information. Underestimation of the actual number of supermarkets or fast food restaurants is only possible if more than 10 sites of each type are present within a 1.2 km radius from each point in our grid. Such underestimation will only occur in densely populated areas with a very high number of supermarkets and fast food restaurants. Hence, when measuring the household to supermarket/fast food distance, such an error tends to be very small (max 1.2 km) and random in individual characteristics. Given that the Google algorithm gives priority to the verified, most relevant, and regularly searched and rated businesses, the probability of including large supermarkets among each Google request outcome is higher than the probability of including small ones. In addition, small supermarkets may not always be verified and correctly classified companies, which may lead to a possible underrepresentation in the

[7] Note that the data set with the geocoding of households can only be used in the DataFirst Secure Data Lab at the University of Cape Town to maintain the anonymity of the households and individuals in the sample.

[8] For quite a few individuals, data on body weight or height were missing. This could cause sample selection bias if the missing observations were systematically associated with supermarket or fast food access. We are able to test for such bias by estimating the effect of supermarket and fast food proximity on a dummy variable that is equal to one if anthropometric measures are missing, and zero otherwise. The proximity coefficients in these regressions are not statistically significant.

[9] Most of the supermarkets in South Africa belong to established chains. Traditional grocery shops typically do not belong to a chain, but they are quite different and therefore do not fall into the category "modern supermarkets". There may be a few smaller supermarkets that do not belong to a chain and that we were unable to capture with the Google data. Possible implications are discussed below.

database. However, missing Big Food supermarkets or modern fast food chain outlets is unlikely.

In our data collection procedure, which was conducted between 14 and 22 September 2017, we sent a total of 927,146 geocoded queries to Google using the "*jsonlite*" and "curl" R libraries and R version 3.4.2. The program was run on a GNU+Linux 64bit machine with 8 GB of RAM, which sent an average number of 1.9 requests per second. After the deletion of duplicates,[10] and the exclusion of erroneously captured stores that only sell liquor, clothes, or other non-food items, our final dataset contained the names and geolocalization of 5,909 supermarkets and 4,450 fast food outlets. Their spatial distribution among district municipalities is shown in Fig. 1.

We consider both our supermarket and fast food outlet datasets to be reasonably representative for two reasons. First, the provision of Google-based store location services on the websites of all the large food retailers and fast food restaurants in South Africa implies that all have a major incentive to register their stores with Google. Second, as Table 1 shows, the number of stores per supermarket and fast food chain in our dataset is close to the aggregate numbers reported on the companies' websites (at the time of data collection). Interestingly, for some of the supermarket chains we find a larger number of stores in our database than reported on the companies' websites, probably because the reported aggregate numbers on the websites are not always up to date. This is not of any concern and rather means that our data is more complete. Only for Shoprite, Boxer, and OK, our data collection strategy resulted in a smaller number of stores than what was declared on the companies' websites.

For Boxer and OK, the aggregate number of supermarket stores reported on the company websites may be larger because these chains also have outlets that we do not include in our supermarket category, such as hardware stores (Boxer Build), liquor shops (Boxer Liquors, OK Liquors and Friendly Liquor Market), and furniture outlets (OK Power Express, OK Furniture and Home & House). These two companies also own a considerable number of minimarkets (Friendly, Boxer Express, OK Express, OK Minimart), which are underrepresented in our sample. Comparing the numbers in Table 1 only for those companies where our Google search resulted in a lower number suggests we are missing 8% of the stores declared by the companies. However, when excluding Boxer and OK (for which the total number includes liquor and furniture stores), the missing number reduces to only 3%.

For the number of restaurants owned by the largest fast food chains, we draw on BusinessTech's yearly review of developments in South Africa's fast food landscape (www.businesstech.co.za), which derives its number of outlets either directly by contacting the brand, or indirectly based on the number of franchises reported by brand holders or the store counts recorded on the various franchise websites. Table 2 lists the number of outlets reported by BusinessTech in 2017 for the twenty-five largest fast food chains, together with our corresponding number collected via the Google Places API.

4.3. Proximity to modern food retailers

Using the household geolocalization provided in the secure version of NIDS together with our georeferenced data on supermarkets and fast food restaurants, we constructed three different measures of proximity. The first measures the minimum Euclidian distance from each household dwelling to the closest supermarket or fast food restaurant.[11] The other two separately measure the distance from the household to the closest supermarket and from the household to the closest fast food outlet. We decided not to construct any measures of distance to specific supermarket types (e.g., discount and convenience stores, hypermarkets, malls), as this would more likely be associated with non-random selection issues.

In South Africa, different retail chains often cater for different types of customers. For instance, Woolworth is relatively highly priced and mainly targeted at high-income consumers, whereas Shoprite has lower prices to also attract lower-income customers. However, targeting higher- or lower-income customers is not necessarily associated with selling more or less healthy foods. so that what consumers actually buy in these supermarkets remains an individual choice. Trying to categorize supermarkets would introduce a certain level of arbitrariness. Hence, we decided not to differentiate further and estimate an average proximity effect across all supermarket formats. However, in our regression framework we control for household income, regional fixed effects, and several other variables (see details below) to account for possible factors that may simultaneously influence the types of foods available in the local context, individual food choices, and nutrition status.

4.4. Anthropometric measures

NIDS collects several anthropometric measures from all sampled adult respondents, including body weight, waist circumference, and height. We calculate the BMI by dividing weight in kilograms by the square of height in meters. Following the WHO guidelines for individuals aged 20 and older, we classify individuals with a BMI ≥ 25 kg/m^2 as overweight or obese, and individuals with a BMI ≥ 30 kg/m^2 as obese (WHO, 2020). In addition, we use the waist circumference and the sex-specific cut-off points of 94 and 80 cm for men and women, respectively. Individuals above this threshold are at increased risk of metabolic complications (WHO, n.d.). We also calculate the waist to height ratio (WtHR). Individuals with a WtHR above the common cut-off of 0.5 are characterized with abdominal obesity, which proved to be a better prognostic parameter for cardiovascular diseases than the BMI (Schneider et al., 2010).

4.5. Confounding variables

Individual nutrition status does not only depend on the food environment, but also on a number of other socioeconomic variables, which we need to control for in a regression framework. One key explanatory variable that we control for is household income, which we measure as monthly disposable income net of taxes, calculated as an aggregate of labor market income, government transfers (i.e., pensions, disability, child support, foster care, and care dependency), workers' compensation, unemployment benefits, and remittances received. We adjust this aggregate based on household structure and equalize it using the square root of household size as an equivalence scale.

Because higher levels of physical activity increase metabolism and energy expenditure (Chiolero, Faeh, Paccaud, & Cornuz, 2008), physical activity plays a central role in body weight regulation and is thus an essential inclusion in any study on overweight and obesity. To measure the degree of physical activity, we employ a set of dummy variables based on the 5-point scale used by NIDS to measure how regularly respondents exercise: (i) never, (ii) less than once a week, (iii) once a week, (iv) twice a week, and (v) three or more times a week. Furthermore, given that smoking can affect metabolic functions and potentially lower caloric intake through

[10] Duplicates are generated by overlap in the radii of queries for contiguous grid points.

[11] We opted for Euclidian distance as opposed to travel time (or actual road distance) due to data availability considerations. The accuracy of other indicators would depend on the availability and precision of road network data for each single grid point. We were not able to find a high-resolution road network shapefile for the entire country of South Africa.

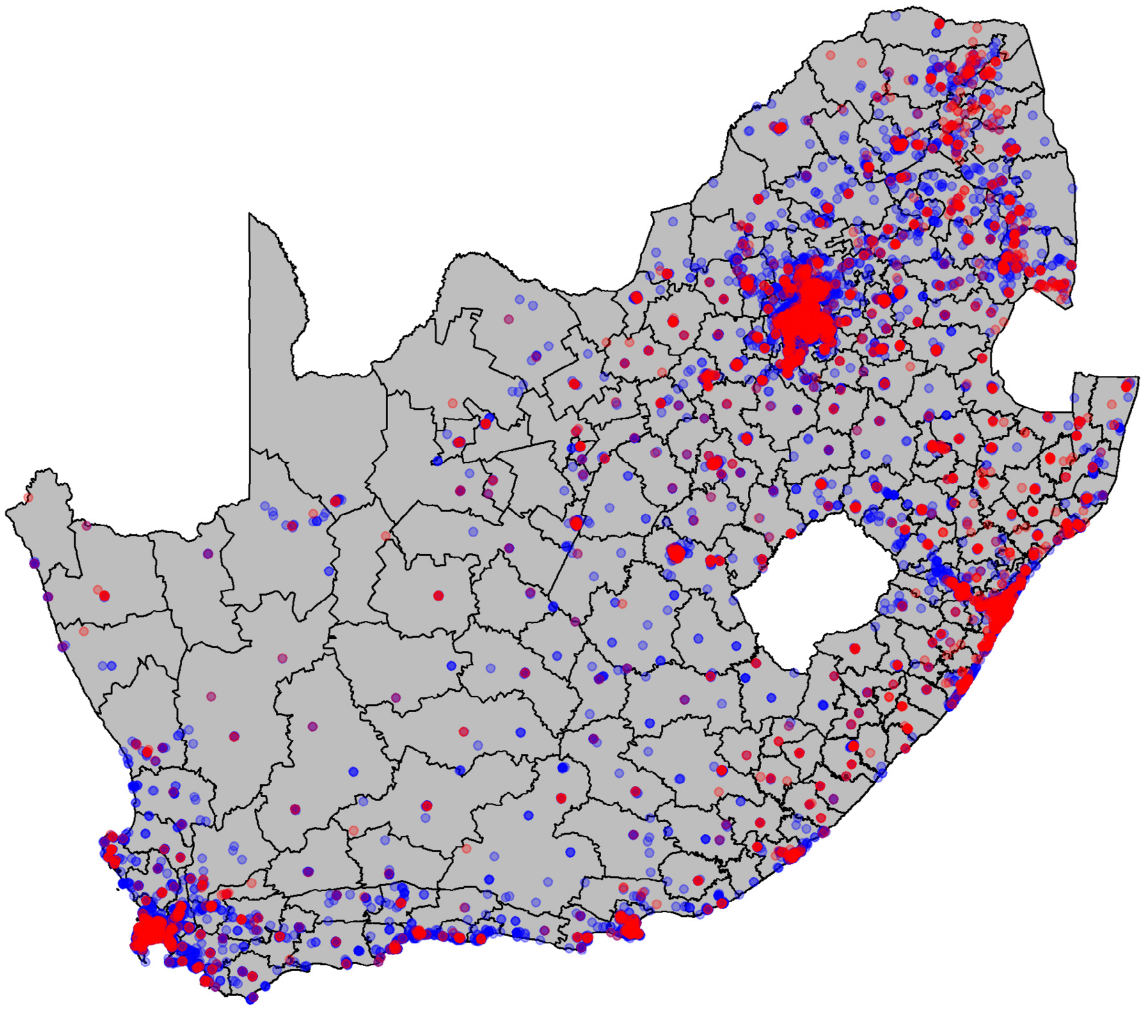

Fig. 1. Spatial distribution of supermarkets and fast food restaurants in South Africa Note: Red dots denote supermarkets and blue dots denote fast food restaurants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table1
Number of supermarkets in South Africa (2017).

Company	Declared on company website	Collected with Google data
Shoprite	767	696
Checkers	243	314
OK	214	129
Pick n Pay	417	483
Boxer	143	72
Spar	850	879
Woolworth		289
Total	2634	2862

Note: OK owns 21 wholesale stores (Megasave), which are not included in the count above. Figures regarding OK include also Sentra and Friendly supermarkets. Declared number of supermarkets are taken from: http://www.boxer.co.za/store-locator/ (Boxer); https://www.checkers.co.za/store-locator.html (Checkers); https://www.okfoods.co.za/our-stores/ (OK); https://www.picknpay-ir.co.za/store-footprint-format.php (Pick n Pay); https://www.shopriteholdings.co.za/group.html (Shoprite); https://www.spar.co.za/Store-Finder (Spar).

nicotine-induced reduction of appetite (Chiolero et al., 2008; Stojakovic, Espinosa, Farhad, & Lutfy, 2017), we include a dummy for whether or not an individual smokes.

Food consumption and dietary patterns are also directly linked to subsistence farming in the sense that households that produce their own food should be less dependent on food purchases and thus also less affected by food price fluctuations (Baiphethi & Jacobs, 2009). Subsistence farming households, which we capture through a dummy variable, may therefore be expected to consume less processed and packaged foods. As a measure of geographic dispersion, we also control for population density at the local municipality level. Finally, we include a set of controls for socioeconomic and sociodemographic factors, such as gender, age, population group, household size, and employment status.

Another important factor is the geographic classification of the household's residence. As mentioned above, in most low- and middle income countries, including South Africa, the prevalence of overweight and obesity is higher in urban than rural areas. Hence, based on 2011 census data, we control for geographic classification of residence by including a set of dummy variables for whether a respondent lives in a traditional, farming, or urban area (reference category). Traditional areas are closely related to the Black homelands of the apartheid era and are under the jurisdiction of traditional leaders (Noble & Wright, 2013). Farming areas are geographic areas in which land is allocated and used for commercial farming, while urban settings comprise continuously built-up settlements, such as cities, towns, townships, small towns, and hamlets.

4.6. Regression framework

To estimate the association between access to supermarkets and/or fast food restaurants and overweight/obesity we use regression models of the following type:

Table 2
Number of western-style fast food restaurants in South Africa (2017).

	Declared by BusinessTech	Collected with Google data
KFC	840	955
Steers	542	612
Debonairs	473	526
Wimpy	492	440
Nando's	300	273
Mc Donald's	241	259
Chicken Licken	240	148
Fishaway	213	152
Roman's Pizza	202	295
Chesa Nyama	183	115
The Fish and Chip Co	163	39
Hungry Lion[#]	130	71
Domino's	125	87
Pizza Perfect	99	73
Panarottis	80	96
Mochachos	78	48
Burger King	70	66
Barcelo's	69	57
Milky Lane	59	14
Simply Asia	56	38
Zebro's	55	21
Rocomamas	48	4
Maxis	36	6
Pizza Hut	35	50
Wakaberry	34	5
Total	4863	4450

Note: [#] The number of Hungry Lyon outlets is based on own store count using the store locator on the company website.

$$y_i = \alpha + \beta DIST_i + \gamma \boldsymbol{IC}_i + \delta \boldsymbol{HC}_i + \theta \boldsymbol{PC}_j + \varepsilon_i$$

where y_i is either a continuous (BMI) or binary (BMI $\geq$ 25, BMI $\geq$ 30, WtHR > 0.5) variable of anthropometric outcomes of individual i, while $DIST_i$, the distance to the closest supermarket or fast food restaurant, is our main explanatory variable of interest. $\boldsymbol{IC}_i$ is a vector of individual characteristics (age, gender, population group, educational level, employment status, physical activity, and smoking behavior), $\boldsymbol{HC}_i$ is a vector of household characteristics (household income, household size, subsistence farming, and geographical classification), and $\boldsymbol{PC}_j$ is a vector that controls for provincial dummies and population density at the district municipal level j in which individual i resides.

We use the OLS estimator for models with BMI as dependent variable, and logistic regression estimators for models with binary outcome variables (obesity, overweight and obesity). In all models, we take the NIDS survey design into account by post-stratification weighting of observations to assure nationally representative estimates, and by computing clustered robust standard errors at the primary sampling unit to correct for the sample's clustered nature (Chinhema et al., 2016).

As regards our main explanatory variable of interest (distance to closest supermarket or fast food restaurant), food retailers and fast food chains establish their stores in a non-random process. Their decisions are typically based on market potential and the socioeconomic characteristics of potential customers. Company decisions may also be driven by regional factors, such as population density and degree of urbanization. These demand side factors are all observable and controlled for in our models. Individual consumers cannot influence the location of supermarkets and fast food restaurants (Rischke et al., 2015). Given the high speed of changes in South Africa's food retail environments, it is also unlikely that many households chose their residence primarily based on the location of supermarkets or fast food restaurants. Hence, the distance between households and supermarkets/fast food restaurants can be considered exogenous in our regression framework.

Fig. 1 shows that many of the supermarkets in South Africa are quite close to one or more fast food restaurants. Such a spatial clustering of supermarkets and fast food outlets has also been observed in other countries (Lamichhane et al., 2013). The correlation between distance to supermarkets and distance to fast food restaurants prevents us from claiming reciprocal exogeneity of these two proximity variables. Nevertheless, when analyzing supermarket effects on nutrition and health, we need to account for potential confounding effects of fast food restaurants because both types of outlets can play a role and their effects are not necessarily identical.

We also stress again that we only consider distances to bigger-chain supermarkets and fast food restaurants, thus ignoring informal and traditional retailers. In areas where formal and informal retailers are spatially clustered (e.g., in densely populated neighborhoods), our results could possibly be imprecise. To partially address this issue, we do not only control for a wide range of household socioeconomic characteristics but also estimate the distance coefficient conditionally on local municipality population density.

In our main model specifications, we use a combined continuous variable that measures the distance of the individual's household to the closest supermarket or fast food restaurant. In additional robustness checks, we include the distance to supermarkets and fast food restaurants as separate continuous variables to see how the effects may possibly change and to get some indication of which of the two types of food retailers may have the stronger effect on nutrition status. We also carry out several other robustness checks, which are explained in more detail below.

5. Results

5.1. Main results

Table A1 in the Appendix shows sample descriptive statistics for all variables used in the regressions. A mean BMI of 27.2 kg/m^2 and an overweight/obesity prevalence of 55% among the adult population underline that overconsumption of calories is widespread in South Africa, as described above. The average distance of households to a modern, bigger-chain supermarket and a fast food chain restaurant is 7.6 and 8.7 km, respectively.

Fig. 1 graphs the distribution of supermarkets and fast food restaurants in South Africa and illustrates their spatial clustering. Every supermarket has on average 1.4 and 2.3 fast food outlets within a radius of 500 and 1,000 meters, respectively. Because of this spatial clustering, in our main regression models we combine both types of modern food outlets in one explanatory variable, estimating the effect of proximity to the closest supermarket or fast food restaurant on BMI and the probability of being overweight or obese. The estimation results are shown in Table 3. The first important insight is that the distance to the closest supermarket or fast food restaurant has a negative coefficient that is statistically significant in all models shown in columns (1) to (5) of Table 3. Each kilometer of additional distance between the household and supermarkets or fast food outlets decreases BMI by 0.014 kg/m^2 (OLS estimates) and also significantly reduces the probability of overweight and obesity (logistic regressions). In other words, closer proximity and thus better access to supermarkets and fast food restaurants increases BMI and the probability of overweight and obesity.

Because the logistic regression models employed for the binary dependent variables (overweight/obese) are nonlinear, we calculate the predicted probabilities as a function of distance to supermarkets at 10 km intervals. These results are shown in Table A2 in the Appendix. From a 56% probability of being overweight or

Table 3
Distance to supermarket or fast food outlet and obesity (OLS and logistic regressions).

	BMI (1)	Obese (dummy) (2)	Overweight (dummy) (3)	Central obesity (dummy) (4)	WtHR (dummy) (5)
Distance to closest supermarket or fast food (km)	−0.0144***	−0.00561**	−0.00604***	−0.00469**	−0.00597***
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
Equalized household income	0.0546***	0.0108*	0.0292***	0.0274***	0.0271***
	(0.018)	(0.006)	(0.009)	(0.009)	(0.010)
Household size	0.0751	0.0127	0.0187**	0.0111	0.0223**
	(0.049)	(0.012)	(0.009)	(0.010)	(0.009)
Female	4.821***	1.606***	1.357***	2.472***	1.636***
	(0.172)	(0.070)	(0.064)	(0.076)	(0.072)
Female household head	−0.0805	0.01	−0.034	−0.0241	−0.0658
	(0.175)	(0.069)	(0.061)	(0.067)	(0.066)
Partnered (ref.: no partner)	1.572***	0.562***	0.572***	0.584***	0.609***
	(0.178)	(0.066)	(0.056)	(0.065)	(0.068)
Age	0.0937***	0.0280***	0.0323***	0.0546***	0.0583***
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)
Educational level (ref.: no schooling)					
Primary school	1.662***	0.398***	0.265***	0.540***	0.338***
	(0.304)	(0.113)	(0.100)	(0.128)	(0.130)
Secondary school	2.694***	0.720***	0.843***	1.005***	0.820***
	(0.320)	(0.119)	(0.109)	(0.128)	(0.139)
Higher education	3.285***	0.900***	1.021***	1.207***	0.932***
	(0.348)	(0.127)	(0.121)	(0.136)	(0.146)
Population group (ref.: African)					
Colored	−0.127	0.118	0.240*	0.337***	0.271**
	(0.329)	(0.097)	(0.129)	(0.112)	(0.118)
Indian/Asian	−2.366***	−0.812***	−0.390***	−0.0538	0.274
	(0.450)	(0.184)	(0.144)	(0.185)	(0.236)
White	−0.773**	−0.278*	−0.126	0.214	−0.216
	(0.389)	(0.144)	(0.135)	(0.146)	(0.161)
Employed	1.014***	0.296***	0.343***	0.302***	0.301***
	(0.149)	(0.055)	(0.055)	(0.065)	(0.066)
Exercise (ref.: never)					
Less than once a week	−0.127	−0.116	0.0671	−0.0364	−0.0714
	(0.255)	(0.106)	(0.097)	(0.109)	(0.104)
Once a week	−0.154	−0.0584	−0.0789	−0.203	−0.193
	(0.316)	(0.118)	(0.112)	(0.124)	(0.123)
Twice a week	−0.509**	−0.203*	−0.177	−0.377***	−0.307***
	(0.256)	(0.107)	(0.108)	(0.114)	(0.111)
Three or more times a week	−0.759***	−0.514***	−0.266***	−0.546***	−0.330***
	(0.191)	(0.096)	(0.091)	(0.108)	(0.084)
Being a smoker (ref.: non-smoking)	−2.600***	−0.870***	−0.963***	−0.856***	−0.876***
	(0.168)	(0.095)	(0.068)	(0.080)	(0.074)
Subsistence farming (ref.: no)	−0.160	−0.0126	−0.00543	−0.100	−0.0352
	(0.221)	(0.078)	(0.072)	(0.078)	(0.083)
Population density	−0.000108	−6.86E-05	−5.50E-05	−0.000146***	−0.000107**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Geographic classification (ref.: urban)					
Traditional	−0.663***	−0.202**	−0.175**	−0.195**	−0.186**
	(0.231)	(0.085)	(0.074)	(0.086)	(0.078)
Farms	−0.817**	−0.379***	−0.238**	−0.305**	−0.0619
	(0.319)	(0.135)	(0.120)	(0.138)	(0.121)
Constant	18.29***	−3.666***	−2.694***	−4.197***	−3.246***
	(0.612)	(0.213)	(0.201)	(0.214)	(0.224)
Observations	19,091	19,091	19,091	19,151	19,123
R-squared	0.248				
F stat	105.9	41.82	54.9	74.61	67.42
F *p*-value	0.000	0.000	0.000	0.000	0.000

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clustered standard errors in parentheses (2,841 enumeration areas) | Regressions include dummy variables at the province level . WtHR, waist to height ratio above cut-off of 0.5 for abdominal obesity.

obese when living close (0–10 km) to a supermarket or fast food restaurant, the probability decreases by around 1.2 percentage points for every 10 km of additional distance. For the other obesity indicators, the marginal effects are similar.

The estimates for the other explanatory variables in Table 3 show statistically significant positive associations between household size, being female, having a partner, and household income and all measures of overweight and obesity. However, for household size the associations are only significant with respect to overweight and the waist to height ratio. Being employed and having higher levels of education show a positive association with all our nutrition status variables. For education, the positive effect may surprise, especially when also controlling for income, because better educated people are expected to be more knowledgeable about healthy nutrition and therefore less affected by overweight and obesity. However, positive associations between education and obesity were also found in other countries of Southern and Eastern Africa (Kimenju et al., 2015; Wittenberg, 2013). This result may indicate that overweight and obesity are not yet widely perceived as unhealthy conditions in the local context.

Subsistence farming has negative estimation coefficients in Table 3. This is expected, as households deriving substantial portions of food from own production will typically consume fewer highly processed foods and thus be less affected by overweight and obesity. However, the negative coefficients are small and statistically insignificant. Even when the geographical control variables are dropped, the subsistence farming coefficients remain insignificant (Table A3 in the Appendix). The insignificance may be due to the very broad definition of subsistence farming in the NIDS data. It is based on the question "have you participated in growing own food or raising livestock other than as part of paid employment during in the last 12 months?", meaning that also households with small kitchen gardens are included. In addition to this broad definition issue, other recent research suggests that subsistence agriculture contributes relatively little to food security (Rogan, 2018) and rural livelihoods in South Africa (Rogan, 2020).

Considering the coefficient estimates for the different population groups in Table 3, overweight and central obesity are more prevalent among Colored people than among Africans (the reference category). In contrast, Indians/Asians and Whites have a significantly lower BMI and a lower risk of being obese (BMI $\geq$ 30). In terms of physical exercise, we do not find significant effects of physically exercising once a week or less, whereas more frequent physical exercise has a decreasing effect on all measures of overweight and obesity. For the dummy variable of being a smoker, we find significantly negative effects in all five regressions in Table 3.

Finally, relative to urban areas, the association between living in traditional areas and in farming areas and the probability of obesity is negative. These negative coefficients in spite of controlling for income, education, subsistence farming, distance to supermarkets, and various other factors suggests that living in traditional areas and farming areas has additional facets that reduce the risk of overweigh and obesity, such as more physical labor.

5.2. Robustness checks and limitations

The main estimation results presented in the previous subsection suggest that closer proximity and thus better access to supermarkets and fast food restaurants increases people's BMI and their probability of being overweight or obese. This is plausible, given that modern supermarkets and fast food restaurants sell more highly processed and obesogenic foods than more traditional retail formats. However, there are a couple of aspects related to data quality, variable definitions, and model specifications that deserve some further attention. In this subsection, we discuss potential data and specification issues and carry out related robustness checks.

A first issue that deserves further attention is the definition of our proximity variable. In the main models above we combined proximity to supermarkets or fast food restaurants in one single continuous variable, which was convenient because of the spatial clustering of supermarkets and fast food outlets. However, in spite of this spatial clustering there is no perfect correlation, so that it is also possible to use two separate continuous variables for the distance to supermarkets and the distance to fast food restaurants. Table 4 shows alternative model estimates with these two distance measures as explanatory variables. The upper part of Table 4 refers to models where only distance to the closest supermarket is included. These estimations also yield significantly negative coefficients, meaning that living further away from a supermarket decreases the probability of overweight and obesity. The middle part of Table 4 shows results of models where only distance to the closest fast food restaurant is included. Also in these specifications, the distance coefficients remain negative but they are only significant for BMI, being overweight (BMI $\geq$ 25), and abdominal obesity (WtHR $\geq$ 0.5).

In the lower part of Table 4, we include both distance variables together in the same regressions. In these models, the supermarket distance coefficients remain negative and significant for most nutrition outcomes, while the fast food distance coefficients turn insignificant for all outcomes. Because of the spatial clustering of supermarkets and fast food restaurants, we use variance inflation factors (VIFs) to test for multicollinearity. With VIFs of 2.4 for supermarkets and 2.7 for fast food outlets we find no evidence of multicollinearity. We can therefore assume that the loss of statistical significance in the fast food distance coefficient is not driven by inflated standard errors.

The estimates in Table 4 underline that our results regarding supermarkets are robust, even though the null result for fast food restaurants in the lower part models should probably not be over-interpreted. The spatial clustering can certainly lead to confounding effects. The advantage of our Google data collection technique is that supermarket and fast food locations can both be captured to control for spatial clustering. This may also be an advantage in other countries and situations, as different types of retailers often co-evolve with changing food environments.

A second issue that deserves some more reflection is potential measurement error with respect to the supermarket and fast food distance variable. As explained above, with our Google search strategy we capture supermarkets and fast food restaurants belonging to larger chains quite well, whereas we do not capture smaller outlets that may still qualify as modern supermarkets or fast food restaurants but do not belong to any of the more widely observed chains. If such non-chain outlets exist, the resulting measurement error could potentially lead to bias. However, we argue that – if existent - the resulting bias will be small. First, according to our own observations, the number of retail outlets that would qualify as modern supermarkets or fast food restaurants and do not belong to a larger chain seems to be small. Second, especially in urban areas, where many chain supermarkets and fast food restaurants exist, the few non-chain counterparts typically cluster in the same locations, so that the distance measures that we use in our regressions would not change much, even if we could include non-chain modern outlets. This is a bit different in more remote rural locations (traditional and farming areas), where the density of modern chain outlets is lower. In such locations, not capturing non-chain outlets may lead to overestimated distance variables. If the distance to the closest supermarket or fast food restaurant is overestimated, the proximity effect on overweight and obesity would be underestimated, meaning that the true effect would be larger in absolute terms. Hence, we can conclude that measurement error – if relevant – would not change our general finding that closer proximity to modern supermarkets and fast food restaurants contributes to rising overweight and obesity.

A third issue concerns the question whether we properly control for household living standard with our income measure. While income is generally considered a good proxy of living standard, the current income of households may fluctuate, so that only looking at income in one particular year, as we do, may not be a perfect indicator. Fortunately, NIDS has a longitudinal structure, so that for most households in the sample we also have income data from previous NIDS waves, which we use to construct an average (deflated) household income variable over time for each individual (3,110 observations that entered the panel survey only in the 2017 wave are excluded from this analysis). In a robustness check, we rerun the regression models by controlling for both current income and average income over time. The results are shown in Table A4 in the Appendix. As expected, average income is positively associated with BMI and the probability of being overweight and obese. However, the effect of distance to the closest supermarket or fast food restaurant remains virtually unchanged in comparison to the main model results in Table 3. Another advantage of

Table 4
Distance to supermarket and fast food outlet and obesity, separate supermarket and fast food variables (OLS and logistic regressions).

	BMI (1)	Obese (dummy) (2)	Overweight (dummy) (3)	Central obesity (dummy) (4)	WtHR (dummy) (5)
Supermarkets only					
Distance to closest supermarket	−0.0138*** (0.005)	−0.00571*** (0.002)	−0.00558*** (0.002)	−0.00402* (0.002)	−0.00482** (0.002)
Fast food only					
Distance to closest fast food	−0.00864* (0.005)	−0.0026 (0.002)	−0.00428*** (0.002)	−0.00222 (0.002)	−0.00372** (0.002)
Supermarkets and fast food					
Distance to closest supermarket	−0.0129** (0.006)	−0.00670*** (0.002)	−0.00382* (0.002)	−0.00425 (0.003)	−0.0032 (0.003)
Distance to closest fast food	−0.00102 (0.006)	0.00118 (0.002)	−0.00205 (0.002)	0.000273 (0.002)	−0.00183 (0.002)
Observations	19,091	19,091	19,091	19,151	19,123

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clusterd standard errors in parentheses | Separate regressions estimated with only supermarket distance, only fast food distance, and both variables together included. All regressions also include the same set of control variables as the models in Table 3. WtHR, waist to height ratio above cut-off of 0.5 for abdominal obesity.

this robustness check is that the additional inclusion of average income reduces the possibility of time-invariant unobserved heterogeneity biasing the results. This further adds to the credibility of our main findings.

A final point worth discussing and testing is the potential heterogeneity of the supermarket and fast food restaurant effects by population groups. South Africa is a country with several ethnic groups that differ in terms of average living standards, lifestyles, and food consumption behavior. In order to test whether the effects vary by ethnic group, we rerun our models only for Africans and Coloreds, that is, excluding observations from Whites and Indian/Asian individuals. This stratification is common in the empirical literature on South Africa, because Whites and Indians/Asians are better off on average. The estimation results are shown in Table 5. Also for this smaller subsample of Africans and Coloreds, the supermarket and fast food effects remain very similar to the full sample effects in Table 3.

Overall, our main finding of a significant relationship between distance to modern retailers and individual nutrition status is robust to a variety of model specifications and the inclusion or exclusion of additional explanatory variables. Of course, omitted variable bias could still play some role, so that further research is warranted. In any case, our results support recent other studies (Demmler, Ecker, & Qaim, 2018; Khonje, Ecker, & Qaim, 2020; Kimenju, Rischke, Klasen, & Qaim, 2015) suggesting that modern retailers contribute to the obesity epidemic among adult populations in Africa.

6. Discussion and conclusion

As the rapid nutrition transition and the rising obesity epidemic in high-, middle-, and low-income countries require urgent attention, a growing body of literature is investigating the relation between changing food environments and people's health and nutrition outcomes. The GBD 2015 Obesity Collaborators (2017) study has identified changing food systems and food environments as the major drivers of the obesity epidemic worldwide, but has also emphasized major research gaps, especially in low- and middle-income countries. The number of studies analyzing the effects of changing food environments on obesity in low- and middle-income countries is small, and the few existing studies yield somewhat mixed findings on the link between modern food retailers, nutrition, and health. In this study, we aimed to investigate these links in South Africa, a middle-income country with a high and further growing prevalence of overweight and obesity and rapid changes in local food environments characterized by a dominant role of Big Food supermarkets and fast food chains.

For the analysis, we proposed a novel methodology that enables the merging of household survey information with publicly available geospatial (Google) data on modern food retailer locations. In principle, this methodology can be used in any country where georeferenced household data and Google data are available. Applying this methodology to South Africa, we demonstrated that proximity to bigger-chain supermarkets and fast food restaurants is significantly associated with overweight and obesity.

Our estimation results suggest that a 10 km decrease in the distance to the closest supermarket or fast food restaurant increases adult BMI by 0.14 kg/m^2 and the probability of overweight and obesity by 1.2 percentage points. These effects remained robust in a variety of model specifications. The magnitude of the effects is relatively small, but it should be stressed that our models control for household socioeconomic status, physical activity levels, ethnicity, and various other location factors that all influence people's nutrition status as well. Hence, we conclude that Big Food and fast food are not necessarily the main drivers of the obesity epidemic in South Africa, but they likely contribute to the problem.

Our results from South Africa are consistent with earlier research in other countries of Africa showing that modern retailers contribute to higher consumption of processed and energy-dense foods (Demmler, Ecker, & Qaim, 2018; Hawkes, 2008; Khonje, Ecker, & Qaim, 2020; Khonje & Qaim, 2019; Rischke, Kimenju, Klasen, & Qaim, 2015). However, unlike our study, these previous studies did not use nationally representative data but focused on specific urban environments only. While overweight and obesity in Africa is still more prevalent in urban than rural areas, our data from South Africa suggest that the observed effects of modern retailers are not confined to urban environments. Modern supermarkets and fast food restaurants have different food offers and marketing strategies than traditional retailers, which seems to influence people such that they consume more calories and more highly-processed foods. This has important policy implications. While banning modern retailers would be inappropriate, regulating food environments in such a way that consumers find it easier to make more healthy food choices should be possible. Studies from different countries demonstrate that interventions such as taxes, subsidies, changes of in-store placements of healthy and less healthy foods, and regulation of information and advertisement campaigns can influence consumer food choices significantly (Adam & Jensen, 2016).

Beyond the methodological innovation in terms of using Google data, another novelty of our study is to examine access to supermarkets and fast food restaurants simultaneously. To the best of our knowledge there is no previous work – neither in high-income nor developing countries – that has jointly examined both components of modern retail environments and their association with overweight and obesity. This is a significant contribution given the recent recognition of the importance of investigating fast food outlets 'as a part of the total food environment' (Chennakesavalu & Gangemi, 2018: 379). To examine either supermarkets or fast food in isolation means ignoring the spatial links

Table 5
Distance to supermarket or fast food outlet and obesity, Africans and Coloreds only.

	BMI (1)	Obese (2)	Overweight (3)	Central obesity (4)	WtHR (5)
Distance to closest supermarket or fast food	−0.0132** (0.005)	−0.00538** (0.002)	−0.00533*** (0.002)	−0.00396* (0.002)	−0.00542*** (0.002)
Observations	17,763	17,763	17,763	17,814	17,790
R-squared	0.274				
F stat	122.5	46.56	62.94	81.3	74.26
F p-value	0.000	0.000	0.000	0.000	0.000
Clusters	2,841	2,841	2,841	2,841	2,841

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clustered standard errors in parentheses | Regressions include the same set of control variables as models in Table 3.

between the two that exist in many contexts and to focus narrowly on only one component of rapidly expanding, globalized, and concentrated food systems.

The global burden of obesity and non-communicable diseases is increasingly shifting from high-income to low- and middle-income countries. Likewise, the most rapid dietary changes are currently observed in low- and middle-income countries. While the dynamics and consequences of Big Food and fast food for developing countries have been acknowledged in the literature, perhaps a better term to describe the ongoing changes is simply the expansion of 'modern global food systems'. Modern global food systems are characterized by a decreasing price differential between highly processed foods and beverages relative to fresh and unprocessed foods, and by industry efforts to make these highly processed foods desirable for large parts of the population. Foreign food and soft drink companies, supermarket chains, and fast food restaurants all play important roles, as well as their domestic counterparts that mimic global brands at lower prices (Traill, 2017). Better understanding the links between modernizing food systems and health outcomes is crucial to ensure that the gains from economic development are not eroded through a deterioration of public health and life expectancy.

CRediT authorship contribution statement

Steffen Otterbach: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Hamid Reza Oskorouchi:** Google API R Coding, Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Michael Rogan:** Conceptualization, Methodology, Validation, Investigation, Writing - original draft, Writing - review & editing. **Matin Qaim:** Conceptualization, Methodology, Validation, Investigation, Writing - original draft, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study uses the secure version of the South African National Income Dynamics Study (NIDS). All analyses were carried out at the DataFirst Secure Research Data Centre at the University of Cape Town. We wish to thank Alex Montgomery, Martin Wittenberg and Lynn Woolfrey from DataFirst and Tim Brophy from the NIDS SALDRU team for granting access to and supporting us with the secure version of NIDS. For useful comments and suggestions we are grateful to Alfonso Sousa-Poza, the University of Goettingen GobalFood team, participants at the 2019 Seminar Series in Economics of the University of Hohenheim, and two anonymous reviewers of this journal. We also appreciate that travel funding to the Secure Research Data Centre at the University of Cape Town (South Africa) was granted from the Universitätsbund Hohenheim e.V.

Appendix

Table A1
Descriptive statistics (NIDS 2017).

	Mean	Std. Dev.
BMI	27.160	7.108
Obese (BMI $\geq$ 30)	0.298	0.457
Overweight (BMI $\geq$ 25)	0.548	0.498
Central obesity (dummy)	0.563	0.496
Waist to height ratio (dummy)	0.558	0.109
Distance to closest supermarket (km)	7.567	12.019
Distance to closest fast food restaurant (km)	8.677	14.590
Distance to closest supermarket or fast food (km)	6.683	11.417
Net equivalized HH income (1000s ZAR)	3.428	6.264
HH size	4.375	3.156
Female (dummy)	0.559	0.496
Female HH head (dummy)	0.539	0.498
Having a partner (dummy)	0.382	0.486
Age (years)	40.027	15.280
Educational level (dummies)		
No schooling	0.053	0.224
Primary school	0.136	0.342
Secondary school	0.555	0.497
Higher Education	0.257	0.437
Population group (dummies)		
African	0.807	0.394
Colored	0.095	0.293
Indian/Asian	0.022	0.148
White	0.075	0.263
Employed	0.506	0.500
Doing exercise (dummies)		
Never	0.663	0.473
Less than once a week	0.080	0.272
Once a week	0.054	0.225
Twice a week	0.071	0.257
Three or more times a week	0.132	0.339
Being a smoker (dummy)	0.205	0.404
Subsistence farming (dummy)	0.112	0.315
Population density	639.260	850.126
Geographic classification (dummies)		
Urban	0.312	0.463
Traditional	0.644	0.479
Farms	0.044	0.205
Province (dummies)		
Western Cape	0.124	0.330
Eastern Cape	0.115	0.319
Northern Cape	0.029	0.167
Free State	0.055	0.228
KwaZulu-Natal	0.192	0.394
Northwest	0.055	0.228
Gauteng	0.248	0.432
Mpumalanga	0.089	0.284
Limpopo	0.093	0.291

Note: Data are weighted using post-stratification weights. The number of observations is 19,091.

Table A2
Marginal effects of distance to supermarket or fast food outlet.

Distance (km)	(1) Obese	(2) Overweight	(3) Central obesity	(4) WtHR
0	0.304***	0.556***	0.568***	0.653***
	(0.006)	(0.006)	(0.005)	(0.005)
10	0.295***	0.544***	0.561***	0.643***
	(0.005)	(0.006)	(0.005)	(0.005)
20	0.285***	0.533***	0.554***	0.634***
	(0.007)	(0.007)	(0.006)	(0.006)
30	0.276***	0.521***	0.547***	0.624***
	(0.010)	(0.010)	(0.009)	(0.009)
40	0.267***	0.509***	0.539***	0.614***
	(0.013)	(0.014)	(0.012)	(0.012)
Observations	19,090	19,090	19,150	19,122

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clustered standard errors in parentheses | Marginal effects are based on the regressions in Table 3.

Table A3
Distance to closest supermarket or fast food outlet, without geographic controls.

	BMI (1)	Obese (2)	Overweight (3)	Central obesity (4)	WtHR (5)
Distance to closest supermarket or fast food	−0.0234***	−0.00840***	−0.00850***	−0.00758***	−0.00765***
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
Subsistence farming (dummy)	−0.303	−0.0573	−0.043	−0.156**	−0.106
	(0.223)	(0.081)	(0.070)	(0.076)	(0.078)
Observations	19,091	19,091	19,091	19,151	19,123
R-squared	0.241				
F stat	148.4	60.78	80.84	106.1	96.27
F p-value	0.000	0.000	0.000	0.000	0.000
Clusters	2,841	2,841	2,841	2,841	2,841

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clustered standard errors in parentheses | Regressions include the same set of control variables as models in Table 3.

Table A4
Distance to closest supermarket or fast food outlet, additionally including average household income over time.

	BMI (1)	Obese (2)	Overweight (3)	Central obesity (4)	WtHR (5)
Distance to closest supermarket or fast food	−0.0135**	−0.00595**	−0.00546***	−0.00433*	−0.00592***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
Current income	0.0659*	0.0114	0.0380**	0.0281	0.0375*
	(0.035)	(0.012)	(0.019)	(0.018)	(0.022)
Average income over time	0.142***	0.0478**	0.0738***	0.0599**	0.0659***
	(0.049)	(0.019)	(0.026)	(0.026)	(0.023)
Observations	15,981	15,981	15,981	16,027	16,004
R-squared	0.273				
F stat	105.4	38.78	49.48	63.68	59.51
F p-value	0.000	0.000	0.000	0.000	0.000
Clusters	2,841	2,841	2,841	2,841	2,841

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ | Clustered standard errors in parentheses | Regressions include the same set of control variables as models in Table 3.

References

Adam, A., & Jensen, J. D. (2016). What is the effectiveness of obesity related interventions at retail grocery stores and supermarkets?—a systematic review. *BMC Public Health, 16*, 1247.

Asfaw, A. (2008). Does supermarket purchase affect the dietary practices of households? Some empirical evidence from Guatemala. *Development Policy Review, 26*, 227–243.

Bai, Y., Naumova, E. N., & Masters, W. A. (2020). Seasonality of diet costs reveals food system performance in East Africa. *Science Advances, 6* eabc2162.

Baiphethi, M. N., & Jacobs, P. T. (2009). The contribution of subsistence farming to food security in South Africa. *Agrekon, 48*, 459–482.

Battersby, J., & Peyton, S. (2014). The geography of supermarkets in Cape Town: Supermarket expansion and food access. *Urban Forum, 25*, 153–164.

Brophy, T., Branson, N., Daniels, R. C., Leibbrandt, M., Mlatsheni, C., & Woolard, I. (2018). National income dynamics study panel user manual. In Release 2018. Version 1. Cape Town: Southern Africa Labour and Development Research Unit.

BusinessTech South Africa. (2018). Accessed 31 August 2020. https://businesstech.co.za/news/business/220551/these-are-the-5-biggest-retailers-in-south-africa.

Campbell, M. (2016). South African supermarket expansion in sub-Saharan Africa. *Third World Thematics: A TWQ Journal, 1*, 709–725.

Chennakesavalu, M., & Gangemi, A. (2018). Exploring the relationship between the fast food environment and obesity rates in the US vs. abroad: A systematic review. *Journal of Obesity & Weight loss Therapy, 8*, 366–383.

Chinhema, M., Brophy, T., Brown, M., Leibbrandt, M., Mlatsheni, C., & Woolard, I. (2016). National income dynamics study panel user manual. In. Cape Town: Southern Africa Labour and Development Research Unit.

Chiolero, A., Faeh, D., Paccaud, F., & Cornuz, J. (2008). Consequences of smoking for body weight, body fat distribution, and insulin resistance. *The American Journal of Clinical Nutrition, 87*, 801–809.

D'Haese, M., & van Huylenbroeck, G. (2005). The rise of supermarkets and changing expenditure patterns of poor rural households case study in the Transkei area, South Africa. *Food Policy, 30*, 97–113.

Debela, B. L., Demmler, K. M., Klasen, S., & Qaim, M. (2020). Supermarket food purchases and child nutrition in Kenya. *Global Food Security, 25* 100341.

Demmler, K. M., Ecker, O., & Qaim, M. (2018). Supermarket shopping and nutritional outcomes: A panel data analysis for urban Kenya. *World Development, 102*, 292–303.

Demmler, K. M., Klasen, S., Nzuma, J. M., & Qaim, M. (2017). Supermarket purchase contributes to nutrition-related non-communicable diseases in urban Kenya. *PLOS ONE, 12* e0185148.
Fleischhacker, S., Evenson, K., Rodriguez, D., & Ammerman, A. (2011). A systematic review of fast food access studies. *Obesity Reviews, 12*, e460–e471.
Fraser, L. K., Edwards, K. L., Cade, J., & Clarke, G. P. (2010). The geography of fast food outlets: A review. *International Journal of Environmental Research and Public Health, 7*, 2290.
Hawkes, C. (2008). Dietary implications of supermarket development: A global perspective. *Development Policy Review, 26*, 657–692.
Khonje, M. G., Ecker, O., & Qaim, M. (2020). Effects of modern food retailers on adult and child diets and nutrition. *Nutrients, 12* 1714.
Khonje, M. G., & Qaim, M. (2019). Modernization of African food retailing and (un) healthy food consumption. *Sustainability, 11*, 4306.
Kimani-Murage, E., Kahn, K., Pettifor, J., Tollman, S., Dunger, D., Gómez-Olivé, X., & Norris, S. (2010). The prevalence of stunting, overweight and obesity, and metabolic disease risk in rural South African children. *BMC Public Health, 10*.
Kimenju, S. C., Rischke, R., Klasen, S., & Qaim, M. (2015). Do supermarkets contribute to the obesity pandemic in developing countries?. *Public Health Nutrition, 18*, 3224–3233.
Kwate, N. (2008). Fried chicken and fresh apples: Racial segregation as a fundamental cause of fast food density in black neighborhoods. *Health & Place, 14*.
Lamichhane, A. P., Warren, J., Puett, R., Porter, D. E., Bottai, M., Mayer-Davis, E. J., & Liese, A. D. (2013). Spatial patterning of supermarkets and fast food outlets with respect to neighborhood characteristics. *Health & Place, 23*, 157–164.
Maumbe, B. (2012). The rise of South Africa's quick service restaurant industry. *Journal of Agribusiness in Developing and Emerging Economies, 2*, 147–166.
Mayosi, B., Flisher, A., Lalloo, U., Sitas, F., Tollman, S., & Bradshaw, D. (2009). The burden of non-communicable diseases in South Africa. *Lancet, 374*, 934–947.
McClelland, W. G. (1962). Economics of the supermarket. *The Economic Journal, 72*, 154–170.
Mergenthaler, M., Weinberger, K., & Qaim, M. (2009). The food system transformation in developing countries: A disaggregate demand analysis for fruits and vegetables in Vietnam. *Food Policy, 34*, 426–436.
NCD Risk Factor Collaboration. (2019). Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature, 569, 260-264.
Neven, D., Reardon, T., Chege, J., & Wang, H. (2006). Supermarkets and consumers in Africa. *Journal of International Food & Agribusiness Marketing, 18*, 103–123.
Noble, M., & Wright, G. (2013). Using indicators of multiple deprivation to demonstrate the spatial legacy of apartheid in South Africa. *Social Indicators Research, 112*, 187–201.
Odegaard, A., Koh, W. P., Yuan, J.-M., Gross, M., & Pereira, M. (2012). Western-style fast food intake and cardiometabolic risk in an Eastern country. *Circulation, 126*, 182–188.
Okop, K. J., Ndayi, K., Tsolekile, L., Sanders, D., & Puoane, T. (2019). Low intake of commonly available fruits and vegetables in socio-economically disadvantaged communities of South Africa: Influence of affordability and sugary drinks intake. BMC Public Health, 19.
Peyton, S., Moseley, W., & Battersby, J. (2015). Implications of supermarket expansion on urban food security in Cape Town, South Africa. *African Geographical Review, 34*, 36–54.
Pingali, P. (2006). Westernization of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy, 32*, 281–298.
Popkin, B. M., Adair, L. S., & Ng, S. W. (2012). Global nutrition transition and the pandemic of obesity in developing countries. *Nutrition Reviews, 70*, 3–21.
Popkin, B. M., Corvalan, C., & Grummer-Strawn, L. M. (2020). Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet, 395*, 65–74.
Popkin, B. M., & Reardon, T. (2018). Obesity and the food system transformation in Latin America. *Obesity Reviews, 19*, 1028–1064.
Popkin, B. M., & Slining, M. (2013). New dynamics in global obesity facing low- and middle-income countries. *Obesity Research, 14*, 11–20.
Prentice, A. (2006). The emerging epidemic of obesity in developing countries. *International Journal of Epidemiology, 35*, 93–99.
Qaim, M. (2017). Globalisation of agrifood systems and sustainable nutrition. *Proceedings of the Nutrition Society, 76*, 12–21.
Reardon, T., Timmer, C. P., Barrett, C. B., & Berdegué, J. (2003). The rise of supermarkets in Africa, Asia, and Latin America. *American Journal of Agricultural Economics, 85*, 1140–1146.
Rischke, R., Kimenju, S. C., Klasen, S., & Qaim, M. (2015). Supermarkets and food consumption patterns: The case of small towns in Kenya. *Food Policy, 52*, 9–21.
Rogan, M. (2018). Food poverty, hunger and household production in rural Eastern Cape households. *Development Southern Africa, 35*, 90–104.
Rogan, M. (2020). Agricultural production, the household 'development cycle' and migrant remittances: Continuities and change in the Eastern Cape hinterland. In L. Bank, D. Posel, & F. Wilson (Eds.), *Migrant labour after apartheid: The inside Story Cape Town*. HSRC Press.
Ronquest-Ross, L. C., Vink, N., & Sigge, G. (2015). Food consumption changes in South Africa since 1994. *South African Journal of Science, 111*, 1–12.
Sartorius, B., Sartorius, K., Taylor, M., Aagaard-Hansen, J., Dukhi, N., Day, C., ... Hofman, K. (2017). Rapidly increasing body mass index among children, adolescents and young adults in a transitioning population, South Africa, 2008–15. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyx263, 1-11.
Schneider, H. J., Friedrich, N., Klotsche, J., Pieper, L., Nauck, M., John, U., ... Wittchen, H.-U. (2010). The predictive value of different measures of obesity for incident cardiovascular events and mortality. *The Journal of Clinical Endocrinology & Metabolism, 95*, 1777–1785.
Statistics South Africa (2017). *South Africa demographic and health survey 2016: Key Indicator Report, Statistics South Africa*. Pretoria: South African National Department of Health, Statistics South Africa and the South African Medical Research Council.
Steyn, N. P., Labadarios, D., & Nel, J. H. (2011). Factors which influence the consumption of street foods and fast foods in South Africa-A national survey. *Nutrition Journal, 10*, 104.
Stojakovic, A., Espinosa, E. P., Farhad, O. T., & Lutfy, K. (2017). Effects of nicotine on homeostatic and hedonic components of food intake. *Journal of Endocrinology, 235*, R13–R31.
Stuckler, D., McKee, M., Ebrahim, S., & Basu, S. (2012). Manufacturing epidemics: The role of global producers in increased consumption of unhealthy commodities including processed foods, alcohol, and tobacco. *PLoS Med, 9*, | e1001235.
Stuckler, D., & Nestle, M. (2012). Big food, food systems, and global health. *PLoS Med, 9*, 1–4.
Swinburn, B. A., Kraak, V. I., Allender, S., Atkins, V. J., Baker, P. I., Bogard, . . . Dietz, W. H. (2019). The global syndemic of obesity, undernutrition, and climate change: The Lancet Commission report. The Lancet, 393, 791-846.
Temple, N. J., & Steyn, N. P. (2009). Food prices and energy density as barriers to healthy food patterns in Cape Town, South Africa. *Journal of Hunger & Environmental Nutrition, 4*, 203–213.
Temple, N. J., & Steyn, N. P. (2011). The cost of a healthy diet: A South African perspective. *Nutrition, 27*, 505–508.
Temple, N. J., Steyn, N. P., Fourie, J., & De Villiers, A. (2011). Price and availability of healthy food: A study in rural South Africa. *Nutrition, 27*, 55–58.
Tessier, S., Traissac, P., Maire, B., Bricas, N., Eymard-Duvernay, S., El Ati, J., & Delpeuch, F. (2008). Regular users of supermarkets in Greater Tunis have a slightly improved diet quality. *Journal of Nutrition, 138*, 768–774.
The GBD 2015 Obesity Collaborators (2017). Health effects of overweight and obesity in 195 countries over 25 Years. *The New England Journal of Medicine, 377*, 13–27.
Timmer, C. P. (2009). Do supermarkets change the food policy agenda?. *World Development, 37*, 1812–1819.
Traill, W. B. (2017). Transnational corporations, food systems and their impacts on diets in developing countries. In Trade Policy Technical Note. New York: Food and Agriculture Organization of the United Nations.
Umberger, W. J., He, X., Minot, N., & Toiba, H. (2015). Examining the relationship between the use of supermarkets and over-nutrition in Indonesia. *American Journal of Agricultural Economics, 97*, 510–525.
Weatherspoon, D. D., & Reardon, T. (2003). The rise of supermarkets in Africa: Implications for agrifood systems and the rural poor. *Development Policy Review, 21*, 333–355.
WHO. (2020). Body Mass Index – BMI. Accessed 31 August 2020. https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi.
WHO. (n.d.). Recognizing adolescence. Accessed 28 August 2020. https://apps.who.int/adolescent/second-decade/section2/page1/recognizing-adolescence.html.
Wingrove, C., & Urban, B. (2017). Franchised fast food brands: An empirical study of factors influencing growth. *Acta Commercii, 17*, 431–439.
Wittenberg, M. (2013). The weight of success: The body mass index and economic well-being in Southern Africa. *Review of Income and Wealth, 59*, S62–S83.